THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PARIS-EST

Spécialité : **Génie Civil**

Présentée par

Alexandre CURY

pour obtenir le titre de Docteur
de l'Université Paris-Est

Titre

TECHNIQUES D'ANORMALITÉ APPLIQUÉES À LA SURVEILLANCE DE SANTÉ STRUCTURALE

soutenue le 16 décembre 2010, devant le jury composé de :

Prof. Edwin DIDAY	Université Paris-Dauphine / SYROKKO	*Président*
Prof. Jean-Claude GOLINVAL	Université de Liège	*Rapporteur*
Prof. Menad SIDAHMED	ENSIIE	*Rapporteur*
M. Bernard BASILE	Advitam	*Examinateur*
Prof. Fabrice THOUVEREZ	École Centrale de Lyon	*Examinateur*
Prof. Flávio BARBOSA	Université Fédérale de Juiz de Fora	*Examinateur*
M. André ORCESI	LCPC	*Examinateur*
M. Christian CREMONA	MEDDTL	*Directeur de thèse*

Remerciements

En tout premier lieu, je tiens à remercier le Professeur Edwin DIDAY, de l'Université Paris-Dauphine et directeur scientifique de la société SYROKKO, d'avoir accepté de présider ce jury, ainsi que les Professeurs Jean-Claude GOLINVAL, de l'Université de Liège et Menad SIDAHMED, de l'École Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise, d'avoir accepté d'être rapporteurs de mon travail.

Toute ma gratitude s'adresse aussi aux membres du jury, le Professeur Fabrice THOUVEREZ, de l'École Centrale de Lyon, le Professeur Flávio BARBOSA, de l'Université Fédérale de Juiz de Fora au Brésil et M. Bernard BASILE, de la societé Advitam, qui m'ont fait l'honneur de participer à la soutenance de cette thèse. En particulier, j'adresse mes plus sincères remerciements au Professeur Flávio BARBOSA, sans lequel mon séjour en France n'aurait pu être possible.

Je suis largement redevable à Monsieur Christian CREMONA, mon directeur de thèse, qui a encadré ma thèse avec une attention et une rigueur remarquables. Son assurance et sa connaissance scientifique m'ont beaucoup apporté tout au long de mon parcours. Avoir pu travailler avec lui a avant tout été un honneur. Les discussions, toujours fructueuses, ont permis non seulement d'atteindre les objectifs de cette thèse, mais aussi d'apporter une base solide à ma formation scientifique. Christian CREMONA a été un directeur de thèse exemplaire et j'ai pu apprécier ses grandes qualités tant sur le plan professionnel que personnel. Ceci renforce d'autant plus ma reconnaissance et mon admiration pour lui. Je lui en suis profondément reconnaissant et lui témoigne ici toute mon amitié.

Je tiens à remercier M. André ORCESI, qui m'a conseillé tout au long de ma thèse au Laboratoire Central des Ponts et Chaussées. André ORCESI a fait preuve de rigueur scientifique et professionnelle et je lui adresse mes remerciements les plus sincères pour ses remarques toujours pertinentes et ses relectures minutieuses de mon manuscrit. Qu'il trouve ici l'expression de ma profonde gratitude.

J'exprime toute ma reconnaissance à M. Bruno GODART, chef du Département Structures et Ouvrages d'Art, à M. François TOUTLEMONDE, adjoint au chef du Département Structures et Ouvrages d'Art, et à M. Jean-François SEIGNOL, chef de l'Unité Sécurité et Durabilité des Ouvrages (SDO), pour m'avoir accueilli et soutenu durant la thèse. Je tiens à remercier Jean-Luc CLEMENT et Agnès VAILLÉ au LCPC ainsi que Marine DANIEL à l'Université Paris-Est, pour les nombreux services que je leur ai demandés et pour leur amitié. Je remercie également M. Yves GAUTIER, de la DDTM 85, et M. John DUMOULIN, du CETE du Sud-Ouest (LR de Bordeaux), pour leur aide et leur disponibilité lors de l'exploitation des campagnes de mesures réalisées sur le pont PI-57 sur l'Oise. J'adresse également un grand merci à Filipe AFONSO, de la société SYROKKO, qui m'a beaucoup aidé au début de ma thèse.

Il est également important de préciser que j'ai trouvé un environnement scientifique et un cadre de travail d'excellente qualité dans l'Unité SDO. Ceci m'a permis d'achever ma thèse en toute sérénité et je suis ravi d'avoir pu travailler avec les agents de l'unité SDO pendant ces trois années. Ils m'ont accueilli comme si j'étais un membre de leur famille et j'en suis

vivement reconnaissant. Merci à Bruno KOUBI pour les discussions pertinentes sur les avancées informatiques, à Jean DAUTHUILLE, mon collègue de bureau, pour sa bonne humeur constante, à Lucas ADELAIDE pour sa disponibilité, ses conseils et toutes les discussions fructueuses que nous avons eues, à Minh ORCESI pour sa gentillesse et son aide précieuse au quotidien, et à Véronique BOUTEILLER pour sa sympathie et son soutien régulier. Je tiens aussi à remercier Franziska SCHMIDT, Jean-François CHERRIER, Binh TA, Nadia KAGHO et Mohamed SAHAL pour leurs encouragements et leur amitié.

Au cours de ma thèse j'ai eu la chance de pouvoir sympathiser avec de nombreuses personnes au LCPC qui m'ont aidé à rendre mon parcours plus aisé. Je pense en particulier à Alexandre DEMAN, Ali GANDOMZADEH, Alice SAAD (merci à elle pour son aide lors de l'impression des manuscrits), Dominique SIEGERT, Fernanda GOMES, Omar KHEMOUDJ, Raouf BARBARI, aux « membres-associés » de l'Hirondelle Café, Alexandre NASSIOPOULOS, André BENINE, Benjamin RICHARD, Eric MARTIN, Josselin PETIT, Romain GALLEN et Thierry BORE, et aux bons amis des « doctoriales » Adrien ACQUISTAPACE et Nicolas RADOMSKI.

Qu'il me soit enfin permis de remercier ma famille et mes amis au Brésil (en particulier, Raphael INFANTE et Janaína OLIVEIRA). En paraphrasant l'écrivain anglais A. Clarke, « *la distance rend toute chose infiniment plus précieuse* », je leur suis profondément reconnaissant pour leur amour et leur soutien constant.

Il faut préférer ce qui est impossible mais vraisemblable
à ce qui est possible, mais incroyable...

(*Aristote*)

À mes parents.

Résumé

Le paradigme de la surveillance de santé structurale repose sur l'introduction d'indicateurs fiables et robustes permettant de détecter, localiser, quantifier et prédire un endommagement de manière précoce. En effet, la détection d'une modification structurale susceptible de devenir critique peut éviter l'occurrence de dysfonctionnements majeurs associés à des conséquences sociales, économiques et environnementales très importantes.

Ces dernières années, de nombreuses recherches se fait de l'évaluation dynamique un élément de diagnostic. La plupart des méthodes reposent sur une analyse temporelle ou fréquentielle des signaux pour en extraire une information compressée au travers de quelques caractéristiques modales ou d'indicateurs évolués construits sur ces caractéristiques. Ces indicateurs ont montré leur efficacité, mais le problème de leur sensibilité, de la nécessité de disposer d'un état de référence, et de leur fiabilité en terme de la probabilité de détection et de fausses alarmes, reste entier. De plus, le fait d'utiliser des mesures dynamiques (particulièrement si plusieurs voies de mesures sont considérées) mène au stockage de grands volumes de données.

Dans ce contexte, il est important d'employer des techniques permettant d'utiliser autant des données brutes que les propriétés modales de manière pratique et pertinente. Pour cela, des représentations adaptées ont été développées pour améliorer la manipulation et le stockage des données. Ces représentations sont connues sous le nom de « données symboliques ». Elles permettent de caractériser la variabilité et l'incertitude qui entachent chacune des variables. Le développement de nouvelles méthodes d'analyse adéquates pour traiter ces données est le but de l'Analyse de Données Symboliques (ADS).

L'objectif de cette thèse est double : le premier consiste à utiliser différentes méthodes couplées à l'ADS pour détecter un endommagement structural. L'idée est d'appliquer des procédures de classification non supervisée (e.g. divisions hiérarchiques, agglomérations hiérarchiques et nuées dynamiques) et supervisée (e.g., arbres de décision Bayésiens, réseaux de neurones et machines à vecteurs supports) afin de discriminer les différents états de santé d'une structure. Dans le cadre de cette thèse, l'ADS est appliquée aux mesures dynamiques acquises *in situ* (accélérations) et aux paramètres modaux identifiés. Le deuxième objectif est la compréhension de l'impact des effets environnementaux, notamment de ceux liés à la variation thermique, sur les paramètres modaux. Pour cela, des techniques de régression des données sont proposées.

Afin d'évaluer la pertinence des démarches proposées, des études de sensibilité sont menées sur des exemples numériques et des investigations expérimentales. Il est montré que le couplage de l'ADS aux méthodes de classification de données permet de discriminer des états structuraux avec un taux de réussite élevé. Par ailleurs, la démarche proposée permet de vérifier l'importance d'utiliser des techniques permettant de corriger les propriétés modales identifiées des effets thermiques, afin de produire un processus de détection d'endommagements efficace.

Abstract

The paradigm of structural health monitoring is based on the development of reliable and robust indicators able to detect, locate, quantify and predict damage. Studies related to damage detection in civil engineering structures have a noticeable interest for researchers in this area. Indeed, the detection of structural changes likely to become critical can avoid the occurrence of major dysfunctions associated with social, economic and environmental consequences.

Recently, many researches have focused on dynamic assessment as part of structural diagnosis. Most of the studied techniques are based on time or frequency domain analyses to extract compressed information from modal characteristics or based on indicators built from these parameters. These indicators have shown their potentialities, but the problem of their sensitivity, the necessity of a reference state, and their reliability in terms of detection probability and false alarm, still remains. Moreover, the use of raw dynamic measurements (especially if several measurement channels are considered) leads to the storage of large datasets.

In this context, it is important to use techniques capable of dealing not only with raw data but also modal parameters in a practical and relevant way. In order to give some insights to this problem, appropriate representations have been developed to improve both manipulation and storage of data. These representations are known as « symbolic data ». They are used to characterize the variability and uncertainty that exists within each variable. The development of new methods capable of dealing with this type of data is the goal of Symbolic Data Analysis (SDA).

This thesis has two main objectives : the first one is to use different methods coupled with the SDA to detect structural damage. The idea is to employ clustering procedures (e.g., hierarchy-divisive, hierarchy-agglomerative and dynamic clouds) and supervised classification methods (e.g., Bayesien decision trees, neural networks and support vector machines) to discriminate different structural states. In this thesis, SDA is applied to dynamic measurements obtained on site (accelerations) and to the identified modal parameters. The second goal is to study the impact of environmental effects, particularly those related to thermal variation over modal parameters. To this end, a couple of regression techniques are proposed.

In order to attest the efficiency of the proposed approaches, several sensibility studies considering numerical applications and experimental investigations are carried out. It is shown that SDA coupled with classification methods is able to distinguish structural conditions with adequate rates. Furthermore, it is stressed the importance of using techniques capable of correcting modal parameters from thermal effects in order to build efficient procedures for damage detection.

Table des matières

Liste de notations **vi**

Introduction **1**

1 Le paradigme de la surveillance de santé structurale **5**

1.1 Introduction 5

1.2 État de l'art des méthodes de détection d'endommagements 7

1.2.1 Méthodes basées sur le changement des paramètres modaux 8

1.2.2 Méthodes basées sur le changement des indicateurs dits évolués 9

1.2.3 Méthodes basées sur les indicateurs d'anormalité 10

1.3 Étude des effets des variations thermiques 12

1.4 Objectifs 14

2 Analyse de Données Symboliques **17**

2.1 Introduction 17

2.2 Des données classiques aux données symboliques 18

2.3 Bilan 23

3 Classification et régression des données **25**

3.1 Méthodes de classification non supervisée 26

3.1.1 Calcul de mesures de dissimilarité et d'inertie 26

3.1.2 Méthode des divisions hiérarchiques 28

3.1.3 Méthode des agglomérations hiérarchiques 32

3.1.4 Méthode des nuées dynamiques 35

3.1.5 Indices de partitionnement optimal 37

3.1.6 Affectation de nouveaux essais 38

3.1.7 Bilan 41

3.2 Méthodes de classification supervisée 41

3.2.1 Arbres de décision Bayésiens 42

3.2.2 Réseaux de neurones 45

3.2.3 Machines à vecteurs support 51

3.3 Bilan . . . 55
3.4 Analyse des techniques de régression . . . 56
3.4.1 Régression Linéaire Multiple . . . 56
3.4.2 Régression Non Linéaire . . . 58
3.4.3 Applications . . . 58
3.4.4 Bilan . . . 60
3.5 Analyse en Composantes Principales . . . 61
3.5.1 Méthodologie . . . 61
3.5.2 Bilan . . . 63

4 Validation numérique et étude de sensibilité 65
4.1 Présentation du cas d'étude . . . 65
4.2 Analyse classique . . . 68
4.3 ADS couplée aux méthodes de clustering . . . 71
4.3.1 Analyse de la sensibilité au bruit . . . 71
4.3.2 Études complémentaires . . . 84
4.3.3 Indices de partitionnement optimal . . . 86
4.3.4 Bilan . . . 92
4.4 ADS couplée aux méthodes de classification supervisée . . . 93
4.4.1 Analyse de la sensibilité au bruit . . . 94
4.4.2 Études complémentaires . . . 96
4.4.3 Bilan . . . 100
4.5 Conclusions . . . 102

5 Étude du pont-rail PK 075+317 105
5.1 Présentation du cas d'étude . . . 105
5.2 Identification modale . . . 108
5.3 Étude des effets thermiques sur les paramètres modaux . . . 111
5.4 Analyse Classique . . . 113
5.4.1 Sans prise en compte des effets thermiques . . . 113
5.4.2 Avec prise en compte des effets thermiques . . . 116
5.5 ADS couplée aux méthodes de clustering . . . 118
5.5.1 Signaux . . . 119
5.5.2 Fréquences propres . . . 120
5.5.3 Modes propres . . . 124
5.5.4 Étude complémentaire . . . 127
5.5.5 Indices de partitionnement optimal . . . 134
5.5.6 Bilan . . . 139

5.6 ADS couplée aux méthodes de classification supervisée . . . 142
5.6.1 Première étude . . . 142
5.6.2 Deuxième étude . . . 146
5.6.3 Troisième étude . . . 149
5.6.4 Bilan . . . 151
5.7 Affectation de nouveaux essais . . . 154
5.7.1 Étude des effets thermiques sur les paramètres modaux . . . 154
5.7.2 Affectation basée sur les divisions hiérarchiques . . . 157
5.7.3 Affectation basée sur les nuées dynamiques . . . 160
5.7.4 Affectation basée sur le calcul de distances . . . 162
5.7.5 Affectation basée sur les méthodes de classification supervisée . . . 165
5.7.6 Bilan . . . 167
5.8 Conclusions . . . 169

6 Étude du pont PI-57 sur l'Oise 173
6.1 Présentation du cas d'étude . . . 173
6.2 Identification modale . . . 176
6.3 Étude des effets thermiques sur les paramètres modaux . . . 180
6.3.1 Étude des modèles de prédiction . . . 185
6.4 Analyse classique . . . 200
6.4.1 Sans prise en compte des effets thermiques . . . 200
6.4.2 Avec prise en compte des effets thermiques . . . 201
6.5 ADS couplée aux méthodes de clustering . . . 202
6.5.1 Analyse sur l'ensemble des essais . . . 203
6.5.2 Application aux paramètres modaux corrigés . . . 206
6.5.3 Comparaison des phases 1 et 2 mois par mois . . . 208
6.5.4 Indices de partitionnement optimal . . . 215
6.5.5 Bilan . . . 221
6.6 ADS couplée aux méthodes de classification supervisée . . . 223
6.6.1 Analyse sur l'ensemble des essais . . . 225
6.6.2 Comparaison des phases 1 et 2 mois par mois . . . 227
6.6.3 Bilan . . . 231
6.6.4 Conclusions . . . 233

Conclusions 237

Références Bibliographiques 243

A Résultats complémentaires - Application Numérique 253

A.1 Analyse classique . . . 253
A.2 ADS couplée aux méthodes de clustering . . . 256
A.2.1 Analyse de la sensibilité au bruit . . . 256
A.2.2 Bilan des résultats . . . 263
A.2.3 Études complémentaires . . . 265
A.2.4 Indices de partitionnement optimal . . . 269
A.3 Méthodes supervisées . . . 274
A.3.1 Analyse de la sensibilité au bruit . . . 275
A.3.2 Études complémentaires . . . 278
A.3.3 Bilan des résultats . . . 281

B Résultats complémentaires - Pont PK 075 + 317 285
B.1 Identification Modale . . . 285
B.2 Analyse classique avec prise en compte des effets thermiques . . . 286
B.3 ADS couplée aux méthodes de clustering . . . 287
B.3.1 Signaux . . . 287
B.3.2 Fréquences propres . . . 289
B.3.3 Modes propres . . . 293
B.3.4 Étude complémentaire . . . 296
B.3.5 Indices de partitionnement optimal . . . 302
B.4 ADS couplée aux méthodes de classification supervisée . . . 307
B.4.1 Première étude . . . 307
B.4.2 Deuxième étude . . . 310
B.4.3 Troisième étude . . . 312
B.5 Affectation de nouveaux essais . . . 315
B.5.1 Affectation basée sur les divisions hiérarchiques . . . 315
B.5.2 Affectation basée sur les nuées dynamiques . . . 317
B.5.3 Affectation basée sur le calcul de distances . . . 319
B.5.4 Affectation basée sur les méthodes de classification supervisée . . . 321

C Résultats complémentaires - Pont PI-57 sur l'Oise 325
C.1 Étude des effets thermiques . . . 325
C.1.1 Étude des modèles de prédiction . . . 326
C.2 Analyses classiques . . . 330
C.3 ADS couplée aux méthodes de clustering . . . 330
C.3.1 Analyse sur l'ensemble des essais . . . 330
C.3.2 Application aux paramètres modaux corrigés . . . 333
C.3.3 Comparaison des phases 1 et 2 mois par mois . . . 335

C.3.4 Indices de partitionnement optimal 342
C.4 ADS couplée aux méthodes de classification supervisée 346
C.4.1 Analyse sur l'ensemble des essais 347
C.4.2 Comparaison des phases 1 et 2 mois par mois 349

D Méthodes d'identification modale **361**
D.1 Méthode des Réalisations Stochastiques (MRS) 361
D.2 Méthode du Décrément Aléatoire (MDA) 364
D.2.1 Couplage MRS-MDA 366
D.3 Méthode d'Ibrahim 366
D.3.1 Lien avec la MDA 366
D.3.2 Extraction des paramètres modaux 368

Notations

Majuscules latines

B	Inertie inter-clusters
C^*	Indice de partitionnement optimal
$\mathbf{C}^i$	$i^{\text{ème}}$ cluster
CH	Indice de partitionnement optimal Calinski Harabasz
D	Nombre de neurones d'une couche cachée
$\widetilde{\mathbf{E}}$	Essai inconnu à classer
$E_{c,i}$	Échantillon de données i soumis à un test d'hypothèse
$\mathbf{E}_i$	Essai dynamique i
$F(.)$	Notation d'une fonction de répartition
$\mathbf{F}$	Matrice des fréquences propres
$\hat{\mathbf{F}}$	Matrice des fréquences propres reprojetées dans l'espace d'origine
$\mathbf{G}_i$	Prototype d'un cluster $\mathbf{C}^i$
$\mathbf{G}_i^*$	Prototype actualisé d'un cluster $\mathbf{C}^i$
H	Hyperplan séparateur de différentes classes
H_0	Hypothèse nulle d'un test d'hypothèse
H_1	Hypothèse alternative d'un test d'hypothèse
$I(.)$	Inertie intra-cluster
$K(.)$	Fonction noyau
K_s	Valeur de la « statistique de test » d'un test d'hypothèse
$\tilde{L}$	Lagrangien de la formulation duale des SVM
M	Paramètre indicateur d'une classe
N_i	Taille d'un échantillon de données $E_{c,i}$ d'un test d'hypothèse
$\mathbf{P}_r$	Partition avec r clusters
$\mathbf{R}$	Matrice d'auto-corrélation
S^k	Somme des distances entre les k essais d'un cluster considéré
S^k_{max}	Somme des k plus grandes distances parmi tous les essais d'un cluster
S^k_{min}	Somme des k plus petites distances parmi tous les essais d'un cluster
T_{ref}	Température de référence
$\mathbf{T}$	Matrice de températures utilisée pour la construction des modèles de régression
$\widehat{\mathbf{T}}$	Matrice de températures utilisée pour la prédiction des paramètres modaux

$\mathbf{U}$	Matrice de sortie d'une transformation ACP
$\mathbf{V}$	Matrice d'entrée d'une transformation ACP
$V \sim N(0,1)$	Variable gaussienne de moyenne nulle et d'écart-type unitaire
$\mathbf{X}$	Matrice de covariance des fréquences propres
X_i	Réponse dynamique non bruitée mesurée par un capteur i
$X_{i,\text{bruitée}}$	Réponse dynamique bruitée mesurée par un capteur i
$\mathbf{Y}$	Hiérarchie de clusters
W	Inertie intra-cluster totale
$\mathbf{Y}$	Hiérarchie de clusters
$\mathbf{Z}$	Matrice des fréquences propres transformées par ACP

Minuscules latines

i, j, k, p, q, u	Compteurs entiers
$b_i^{q,k}$	Proportion correspondante à la $k^{\text{ème}}$ catégorie du $q^{\text{ème}}$ capteur d'un essai $\mathbf{E}_i$
c_j	Étiquette d'un neurone j
cg	Centre de gravité d'un histogramme
$cov(.)$	Covariance entre variables aléatoires
$d(.)$	Éléments de la matrice de distances
$dg(.)$	Éléments de la matrice de degrés de généralité
$\mathbf{dg}$	Matrice des degrés de généralité
ds_i	Distance seuil d'une distribution de probabilités
e_j	Erreur de sortie d'un neurone j
f_i	Fréquence propre mesurée
$\hat{f}_i$	Prédiction d'une fréquence propre par un modèle de régression
f_{ref}	Fréquence propre de référence
$f(.)$	Fonctions de densité de probabilité
$\mathbf{f}_{\text{corr}}$	Vecteur des fréquences propres corrigées par un modèle de régression
$\mathbf{f}$	Vecteur des fréquences propres identifiées
g	Nombre de caractéristiques d'un essai dynamique
h	Paramètre de lissage des fonctions noyau
(l)	Dénotation d'une couche d'un réseau de neurones
m	Nombre de catégories d'un histogramme
n	Nombre total d'essais
n_{bruit}	Niveau de bruit (en pourcentage) ajouté à une réponse dynamique
n_i	Nombre d'itérations utilisées dans la méthode des nuées dynamiques
n_k	Nombre de tests dans le cluster $\mathbf{C}^k$ ou dans une classe k
n_r	Nombre de répétitions utilisées dans la méthode des nuées dynamiques
$\mathbf{p}$	Matrice de corrélation croisée
p^*	Probabilité d'acceptation d'un test d'hypothèse ($1^{\text{ère}}$ espèce)
q	Dimension réduite de l'espace après transformation par ACP
r	Nombre de clusters considérés

$\bar{\mathbf{r}}$	Vecteur des ratios normalisées des fréquences propres
$\bar{\mathbf{r}}^*_{\text{corr}}$	Vecteur des ratios normalisées des fréquences propres corrigés
s	Nombre de clusters utilisés pour le calcul des indices de partitionnement optimal
(t)	Itération d'actualisation des poids d'un réseau de neurones
$t_{\alpha,\gamma}$	Valeur de la distribution de t-Student à γ degrés de liberté correspondant à la probabilité α
u_j	Valeur de sortie intermédiaire d'un neurone j
v	Nombre de variables (capteurs ou fréquences ou modes utilisées)
$var(.)$	Variance entre variables aléatoires
vc	Valeur de coupure pour la définition du partitionnement optimal
vm	Valeur moyenne d'un intervalle interquartile
w_j	Poids synaptique d'un neurone j
$\mathbf{w}^*$	Vecteur des poids calibrés
$\mathbf{w}^*_{\text{corr}}$	Vecteur des poids calibrés après correction des effets thermiques
y_j	Valeur de sortie d'un neurone j
$z_i^{(l-1)}(t)$	Sortie du neurone i de la couche $(l-1)$ à la $t^{\text{ème}}$ itération

Majuscules grecques

Γ	Indice de partitionnement optimal Γ
$\Gamma_+(P_j)$	Nombre de distances inter-cluster plus grandes que les distances intra-clusters
$\Gamma_-(P_j)$	Nombre de distances inter-cluster plus petites que les distances intra-clusters
Λ	Matrice des valeurs singulières d'une transformation ACP
$\mathbf{\Xi}$	Matrice d'erreur résiduelle d'une transformation ACP
Σ	Fonction de combinaison d'un réseau de neurones
Υ	Opérateur d'échelle utilisé dans les modèles de régression non linéaire
Φ	Opérateur d'une transformation non linéaire
Ψ	Matrice de transformation d'une ACP
$\mathbf{\Omega}$	Ensemble d'essais

Minuscules grecques

α_i	Multiplicateurs de Lagrange
β	Biais d'un neurone
δ_j	Gradient local d'un neurone j utilisé dans la règle de delta modifiée
$\boldsymbol{\epsilon}$	Vecteur erreur du modèle de régression
η	Taux d'apprentissage utilisé dans la règle de delta modifiée
κ	Indice d'endommagement utilisé dans l'ACP
λ	Valeur propre d'une matrice
μ	Constante de moment utilisée dans la règle de delta modifiée
ρ_i	Probabilité *a priori* d'une classe i
σ_{X_i}	Écart-type de la réponse dynamique X_i

$\varphi(.)$	Mesure de distance
$\boldsymbol{\phi}$	Vecteur propre d'une matrice
$\psi(.)$	Fonction d'activation d'un neurone

Abréviations

AH	(Méthode) des Agglomérations Hiérarchiques
ACP	Analyse en Composantes Principales
ADB	Arbres de Décision Bayésiens
DH	(Méthode) des Divisions Hiérarchiques
EMPA	Laboratoire Fédéral d'Essai des Matériaux et de Recherche
LCPC	Laboratoire Central des Ponts et Chaussées
LRPC	Laboratoire Régional des Ponts et Chaussées
MAC	Modal Assurance Criterion
MDA	Méthode du Décrement Aléatoire
MRS	Méthode des Réalisations Stochastiques
ND	(Méthode) des Nuées Dynamiques
RN	Réseau de Neurones
SANEF	Société des Autoroutes du Nord et de l'Est de la France
SNCF	Société Nationale des Chemins de fer Français
SVM	Machines à Vecteur Support (*Support Vector Machines*)

Introduction

Le paradigme de la surveillance de santé structurale repose sur l'introduction d'indicateurs fiables et robustes permettant de détecter, localiser, quantifier et prédire un endommagement. Les études liées à la détection d'endommagements dans les structures du génie civil ont un intérêt perceptible pour les chercheurs dans ce domaine. En effet, la détection d'une modification structurale susceptible de devenir critique peut éviter l'occurrence de dysfonctionnements majeurs associés à des conséquences sociales, économiques et environnementales très importantes. Ces dernières années, de nombreuses recherches ont porté sur l'utilisation de l'évaluation dynamique comme élément de diagnostic par le développement de divers indicateurs basés sur les caractéristiques modales, c'est-à-dire les fréquences propres, les taux d'amortissement et les modes propres identifiés par une analyse modale opérationnelle (Doebling et al. 1996, Salawu 1997, Crémona 2004). Ces quantités sont directement affectées par les changements des propriétés physiques de la structure (impactant par exemple la masse et la raideur de la structure). De manière générale, les processus d'endommagement réduisent principalement la rigidité de la structure et modifient ses caractéristiques vibratoires (Pandey & Biswas 1994).

La plupart de ces indicateurs reposent sur une analyse temporelle ou fréquentielle des signaux pour en extraire une information compressée au travers de quelques caractéristiques modales ou d'indicateurs évolués construits sur ces caractéristiques. Ces indicateurs ont montré leur efficacité, mais le problème de leur sensibilité, de la nécessité de disposer d'un état de référence, et de leur fiabilité en termes de la probabilité de détection et de fausses alarmes, reste entier. Par ailleurs, l'identification des paramètres modaux est un procédé de filtrage, ce qui peut conduire à une perte d'information comparée aux données brutes. Cette perte de l'information peut masquer des changements de faible amplitude dus à une modification structurale. En outre, un inconvénient majeur lié à l'utilisation des paramètres modaux est qu'ils décrivent essentiellement un comportement linéaire de la structure, ce qui n'est pas toujours le cas.

Plusieurs méthodes de détection d'endommagements basées sur des principes de signature existent dans la littérature, mais elles échouent habituellement dans la pratique, faute d'outils efficaces ou des difficultés pour gérer de telles quantités de données brutes. Ainsi, en dépit de la capacité de traitement courante offerte par les ordinateurs, le coût de calcul nécessaire pour manipuler de grands ensembles de données demeure un problème. De plus, le fait d'utiliser des mesures dynamiques (particulièrement si plusieurs voies de mesures sont considérées) mène au stockage de grands volumes de données. Les mesures dynamiques peuvent facilement contenir plusieurs milliers de valeurs rendant ce processus d'analyse délicat, voire même impossible dans certains cas.

Dans ce contexte, il est important d'employer des techniques permettant d'utiliser les données brutes de manière pratique et pertinente. L'Exploitation de Données ou Fouille de Données (ou encore *Data Mining*, en anglais) consiste en un ensemble d'outils qui extraient des caractéristiques cachées dans des grandes bases de données brutes (Hastie et al. 2009). Plus le nombre de données enregistrées augmente, plus l'exploitation de l'information contenue est complexe et justifie le recours à ces outils. Cette démarche est utilisée dans de nombreux domaines, tels que l'économie,

le marketing, la détection de fraudes, etc. Afin de donner quelques éléments de réponse à ce problème, des représentations adaptées ont été développées pour améliorer le stockage des données. Ces représentations sont connues sous le nom de « données symboliques ». Elles permettent de caractériser la variabilité et l'incertitude qui entachent chacune des variables considérées. Le développement de nouvelles méthodes d'analyse adéquates pour traiter ces données est le but de l'Analyse de Données Symboliques (ADS) (Billard & Diday 2006). Actuellement, la plupart des techniques les plus développées dans l'ADS sont des prolongements des méthodes statistiques existantes.

L'objectif de cette thèse est double : le premier consiste à utiliser différentes méthodes de classification des données couplées à l'ADS pour détecter un endommagement structural. L'idée est d'utiliser des procédures de clustering (e.g. divisions hiérarchiques, agglomérations hiérarchiques et nuées dynamiques) et des méthodes de classification supervisée (e.g., réseaux de neurones, arbres de décision Bayésiens et machines à vecteurs supports) afin de discriminer les différents états de santé d'une structure. Dans le cadre de cette thèse, l'approche proposée est appliquée aux mesures dynamiques acquises *in situ* (accélérations) et aux paramètres modaux identifiés, le but étant de détecter des modifications structurales. Le deuxième objectif est la compréhension de l'impact des effets environnementaux, notamment de ceux liés à la variation thermique, sur les paramètres modaux. De nombreuses recherches ont permis démontrer que la prise en compte de ces variations est nécessaire pour différencier efficacement des événements endommageants d'autres non endommageants (Sohn et al. 1999, Yan et al. 2005). Pour cela, des techniques de régression linéaire et non linéaire sont proposées. L'idée est de pouvoir caractériser une variation « normale » des paramètres modaux d'une variation « anormale ». Dans cette thèse, les techniques de régression linéaire et non linéaire sont couplées aux méthodes de classification de données, le but étant d'observer comment la variation thermique peut influencer la détermination des groupes de comportements structuraux. Une approche basée sur l'Analyse en Composantes Principales (ACP) est proposée afin de discriminer des variations fréquentielles liées aux changements thermiques, de celles dues à la présence d'endommagements structuraux lorsqu'aucune mesure de température est disponible. Cette démarche présente donc l'avantage de s'affranchir des mesures de températures. Cette approche est également couplée aux techniques de classification pour l'identification plus fine d'essais « anormaux ».

Afin de répondre à ces enjeux, la thèse s'articule en six parties. La première partie rend compte d'un état de l'art détaillant les principales méthodes utilisées pour la détection d'endommagements par essais dynamiques. Ces méthodes sont divisées en trois groupes : celles basées sur la variation des paramètres modaux, celles basées sur la variation des indicateurs dits « évolués », et celles basées sur de « nouveaux indicateurs » (*novelty detection*). Ces dernières reposent sur l'analyse statistique et la reconnaissance de formes, telles que la fusion de données, les réseaux neuronaux, les réseaux Bayésiens, etc. Par ailleurs, un bref état de l'art des recherches effectuées sur la prise en compte des effets environnementaux sur les fréquences et modes propres est proposé.

Le deuxième chapitre a pour objectif d'introduire les concepts majeurs de l'ADS. Une introduction générale de l'ADS met en évidence la manière dont les données sont manipulées. Dans un premier temps, des aspects généraux sur le *Data Mining* et la création de l'ADS sont présentés. Dans un second temps, des exemples illustrant la transformation de données classiques en données symboliques sont introduits.

Le troisième chapitre porte sur les méthodes de classification et les techniques de régression de données utilisées dans le cadre de la thèse. Les méthodes de classification de données sont regroupées en deux familles : les méthodes dites non supervisées et les méthodes dites supervisées. Une fois les méthodes de classification présentées, deux types de techniques de régression des données sont introduits : linéaire et non linéaire. L'objectif est de construire des modèles de régression

capables de filtrer les effets thermiques des paramètres modaux. Le but est aussi d'appliquer les méthodes de classification de données décrites auparavant aux fréquences et modes propres « corrigés » par les modèles de régression. Par ailleurs, une démarche similaire est appliquée en considérant l'approche basée sur l'ACP.

Le quatrième chapitre présente une application numérique basée sur un modèle de poutre discretisée par éléments finis. L'objectif est de réaliser une étude de sensibilité des approches proposées en simulant plusieurs niveaux d'endommagement et de bruit. Cette étude permet de réaliser des simulations contrôlées afin d'observer comment les méthodes de classification peuvent être influencées par le bruit existant dans les mesures dynamiques.

Les cinquième et sixième chapitres portent sur des applications expérimentales. La première application est basée sur des essais dynamiques effectués sur le pont-rail PK 075+317 (structure à poutrelles enrobées faisant partie de la ligne des trains à grande vitesse reliant Paris à Lyon, en France). L'objectif est de caractériser et vérifier *in situ* les améliorations apportées par une procédure de resserrage des bielles sur le comportement de l'ouvrage. Le but est aussi d'évaluer l'impact des variations environnementales (en particulier celles liées à la température), sur les caractéristiques modales. La deuxième application expérimentale est basée sur deux campagnes de surveillance dynamique réalisées sur l'un des tabliers du pont routier PI-57, qui relie Lille à Paris et franchit l'Oise. Cet ouvrage a subi des travaux de renforcement par précontrainte additionnelle durant l'été 2009. Comme pour la première application expérimentale, l'objectif est ici double. Dans un premier temps, les méthodes de classification proposées sont mises en œuvre afin de discriminer les campagnes de mesures menées avant et après travaux. Dans un deuxième temps, les techniques de régression et l'ACP sont utilisées afin d'apprécier l'influence des effets thermiques sur les paramètres modaux identifiés.

Finalement, les conclusions de l'étude réalisée sont tirées afin de mettre en évidence les apports des travaux menés et les perspectives sur des recherches futures.

Chapitre 1

Le paradigme de la surveillance de santé structurale

L'objectif de ce chapitre est de présenter un bref état de l'art des techniques utilisées pour la surveillance de santé structurale par essais dynamiques. Dans un premier temps, une discussion sur la problématique et l'importance des campagnes de mesures vibratoires sur ouvrages est proposée. Dans un deuxième temps, les principales méthodes existantes dans la littérature, utilisées pour la détection de dommages structuraux, sont exposées. Finalement, la prise en compte des effets environnementaux sur la surveillance des ouvrages est abordée. Cette introduction sur le contexte et les enjeux de la surveillance structurale des ouvrages permet alors de mettre en relief les objectifs majeurs de la thèse.

1.1 Introduction

Ces dernières années, de nombreuses recherches ont été réalisées afin d'introduire des indicateurs fiables et robustes permettant de détecter, localiser, quantifier ou même prédire un endommagement. Les études menées se sont principalement intéressées à l'utilisation de l'évaluation dynamique comme élément de diagnostic pour le développement d'indicateurs basés sur les caractéristiques modales identifiées par une analyse modale opérationnelle. Dans ce contexte, de nombreuses questions se posent. En particulier, les deux questions suivantes sont jugées essentielles : « quelles sont les causes des dysfonctionnements d'une structure ? » et « comment détecter de tels dysfonctionnements ? ».

Répondre à la première question de manière directe est un sujet délicat. En effet, de multiples raisons peuvent expliquer un dysfonctionnement structural. En particulier, la surveillance des ouvrages existants nécessite de prendre en compte les différentes étapes de la vie de l'ouvrage (investigations déjà menées, précédentes actions de maintenance et de réhabilitation, etc.). L'augmentation de la durée de vie de service des ouvrages les plus anciens peut alors justifier une surveillance structurale renforcée. Toutefois, il est noté que même des structures récentes peuvent présenter des défauts de mise en œuvre et/ou subir des événements extrêmes, pouvant entraîner une défaillance localisée, voire globale à l'échelle de la structure. L'apparition de dommages structuraux peut être liée (i) aux modifications des capacités de résistance des ouvrages du fait de la dégradation de ses matériaux constitutifs (e.g. corrosion des aciers, fissures de fatigue, etc.) ou bien (ii) aux différentes sollicitations que l'ouvrage subit. En effet, les structures du génie civil sont généralement soumises à des chargements successifs au cours de leur vie qui peuvent causer

des dommages structuraux de façon continue dans le temps (e.g. fatigue des assemblages métalliques sous l'effet du trafic des véhicules) ou bien de manière discontinue (e.g. occurrence d'un séisme). Une structure peut également dysfonctionner du fait d'actions accidentelles identifiées ou non identifiées.

Un dommage structural peut être défini par des modifications de la structure affectant sa performance (e.g. sa sécurité structurale, son aptitude au service) actuelle et/ou future (EN1990 1990). De façon explicite, le concept d'un dommage n'est généralement significatif que si deux états structuraux sont identifiés et comparés (en considérant alors le premier état comme celui de référence, non endommagé généralement). Pour les systèmes structuraux du génie civil, les effets d'un dommage sont essentiellement limités à des modifications au niveau du matériau et/ou des propriétés géométriques. Par ailleurs, les modifications des conditions limites et de la connectivité du système peuvent aussi affecter la performance structurale. Un processus de fissuration, par exemple, peut produire un changement dans la géométrie pouvant finalement mener à l'effondrement de la structure dans le cas d'un mécanisme de défaillance en chaîne. Par ailleurs, en fonction de la dimension, de la localisation de cette fissure et des charges appliquées à la structure, les effets d'un dommage sur la performance du système peuvent être immédiats ou différés dans le temps. En général, deux échelles distinctes peuvent être introduites pour caractériser un endommagement structural : une première, dimensionnelle, et une deuxième, temporelle. Pour l'échelle dimensionnelle, tout endommagement commence au niveau de l'élément. Selon les scénarios de chargements appliqués, cet endommagement peut alors progresser jusqu'à la défaillance complète de la structure. Pour l'échelle temporelle, le dommage peut s'accumuler graduellement sur de longues périodes (comme les effets dus à la fatigue ou à la corrosion, par exemple). La Figure 1.1 illustre un processus de fissuration d'une des piles du pont Honoré-Mercier à Montréal, au Canada.

FIGURE 1.1 – Pont Honoré-Mercier à Montréal et détail des fissures sur l'une des piles.

La réponse à la deuxième question (« comment détecter de tels dysfonctionnements ? ») est généralement liée à l'évaluation de la performance des ouvrages. Les procédures d'inspection visuelle par des experts servent de base dans la plupart des cas pour la détection d'endommagements et la définition de la maintenance structurale sur une longue période. Toutefois, l'évaluation visuelle ne permet pas de quantifier la performance structurale, mais seulement d'en donner une mesure qualitative. De plus, l'accès aux différents éléments d'un ouvrage peut être difficile (voire impossible dans certains cas), ce qui limite l'utilisation de ce type de procédure. La nécessité d'approches alternatives pour répondre à cet enjeu a été une motivation forte pour

le développement des méthodes de surveillance de santé structurale. De façon générale, deux familles de techniques d'inspection d'ouvrages sont utilisées : les techniques non destructives et celles destructives. La première, qui est généralement la plus recherchée, consiste en un ensemble de méthodes permettant de caractériser l'état d'intégrité des structures ou des matériaux, sans les dégrader (contrairement aux méthodes destructives). Les exemples les plus connus sont les contrôles par ressuage, les émissions acoustiques, les rayons X, les ultrasons, les courants de corrosion, etc.

L'une des techniques d'évaluation non destructive la plus appropriée à la surveillance de santé des ouvrages est celle basée sur le suivi de ses caractéristiques vibratoires. Les méthodes vibratoires considèrent les réponses dynamiques de la structure comme des fonctions de ses propriétés mécaniques. Le principe est que le dommage modifie la rigidité, la masse et/ou les propriétés de dissipation d'énergie du système, ce qui altère la réponse dynamique de la structure. Plus spécifiquement, un processus d'endommagement produit des changements, notamment sur les paramètres modaux de la structure (i.e., les fréquences propres, déformées propres et coefficients d'amortissement). Les modifications de ces paramètres constituent donc intuitivement des indicateurs pertinents pour la détection d'un dommage structural. Par ailleurs, l'un des grands avantages des méthodes vibratoires repose sur la possibilité d'utiliser des vibrations ambiantes comme source d'excitation, ce qui est souvent observée dans la pratique.

1.2 État de l'art des méthodes de détection d'endommagements

La détection de défauts structuraux par méthodes vibratoires renvoie à l'étude des problèmes inverses. Cette démarche est basée sur une procédure d'identification de la localisation du dommage et la quantification de la perte de rigidité du système afin de l'ajuster à la réponse de la structure. La recherche des indicateurs est devenue populaire avec le développement des techniques d'évaluation dynamique et d'identification modale (Ewins 2000). En général, les effets d'un endommagement structural peuvent être classés comme linéaires ou non linéaires. Le premier cas est une situation dans laquelle la structure initialement élastique linéaire reste dans le domaine linéaire après être endommagée. Ceci est le cas des problèmes précoces des ouvrages. Dans le second cas, le comportement de la structure devient non linéaire après l'endommagement. Il est noté que la non linéarité intervient pour des degrés d'endommagement sévères comme par exemple la formation d'une fissure de fatigue sous des cycles de chargement/déchargement dans un environnement normal de vibration, la plastification de certains éléments suite à un choc, etc.

Les méthodes de détection d'endommagements linéaires sont les plus couramment utilisées. Elles sont classées en deux groupes : les méthodes basées sur modèle (MBM) et les méthodes non basées sur modèles (MNBM). Les méthodes du premier groupe utilisent des modèles numériques discretisés par des éléments finis (Zienkiewicz & Taylor 2000) pour développer des algorithmes de détection de défauts. Ces techniques sont majoritairement basées sur l'actualisation de modèles (notamment des matrices de rigidité et masse) par rapport à la structure réelle. Par exemple, il est possible de citer les algorithmes d'actualisation optimale des matrices (Smith & Beattie 1991), les techniques d'ajustement des paramètres modaux (Lim & Kashangaki 1994), les méthodes de sensibilité (Alvin 1997), les techniques basées sur des fonctions d'endommagements (Abdel Wahab et al. 1999, Teughels et al. 2002), etc.

Les MNBM sont essentiellement basées sur le changement des paramètres modaux et leurs dérivées (ou même sur la modification des matrices particulières, comme celles de flexibilité et de rigidité). L'un des avantages de ces techniques est qu'elles ne nécessitent pas la connaissance *a priori* de la localisation de l'endommagement.

Les méthodes de détection d'endommagements sont utilisées pour renseigner quatre niveaux

de connaissances (Rytter 1993) :

Niveau 1 : Y-a-t-il un dommage structural (existence) ?

Niveau 2 : Où le dommage se trouve-t-il (localisation) ?

Niveau 3 : Quelle est la sévérité du dommage (quantification) ?

Niveau 4 : Quelle est la durée de vie résiduelle de l'ouvrage (pronostic) ?

De nombreuses approches de type MNBM sont répertoriées et sont souvent basées sur le changement des fréquences propres, des coefficients d'amortissement et des déformées modales. Par ailleurs, des indicateurs dits « évolués » ont été proposés par plusieurs auteurs ces dernières années (Doebling et al. 1996). L'un des objectifs de ces indicateurs est bien souvent d'améliorer et raffiner l'identification de dommages par rapport à celle fournie par les méthodes basées sur le changement des seuls paramètres modaux. Une vaste étude comparative de ces techniques a été menée par Alvandi (2004). Dans les paragraphes suivants, une brève discussion sur les indicateurs existants est réalisée. Dans un premier temps, un résumé des techniques qui utilisent les caractéristiques modales pour la détection d'endommagement est proposé. Dans un deuxième temps, quelques indicateurs évolués existants dans la littérature sont présentés. Finalement, une discussion sur des nouveaux indicateurs, basés sur des techniques d'anormalité (*novelty detection*) est proposée.

1.2.1 Méthodes basées sur le changement des paramètres modaux

Au cours des premières études expérimentales, menées au XIX$^{\text{ème}}$ siècle, les équipements utilisés étaient relativement peu robustes et pratiquement seules les fréquences propres pouvaient être estimées avec un niveau de précision adéquat. De nombreuses études ont alors utilisé le changement de fréquence comme indicateur d'endommagement. Cette approche repose sur l'hypothèse que les fréquences propres sont des indicateurs sensibles de l'intégrité structurale. Il a en effet été observé que les changements des propriétés structurales causent des modifications sur les fréquences propres de la structure. Dans ce contexte, une analyse périodique des fréquences peut *a priori* servir de base pour une méthode de surveillance structurale. L'une des premières méthodes permettant de lier un endommagement structural à la variation des fréquences propres a été proposée par Cawley & Adams (1979). Dans cette approche, le dommage est considéré comme une réduction locale de la matrice de rigidité de la structure. Bien que l'utilisation du changement des fréquences propres ait permis de localiser l'endommagement, sa quantification n'était pas possible. Des améliorations ont été proposées par (Friswell et al. 1994, Williams et al. 1996) quelques années plus tard. Finalement, Messina et al. (1998) ont proposé une méthode de sensibilité intitulée MDLAC (*Multiple Damage Location Assurance Criterion*), permettant de détecter et quantifier des dommages multiples.

Dans toutes les études, que ce soit sur des ponts métalliques, mixtes ou en béton armé/précontraint, les fréquences propres décroissent en fonction d'un endommagement croissant. Ceci semble logique et intuitif puisque l'endommagement tend à réduire la rigidité de la structure. De nombreuses études traitent de la détection d'endommagements par l'analyse des décalages fréquentiels. Toutefois, la plupart de ces études montrent que l'analyse des décalages fréquentiels semble être d'une utilité pratique très limitée. Ainsi, Kato & Shimada (1986) ont réalisé des mesures dynamiques sous vibration ambiante d'un pont en béton précontraint. Ils ont montré que le niveau de chargement statique appliqué doit être proche du chargement ultime pour détecter une réduction des fréquences propres. Gudmundson (1982) a également mis en évidence que la diminution des fréquences propres est moins importante dans le cas d'une fissure de fatigue qui s'ouvre et qui ferme, que dans celui d'une fissure restant ouverte. Cela implique que la diminution des fréquences est affectée par des facteurs tels que les charges permanentes ou les contraintes résiduelles. En effet, les charges permanentes peuvent maintenir des fissures

ouvertes et accentuer le changement des fréquences. En utilisant des résultats numériques et expérimentaux sur une poutre, Fox (1992) a montré que les changements des fréquences propres sont des indicateurs peu sensibles dans le cas d'une poutre fissurée à la scie. Srinivasan & Kot (1992) sont arrivés aux mêmes conclusions dans le cadre d'étude sur des plaques endommagées.

La modification des paramètres modaux peut de plus ne pas être identique pour chaque mode. D'après Duggan et al. (1980), elle dépend en effet de la nature, de la localisation et de la sévérité de l'endommagement. La faible sensibilité des décalages fréquentiels à l'endommagement nécessite donc une instrumentation précise ou bien d'être en présence de niveaux de dommages importants. L'utilisation des décalages fréquentiels est surtout envisageable lors d'essais en environnement contrôlé, comme des essais réalisés en laboratoire, par exemple. Dans la plupart des cas, il est difficile d'établir une forte corrélation entre un décalage fréquentiel et l'occurrence d'un endommagement. En effet, aucune information spatiale n'est fournie par les fréquences propres. La seule exception réside dans le cas des fréquences très élevées qui sont souvent associées à des réponses locales. Cependant, les limitations dans l'excitation et l'extraction de ces modes rendent très difficile leur identification.

En ce qui concerne les coefficients d'amortissement, peu d'études ont été menées ces dernières années. Salane & Baldwin (1990) ont étudié l'influence d'endommagements sur les coefficients d'amortissement et ont montré que ces coefficients ne pouvaient pas constituer des indicateurs fiables car ils tendent subir des grandes oscillations empêchant une corrélation des effets d'endommagement avec leur variation. Toutefois, des études menées par Ndambi et al. (2000) et Kawiecki (2001) tendent confirmer l'utilité de se servir de l'amortissement comme indicateur de l'intégrité de la structure. Il est noté cependant que ce choix n'a fait l'objet que de peu de travaux de recherches jusqu'à présent.

À l'inverse, de nombreuses recherches ont été menées sur le développement d'indicateurs basés sur le changement des modes propres. La plupart des premières méthodes utilisaient une comparaison directe entre les déformées structurales avant et après l'occurrence d'un dommage. Allemang & Brown (1982) ont proposé un indice, nommé MAC (*Modal Assurance Criterion*) qui consiste à évaluer la corrélation entre deux quantités. Cet indice varie entre 0 et 1, où 0 exprime l'absence de corrélation entre les quantités comparées et 1, une corrélation parfaite. L'évaluation du MAC pour des modes endommagés et non endommagés fournit un moyen quantitatif d'évaluer la déviation de la corrélation des déformées propres endommagées et intactes. West (1984) a été le premier à utiliser les déformées propres pour la localisation d'endommagements sans recourir à un modèle aux éléments finis. Il a ainsi utilisé les coefficients MAC pour déterminer la corrélation entre les déformées propres avant et après endommagement des volets d'une navette spatiale.

Quelques années plus tard, Lieven & Ewins (1988) ont proposé l'indice COMAC (*COordinate Modal Assurance Criterion*). Cet indice est une variation du MAC dans le sens où la corrélation est maintenant calculée pour chaque coordonnée à travers les modes. Il donne ainsi une corrélation entre des déformées propres mesurées pour chaque degré de liberté. Ce coefficient est généralement employé pour identifier l'endroit où les déformées propres d'une série de mesures ne sont pas corrélées les unes avec les autres. Dans le cas d'une corrélation parfaite entre les déplacements d'une coordonnée, le coefficient COMAC est égal à 1. Un grand écart par rapport à 1 peut être interprété comme un endommagement dans la structure.

1.2.2 Méthodes basées sur le changement des indicateurs dits évolués

Les études menées sur l'utilisation directe des paramètres modaux ont mis en évidence la faible sensibilité des fréquences propres et une certaine potentialité des déformées propres à détecter et/ou localiser des endommagements. Pour accentuer la capacité des indicateurs basés sur des déformées modales à détecter des endommagements, diverses approches ont été proposées

(Farrar & Jauregui 1996).

Une autre façon d'utiliser les déformées propres pour obtenir des informations sur la source des changements vibratoires est d'étudier leurs dérivées. Il y a en effet une relation directe entre la courbure des déformées propres et les déformations de flexion. Pandey et al. (1991) ont introduit un nouvel indicateur basé sur la variation de la courbure des modes propres des structures saines et endommagées. Ils ont mis en évidence que les valeurs absolues du changement de la courbure des déformées propres peuvent constituer un bon indicateur d'endommagement pour un modèle aux éléments finis. En utilisant un schéma aux différences centrées, les valeurs de la courbure peuvent être calculées à partir des déformées propres. La valeur absolue de la différence de la courbure de déformées propres endommagées et non endommagées dans la région endommagée doit être maximale. Plus la réduction de la rigidité en flexion est grande (correspondant à un niveau d'endommagement élevé), plus le changement de courbure est important. Toutefois, dans la plupart des campagnes des mesures, il est généralement très difficile de pouvoir mesurer les degrés de liberté de rotation ou même plusieurs degrés de liberté de translation (dû aux limitations de l'équipement, du nombre de capteurs utilisés, etc.). L'utilisation de cette méthode reste donc peu pratique. De plus, Chance et al. (1994) ont montré que le calcul de la courbure obtenue à partir des déformées propres peut aboutir à des niveaux d'erreur très importants.

Une autre classe de méthodes de détection d'endommagements pour estimer des changements sur le comportement d'une structure est basée sur l'utilisation de la matrice de flexibilité (Pandey & Biswas 1994). Aktan et al. (1994) ont proposé l'usage de la flexibilité mesurée comme un indice de l'intégrité relative d'un pont. La méthode n'exige pas un modèle analytique de la structure et seules quelques fréquences et déformées propres, avant et après endommagement, sont nécessaires. La formulation de la matrice de flexibilité par cette méthode est approximative, car seulement un petit nombre de modes (généralement les modes propres associés aux basses fréquences) est mesuré. Étant donné d'une part que la présence d'endommagements diminue la rigidité de la structure, et d'autre part que la flexibilité est l'inverse de la rigidité, la présence d'endommagements doit donc augmenter la flexibilité de la structure.

En combinant certains aspects des méthodes de courbure des déformées propres et de flexibilité, Zhang & Atkan (1995) ont développé une autre méthode de détection d'endommagements. De la même manière que pour la méthode de la courbure, l'idée générale est que la perte de rigidité localisée cause l'augmentation de la courbure au même point. Le changement de courbure s'obtient alors par la flexibilité au lieu des déformées propres.

En 1993, Kim & Stubbs (1993) ont proposé une démarche originale pour la détection et la localisation d'endommagements reposant sur le changement de l'énergie de déformation avant et après endommagement. L'un des avantages de cette méthode est que le nombre de modes de vibration requis pour obtenir de bons résultats est relativement peu élevé. Ce changement, exprimé au moyen d'un ratio, permet de définir un indicateur qui indique l'absence de dommages lorsqu'il reste négatif et la présence d'endommagements, lorsqu'il est positif. Cette méthode est développée pour de faibles endommagements, car elle repose sur une linéarisation des énergies de déformation après endommagement. Plus récemment, Cury, Borges & Barbosa (2010) ont proposé une approche hybride permettant de localiser des endommagements dans un premier temps et ensuite de les quantifier de manière indépendante.

1.2.3 Méthodes basées sur les indicateurs d'anormalité

Les méthodes présentées dans les paragraphes précédents reposent sur une analyse essentiellement fréquentielle des signaux pour en extraire une information compressée au travers de quelques caractéristiques modales ou d'indicateurs évolués construits sur ces caractéristiques. Ces indicateurs ont démontré leur efficacité, mais ont soulevé le problème de leur sensibilité, de

la nécessité de disposer d'un état de référence, et de leur fiabilité en termes de la probabilité de détection et de fausses alarmes. Un aspect également important à considérer est que ces méthodes supposent un comportement mécanique linéaire après endommagement, ce qui n'est pas toujours le cas. Par ailleurs, le processus d'identification modale est avant tout une procédure de filtrage, ce qui peut conduire à une perte d'informations essentielles sur l'état d'une structure. Parallèlement, peu de travaux ont porté sur l'analyse directe de la réponse d'une structure au travers de sa signature lors d'excitations extérieures. En effet, les signaux issus des mesures dynamiques sont rarement utilisés en raison du manque d'outils appropriés pour leur manipulation et de la difficulté à les exploiter directement.

Afin d'apporter des éléments de réponse à ces besoins, des nouvelles démarches basées sur l'analyse statistique et la reconnaissance de formes ont été introduites et des techniques originales pour la détection de dommages ont été proposées. L'une de ces techniques est basée sur le concept de fusion des données. Cette manipulation de données s'est développée lors de procédures militaires aux États-Unis dans les années 50. L'objectif était de déterminer les situations des champs de bataille et d'évaluer les menaces à partir d'informations générées par de nombreuses sources différentes (Klein 1999). Cette démarche s'est rapidement propagée à d'autres domaines comme l'économie, la biologie, la géographie, etc. La fusion de données consiste essentiellement à confronter et intégrer des informations multiples dans le but de réduire l'incertitude sur l'information résultante. Dans le domaine du génie civil, et plus particulièrement celui de la détection d'endommagements, relativement peu d'études ont été menées sur le problème de la gestion de nombreuses données acquises. Guo (2006) et Minor et al. (2007) ont proposé différentes méthodologies pour assembler les informations obtenues par plusieurs capteurs accélérométriques. Ils ont signalé que la robustesse de leurs approches pour identifier des dommages structuraux dépend fortement de la taille de la base de données à traiter. Chun et al. (2005) ont également mis en évidence le problème de gestion de nombreuses données acquises dans une campagne de mesures. Su et al. (2009) ont proposé plusieurs schémas de fusion de données dans lesquels certains capteurs sont préalablement choisis et seules leurs mesures sont exploitées. Les résultats obtenus ont cependant montré que la perte d'information dans ce cas peut conduire à des fausses alarmes dans la détection d'endommagements. Des approches basées sur la méthode des sous-espaces ont été également proposées pour la détection de comportements structuraux anormaux (Basseville et al. 2006). Le principe est d'estimer des résidus issus des comparatifs entre les réponses dynamiques de l'ouvrage sain et endommagé. De manière générale, la grandeur des résidus permettrait d'avoir une idée sur localisation et quantification d'endommagements (Mevel et al. 1999, Bodeux & Golinval 2003).

Une démarche appliquée à la surveillance de santé structurale, ayant récemment suscité un grand intérêt, est celle de la reconnaissance de formes (aussi connue sous le nom de *pattern recognition*). La reconnaissance de formes, appliquée au domaine du génie civil, consiste en un ensemble de méthodes permettant d'identifier des caractéristiques intrinsèques des ouvrages (appelées *features*) à partir de données brutes. Ces caractéristiques permettent de déterminer et discriminer des comportements structuraux distincts ou encore différents états d'endommagements de la structure. Il existe une variété de méthodes employées pour la détection d'endommagements décrites dans la littérature. Ces méthodes sont souvent séparées en deux familles : les méthodes dites non supervisées (telles que les méthodes de partitionnement de données, les tests d'hypothèse, etc.) et les méthodes dites supervisées (les réseaux de neurones, les machines à vecteur support, etc.). Les méthodes non supervisées consistent à regrouper un ensemble d'observations inconnues (non étiquetées) dans des groupes différents. Le but est d'identifier des groupes tels que les observations les plus similaires se retrouvent dans un groupe commun et les plus différentes soient séparées dans des ensembles distincts. Dans le cas des méthodes supervisées, une collection d'étiquettes (connues *a priori*) est fournie comme donnée d'entrée pour la détermination des différentes classes. Le problème consiste à donner une étiquette à une nouvelle observation

sachant que les groupes sont déjà définis. Quelques auteurs ont proposé des approches basées sur ces deux types de méthodes. Trendafilova & Heylen (2003) et Iwasaki et al. (2004) ont utilisé des algorithmes de classification non supervisée pour la détection de dommages sur une poutre encastrée. Pour cela, ils ont employé différentes réponses fréquentielles obtenues à partir d'essais d'endommagement contrôlés. Les résultats obtenus ont montré que l'approche considérée permettait de classer correctement les différents états de la poutre. Haritos & Owen (2004) ont comparé plusieurs techniques statistiques telles que l'analyse en composantes principales et l'utilisation de tests d'hypothèse sur des essais dynamiques expérimentaux réalisés sur une poutre métallique. Les niveaux d'endommagements considérés sont des entailles sur une région donnée de la poutre. Ils ont conclu que les indicateurs n'étaient pas suffisamment sensibles pour détecter de petites entailles. Reda Taha & Lucero (2005) ont utilisé une démarche basée sur la logique floue (*fuzzy logic*) pour la détection de défauts sur un modèle numérique d'un pont en béton précontraint. La technique s'est révélée être efficace pour distinguer un état sain d'un état très dégradé. Dans les cas d'un état de dégradation modéré, l'approche proposée n'était pas capable de le distinguer d'un état sain. Par ailleurs, une étude basée sur le partitionnement « flou » des données (e.g., des algorithmes de type *c-means*) a été proposée par Da Silva et al. (2008). Zhang (2007) a utilisé des indices statistiques basés sur la réponse fréquentielle d'un modèle numérique d'un pont métallique. Il a montré que, pour de faibles niveaux d'endommagement, ces indices peuvent ne pas révéler la présence de défauts structuraux.

Par ailleurs, des études sur les méthodes de classification supervisées ont été menées. La plupart de ces méthodes ne considèrent que les paramètres modaux lors de la procédure de classification. Les réponses vibratoires (signaux) de la structure sont rarement utilisées. Plusieurs auteurs signalent que les mesures dynamiques sont peu utiles pour la détermination des comportements structuraux, du fait de la présence de bruit ou de la difficulté à les manipuler proprement (Kim & Lee 2000, Yan et al. 2007). En effet, Yeung & Smith (2005) ont proposé un algorithme d'apprentissage basé sur la signature d'un modèle numérique d'un pont réel, simulant plusieurs niveaux d'endommagement et de bruit. Il a été observé que pour de faibles niveaux de bruit, l'algorithme peut identifier correctement les scénarios d'endommagement imposés. En revanche, lorsque la quantité de bruit ajoutée aux signaux augmente, la performance de la classification n'est plus satisfaisante. Fang et al. (2005) a étudié d'autres architectures des réseaux de neurones appliqués à la détection de dommages. Dans ce cas, seules les fréquences propres sont utilisées. Les résultats obtenus ont montré que ce paramètre modal peut être utile pour l'identification de modifications structurales. Une étude similaire a été menée par Wen et al. (2005) à partir du couplage de réseaux neuronaux avec la théorie de la logique floue. D'autres méthodes de classification supervisée souvent utilisées sont celles basées sur les machines à vecteur support. Toutefois, peu de travaux ont été réalisés dans le domaine de la surveillance de santé des ouvrages. Hagiwara & Mita (2003) ont utilisé cette méthode pour la quantification d'endommagements imposés à une poutre. Pour cela, seules les fréquences propres ont été considérées pour réaliser la classification des états structuraux. Par ailleurs, Mita & Taniguchi (2004) ont étendu cette approche pour une application sur le contrôle actif d'un pont en béton armé. Ils ont conclu que les conditions environnementales expérimentées durant les essais dynamiques ont empêché la caractérisation des groupes de comportement structuraux différents. Un couplage avec des ondelettes a été envisagé par He & Yan (2005), mais son application est restée limitée à des simulations numériques.

1.3 Étude des effets des variations thermiques

De nombreux facteurs extérieurs peuvent affecter l'identification d'endommagements d'un ouvrage. Les références bibliographiques sur ce sujet sont peu nombreuses et parfois peu complètes.

La plupart des techniques de détection d'endommagements existantes négligent les impacts souvent importants causés par les changements environnementaux sur les ouvrages. Des modifications de la température, de l'humidité, du vent, etc. peuvent produire une modification significative des paramètres modaux des structures Guo et al. (2007). En effet, il existe des cas où cette variation peut être beaucoup plus importante que celle due à un endommagement (Sohn et al. 1999). Les vibrations causées par le vent, par exemple, peuvent modifier le comportement dynamique d'un ouvrage en altérant ses caractéristiques d'amortissement (Fujino & Yoshida 2002). Toutefois, l'un des effets environnementaux qui a le plus suscité de débats dans la littérature est celui causé par les variations thermiques.

Les études existantes ont montré que des variations des fréquences propres de l'ordre de 4 à 5% pour un changement de température de 10 °C sont envisageables (Wahebm & De Roeck 1997). Farrar et al. (1995) ont pu constater une variation diurne de 5% en moyenne sur une période de 9 mois de mesures. Par ailleurs, Kim et al. (2003) et Mita & Taniguchi (2004) ont constaté que la variation des paramètres modaux causée par des changements de température peut masquer un processus d'endommagement structural.

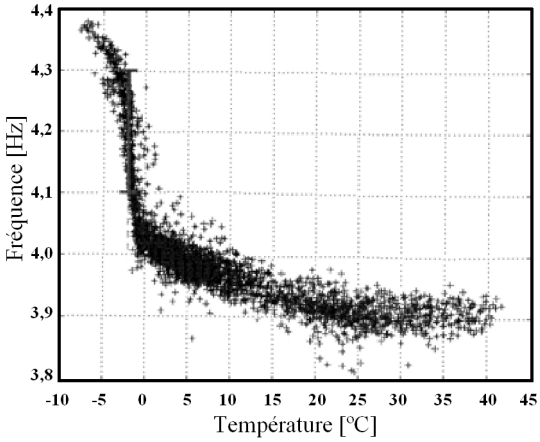
Une correction dans l'estimation des fréquences mesurées est donc nécessaire. Pour cela, quelques études ont été menées en considérant différentes lois de correction. Rohrmann et al. (2000), Leconte et al. (2007) ont proposé un modèle de proportionnalité entre les variations thermiques et les modifications fréquentielles. Peeters & De Roeck (2001) ont proposé des techniques d'autorégression (telles que des modèles ARX, par exemple) et Sohn et al. (2002) ont utilisé des méthodes basées sur les réseaux de neurones pour déterminer une corrélation entre les températures et les fréquences propres. Roberts (1995), mais également Woods (1992) ont suggéré, à la suite de l'étude de plusieurs ouvrages sur des périodes de plus de six mois, une relation linéaire entre la température ambiante et les fréquences propres des ouvrages. Alampalli (1995) a mis en évidence que les fréquences propres d'un pont peuvent changer de 40 à 50% dû au gel au niveau de ses appuis, alors que la variation de la fréquence causée par un endommagement artificiel (entaille sur la semelle d'une poutre) n'est que de 3 à 8%. Il est cependant noté que le pont considéré dans cette étude était relativement court avec une longueur de 6,76 m et une largeur de 5,36 m. Roberts & Pearson (1998) ont réalisé une étude de l'influence thermique sur un pont de 840 m de long, comportant neuf travées. Ils ont mis en évidence un changement fréquentiel de 3 à 4% au cours d'une année. Yan et al. (2005) ont proposé un modèle de correction basé sur l'analyse en composantes principales. En utilisant une modélisation numérique d'un pont, ils ont simulé plusieurs niveaux d'endommagement et ont imposé un gradient thermique sur la structure. Ils ont pu constater que cette correction est satisfaisante pour les cas des variations linéaires et quasi non linéaires. Les modèles suggérés ont montré leur efficacité, mais peu d'attention a été donnée aux applications pour la détection des modifications structurales.

De façon générale, l'établissement d'une loi de correction ne peut se faire qu'au travers d'un suivi simultané de la température et des fréquences propres. Il est cependant utile de comprendre *a priori* pour quelles raisons les paramètres modaux sont sensibles à la température. Le type de construction et les matériaux utilisés ont un effet sur le comportement dynamique d'une structure. Chaque matériau a ainsi un coefficient thermique d'expansion, et sera donc affecté différemment selon les variations de la température extérieure.

Des études expérimentales ont été effectuées par l'EMPA (Laboratoire Fédéral d'Essai des Matériaux et de Recherche, (EMPA 1999)) et l'Université Catholique de Louvain (Peeters et al. 2001) sur le pont Z24 (Figure 1.2(a)). Ce pont est un ouvrage de 63,3 m en béton précontraint de l'autoroute A1 Berne-Zurich reliant Koppigen et Utzenstorf. Les études menées ont apporté quelques renseignements intéressants et complémentaires sur la sensibilité des paramètres modaux aux variations climatiques. En particulier, il a été noté une variation de près de 10% sur la première fréquence (Figure 1.2(b)) pour une variation thermique allant de -5 °C à 35 °C. En

contrepartie, pour des essais réalisés entre 15 °C et 30 °C, cette variation n'était que de 2,5%.

(a) Vue du pont



(b) Évolution de la première fréquence propre avec la température

FIGURE 1.2 – Étude des effets thermiques sur le pont Z24.

Des résultats analogues ont été obtenus sur le pont B15, ouvrage précontraint en béton qui franchit l'autoroute E19 et relie Bruxelles et Anvers. Composé de trois travées pour une longueur totale de 124,6 m, l'ouvrage est supporté par un système d'articulation qui permet des mouvements latéraux au niveau des deux culées. Sur l'analyse de deux séries de mesure en hiver et au printemps, une diminution des fréquences avec la température (de 0 °C à 15 °C) a été constatée. Les coefficients d'amortissement ne suivent pas d'évolution précise, ceci pouvant être expliqué par des erreurs de mesure, de traitement des signaux et par la dépendance de ce coefficient au type d'excitation et à l'amplitude de la vibration. Par ailleurs, des études menées sur la poutre VIPP (Viaducs à travées Indépendantes à Poutres préfabriquées en béton Précontraint par post-tension) de Merlebach ont également montré une diminution sensible des deux premières fréquences propres avec l'augmentation de la température (Crémona 2007, Leconte et al. 2007).

Cette diminution de la fréquence avec la température rencontrée dans les ponts Z24 et B15 n'est pas une règle en soi. Ainsi, à titre de contre-exemple, Farrar et al. (1995) ont mis en évidence, sur l'étude du pont Alamosa Canyon au Nouveau Mexique (USA), une augmentation des fréquences avec la température.

1.4 Objectifs

L'objectif majeur de cette thèse est d'apporter des nouveaux éléments de réponse à la problématique de détection d'endommagements sur les ouvrages d'art. Un premier enjeu est de proposer des nouvelles approches permettant d'identifier un dommage structural non seulement à partir des changements des paramètres modaux, mais également directement à partir de la réponse dynamique de l'ouvrage. Comme mentionné précédemment, les mesures sont peu (voire jamais) exploitées faute d'outils efficaces ou des difficultés rencontrées pour gérer des quantités importantes de données brutes. Un deuxième enjeu est de prendre en compte l'influence des effets thermiques sur le comportement dynamique des ouvrages. De nombreuses recherches ont indiqué que ces effets ne sont pas négligeables et qu'ils doivent être considérés dans les études de détection d'endommagements.

Pour répondre au premier enjeu, une méthodologie récemment développée, basée sur les

concepts de fusion de données et du *Data Mining* est proposée. Cette méthodologie, appelée « Analyse de Données Symboliques (ADS) » permet de manipuler des grandes bases de données alliant un stockage plus compact sans perdre la généralité de l'information originale. Actuellement, l'ADS n'est souvent employée que dans le cadre d'applications économiques, telles que l'analyse de comportements de consommateurs, ventes croisées, etc. Dans cette thèse, l'ADS est utilisée de manière originale pour la détection d'endommagements structuraux. Pour cela, elle est couplée à des méthodes de classification non supervisée (telles que les méthodes de partitionnement de données) et supervisée (arbres de décision Bayésiens, réseaux neuronaux et machines à vecteur support).

Il est important de remarquer que la démarche proposée permet d'utiliser des mesures vibratoires (signaux) et des paramètres modaux pour la discrimination de groupes de comportement structural distincts. En effet, les méthodes décrites précédemment, et qui sont basées sur des indicateurs dits « classiques », ne permettent pas une analyse générale sur un ensemble d'états structuraux. En d'autres termes, il n'est pas possible de distinguer aisément des groupes de comportements distincts. Le chapitre suivant a pour objectif d'introduire l'ADS et montrer comment les données sont manipulées dans cette thèse.

En ce qui concerne le second enjeu de cette thèse, la démarche proposée consiste à développer des nouvelles approches pour étudier les effets thermiques sur les paramètres modaux. L'objectif est de pouvoir éliminer ces effets et être capable de caractériser une variation « normale » des fréquences propres et déformées modales d'une variation « anormale ». Pour cela, des techniques de régression linéaire et non linéaire sont proposées. Il est noté que peu d'études ont été menées avec des modèles de régression non linéaire. Ces techniques sont aussi couplées aux méthodes de classification de données. L'idée principale est d'observer comment la variation thermique peut influencer la détermination des groupes de comportements structuraux. Par ailleurs, une méthodologie basée sur l'Analyse en Composantes Principales (ACP) est proposée, le but étant de discriminer des variations fréquentielles liées aux changements thermiques de ceux dus à la présence d'endommagements structuraux.

Chapitre 2

Analyse de Données Symboliques

Ce chapitre a pour objectif d'introduire les concepts majeurs de l'Analyse de Données Symboliques (ADS). Pour cela, une introduction générale de l'ADS est proposée en mettant en évidence la manière dont les données sont manipulées. Dans un premier temps, des aspects généraux sur le *Data Mining* et la création de l'ADS sont présentés. Dans un second temps, des exemples illustrant la transformation de données classiques en données symboliques sont introduits.

2.1 Introduction

Le *Data Mining* ou « exploitation de données » utilise des outils plus ou moins sophistiqués pour découvrir des modèles et des relations précédemment inconnus (ou « cachés ») dans des grands ensembles de données. Ces outils peuvent inclure des modèles statistiques, des algorithmes mathématiques, des méthodes d'apprentissage, etc. Le *Data Mining* ne se limite pas à un assemblage et un stockage de données. Il inclut également une analyse critique de ces données dans un moment présent, mais aussi une prévision de tendances pour l'avenir.

La démarche du *Data Mining* peut être exécutée sur des données représentées sous formes quantitatives, textuelles ou de type multimédia (comme des images et des sons, par exemple). Les techniques d'extraction de connaissances peuvent nécessiter de nombreux paramètres pour caractériser les données. Elles incluent l'association (modèles où un événement est lié à un autre : acheter un stylo et du papier, par exemple), l'ordre ou l'analyse de conséquences (modèles où un événement mène à un autre : naissance d'un enfant et l'achat d'un berceau), la classification (identification de nouveaux modèles), le clustering (découverte et visualisation de groupes de données : préférences de marques de vêtements) et la prédiction (découverte de modèles à partir desquels des prévisions raisonnables concernant des futures activités peuvent être proposées). L'un des grands apports du *Data Mining* est lié à l'utilisation des approches dites de « découverte », pour lesquelles des algorithmes peuvent être employés pour examiner simultanément plusieurs relations multidimensionnelles de données (Hastie et al. 2009).

Le *Data Mining* est de plus en plus utilisé aujourd'hui tant dans les secteurs privés que publics. Il est souvent considéré comme un outil permettant la gestion d'information sur des clients et la réduction des probabilités de fraude et de perte, par exemple. Bien que les algorithmes d'exploitation de données peuvent être des outils très puissants, ils ne se suffisent pas à eux-mêmes. En ce sens, l'analyse critique de l'utilisateur reste toujours nécessaire.

Les techniques statistiques utilisées pour l'analyse de données ont généralement été conçues pour des cas d'application relativement simples. L'unité de l'analyse est souvent un individu (par

exemple, une personne, un objet, un endroit, etc.) décrit par un type bien défini de variables (qualitatives ou quantitatives), associées chacune à une valeur simple. Lorsque les unités d'étude passent du niveau « individu » au niveau « ensembles d'individus », leur description devient plus délicate. En cela, une nouvelle représentation de données a été conçue. Ces données, contenant une information plus générale mais également plus complexe, sont nommées « données symboliques ». Pour mieux les appréhender et en extraire de nouvelles connaissances, l'ADS a été créée.

2.2 Des données classiques aux données symboliques

Les récents progrès techniques et technologiques rendent désormais possible l'acquisition continue d'informations diverses et le stockage de bases de données volumineuses. Cependant, en dépit de la capacité de traitement courante des ordinateurs, les ressources informatiques nécessaires pour manipuler et analyser des grands ensembles de données sont généralement limitées. Par ailleurs, les méthodes traditionnelles d'analyse de données ont été développées à une époque où la quantité d'information disponible était beaucoup plus limitée qu'elle ne l'est actuellement.

Plusieurs études liées à l'analyse multivariée, à la reconnaissance des motifs, aux bases de données et à l'intelligence artificielle ont été menées depuis des années. Les méthodologies en cours de développement généralisent les techniques d'analyse des données classiques, comme par exemple la visualisation, les techniques factorielles, les arbres de décision, la discrimination, la régression, la classification ou encore les méthodes de partitionnement de données.

Contrairement à l'analyse classique qui se concentre sur l'étude des *individus*, l'analyse symbolique manipule des *concepts*, qui représentent une famille plus riche et moins spécifique de données. Les concepts, considérés comme des nouvelles unités statistiques, sont décrits par des données plus complexes que celles habituellement rencontrées en analyse statistique. Elles sont dites « symboliques », car en exprimant la variation interne inhérente aux concepts et en tenant compte des spécificités de la base de données, elles ne peuvent pas être manipulées uniquement par de simples descriptions numériques ou qualitatives. En cela, elles nécessitent une caractérisation plus riche mais aussi plus compacte. Dans ce contexte, l'extension des méthodes classiques à de telles données représente un enjeu majeur. Le Tableau 2.1 donne un exemple simple d'une comparaison des unités d'étude d'une analyse classique et d'une analyse symbolique.

Données classiques : *individus*	Données symboliques : *concepts*
Joueurs	Équipes
Orchidées, roses	Fleurs
Pigeons, canards	Oiseaux

TABLEAU 2.1 – Exemples d'unités d'étude pour une analyse classique et une analyse symbolique.

La première étape de l'ADS est d'assembler et de décrire les concepts. Ceci peut être réalisé par l'utilisation de variables qui permettent non seulement de bien représenter mais aussi de caractériser ces concepts. La seconde étape consiste à analyser le tableau de données symboliques issu de la première étape. Pour cela, il est nécessaire d'élargir les notions du *Data Mining* et de la statistique classique au cas où les unités sont des concepts décrits par des variables symboliques. Ce type d'analyse est connu sous le nom de *Knowledge Mining* (Bock & Diday 2000). Ainsi, dans cette seconde étape, l'ADS n'a pas pour but d'analyser une seule sorte de données complexes. Au contraire, le but est d'analyser des concepts décrivant la variation interne de leurs instances, décrites par des données de toutes sortes qui peuvent être plus ou moins complexes (Billard & Diday 2006). Les tableaux de données symboliques peuvent comporter des objets de différentes natures (par exemple, un ensemble de plusieurs valeurs), pouvant être pondérés et

liés entre eux par des règles et/ou des taxonomies. Des exemples de ces représentations sont énumérés ci-dessous :

- une seule valeur quantitative. Par exemple, si « hauteur » est la variable et z est un individu : hauteur(z) = 3,5 m.
- une seule valeur catégorique : ville(z) = Paris.
- des données catégoriques multi-valeurs : prix(z) = {haut, moyen, bas} indiquant que le prix général pour un produit z peut être haut, moyen ou bas.
- des données de type intervalle : poids(z) = [20 ; 180] signifie que le poids de z varie dans l'intervalle [20 ; 180].
- des données de type modal multi-valeurs (histogramme) : hauteur (z) = {[0 ; 1,20] (0,225) ; [1,20 ; 1,50] (0,321) ; [1,50 ; 1,80] (0,335) ; [1,80 ; 2,10] (0,119)} indiquant que 22,5% de la population z a sa hauteur comprise entre 0 et 1,20 m, 32,1% entre 1,20 m et 1,50 m, 33,5% entre 1,50 m et 1,80 m, et 11,9% entre 1,80 m et 2,10 m.

Ce type de représentation permet d'exprimer la variabilité qui entache chacun des concepts. En étendant l'analyse de données traditionnelles aux tableaux de données symboliques, l'ADS permet d'extraire des descriptions symboliques munies d'un opérateur de comparaison et d'une fonction de reconnaissance formant un triplet appelé « objet symbolique ». Ces objets symboliques décrivent les concepts par leurs propriétés communes et leur extension dans l'ensemble des individus qu'ils représentent (Billard & Diday 2006).

Les principaux avantages des descriptions symboliques peuvent se résumer comme suit :

- elles fournissent un résumé de la base de données plus générale que les données agrégées habituelles, tenant compte de la variation interne et permettant de construire un nouveau tableau de données plus riche en information pour lequel une ADS peut être appliquée ;
- afin de modéliser des concepts, elles peuvent aisément exprimer des propriétés des variables provenant de plusieurs tableaux associés à différentes populations. Par exemple, pour construire une description symbolique associée à une ville, il est possible d'utiliser des propriétés issues d'une relation décrivant les habitants de chaque ville et une autre décrivant les écoles de chaque ville ;
- plutôt que de fusionner plusieurs bases pour étudier ensuite la base synthétique obtenue, il peut être plus pertinent d'extraire d'abord des descriptions symboliques de chaque base puis d'étudier l'ensemble des descriptions symboliques ainsi obtenues.

Afin d'illustrer la différence entre un tableau de données classiques et un tableau de données symboliques, l'exemple d'un groupe d'oiseaux comprenant 300 hirondelles, 100 pingouins et 100 autruches est considéré. Le Tableau 2.2 représente un tableau classique où chaque ligne décrit un oiseau (individu) d'après trois variables (Diday & Noirhomme-Fraiture 2007) : espèce, capacité de voler et poids.

Oiseau	Espèce	Vol	Poids (kg)
1	Pingouin	Non	35
2	Hirondelle	Oui	0.05
⋮	⋮	⋮	⋮
500	Autruche	Non	100

TABLEAU 2.2 – Description classique de 500 oiseaux

Une analyse plus générale peut être menée en considérant les espèces d'oiseaux comme unités d'étude (les espèces d'oiseaux sont alors les concepts). Le Tableau 2.3 classifie les espèces par rapport aux caractéristiques décrites précédemment (capacité de voler et poids).

Espèce	Nombre d'individus	Vol	Poids (kg)
Hirondelle	300	Oui	[0,01 ; 0,08]
Pingouin	100	Non	[30 ; 100]
Autruche	100	Non	[80 ; 160]

TABLEAU 2.3 – Description symbolique de 500 oiseaux

À la question « Quel est le pourcentage d'oiseaux qui volent ? », deux réponses sont possibles. La première, si la description classique est prise en compte et les unités d'étude sont représentées par les individus, est la suivante « $300/500 = 60\%$ » (vu que seulement les 300 hirondelles volent). Si la description symbolique est considérée et les espèces (concepts) sont utilisées comme unités d'étude, la réponse est alors « $1/3 = 33,33\%$ » (une espèce sur trois).

Par ailleurs, il est important d'observer que dans le Tableau 2.3 les valeurs de la variable « Poids » ne sont pas des scalaires, mais des intervalles. Ce type de notation permet de construire des tableaux compactant l'information et plus généraux que des tableaux classiques, étant donné qu'ils permettent d'analyser, d'une certaine façon, la dispersion des variables. L'ADS permet d'organiser (voire même de compresser) les données originales de manière contrôlée afin d'être capable de les mieux manipuler. L'enjeu est donc de savoir jusqu'à quel niveau cette compression peut être réalisée sans dénaturer ou déformer l'information originale. En effet, chaque application entraînera des descriptions et transformations de données spécifiques et c'est à l'utilisateur de juger celle la plus adéquate.

L'étude de la surveillance dynamique d'une structure est maintenant considérée. Un essai hypothétique $\mathbf{E}$ contenant 10000 valeurs mesurées par 10 capteurs accélérométriques est introduit et illustré à la Figure 2.1.

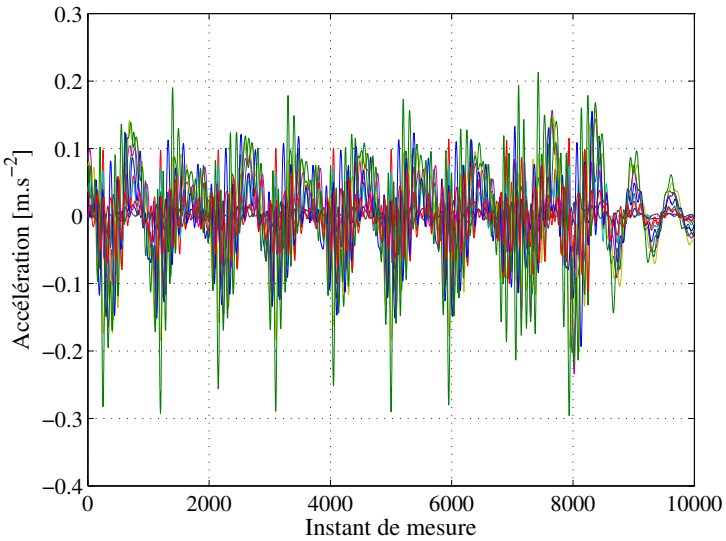
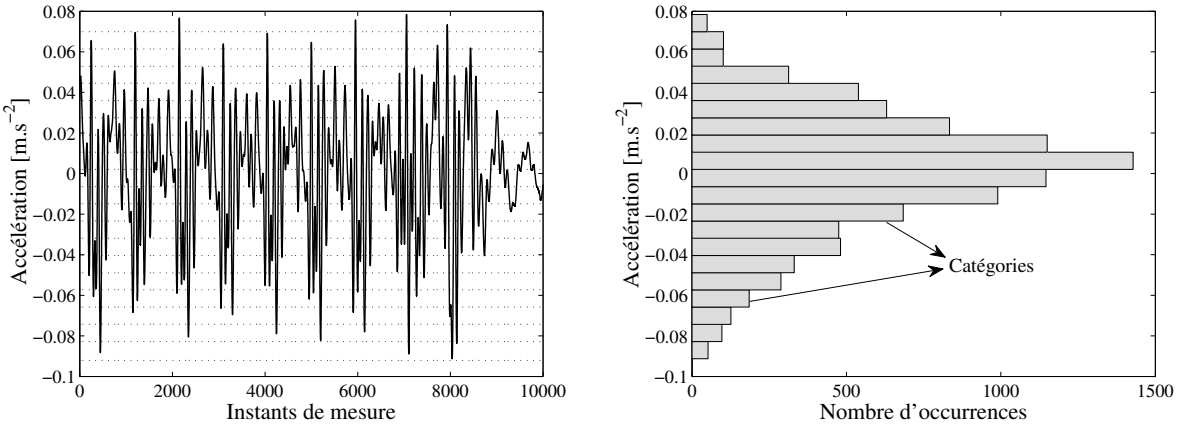


FIGURE 2.1 – Exemple d'un essai dynamique.

Cet essai peut être représenté de manière symbolique. Dans cet exemple, les valeurs mesurées sont les **individus** et les capteurs sont les **variables** qui décrivent les essais dynamiques (ici, les **concepts**). Le signal mesuré par le capteur 7 est maintenant considéré. La Figure 2.2 illustre le passage de la représentation classique à la représentation symbolique. Dans cette transformation, chaque valeur mesurée est projetée sur l'axe des ordonnées. Un histogramme est alors construit, en tenant compte de la proportion de valeurs comprises entre les bornes de chaque division. Ces divisions sont nommées *catégories*. Le nombre de catégories est défini *a priori* selon le type

d'étude menée et détermine la résolution ou le degré de raffinement pour la description de chaque concept.



(a) Signal enregistré par le capteur 7.

(b) Histogramme à 20 catégories.

FIGURE 2.2 – Représentation symbolique d'un signal classique par un histogramme de valeurs.

Une autre manière de représenter symboliquement cet essai est de le transformer en intervalles absolus ou interquartiles. Le premier prend en compte les valeurs minimales et maximales absolues. Le deuxième considère les valeurs correspondantes à 25% et à 75% de la distribution [1]. Le Tableau 2.4 illustre une description en données symboliques pour l'essai **E** à partir des intervalles interquartiles et des histogrammes à 20 catégories. Pour ce dernier, les valeurs entre parenthèses indiquent les proportions calculées pour chacune des catégories définies initialement.

Capteur	Intervalle interquartile [m·s^{-2}]	Histogramme
1	[-0,0142 ; 0,0176]	[1(0,002), ..., 10(0,31), ..., 20(0,001)]
2	[-0,0517 ; 0,0675]	[1(0,001), ..., 10(0,28), ..., 20(0,005)]
⋮	⋮	⋮
10	[-0,0051 ; 0,0056]	[1(0,003), ..., 10(0,25), ..., 20(0,004)]

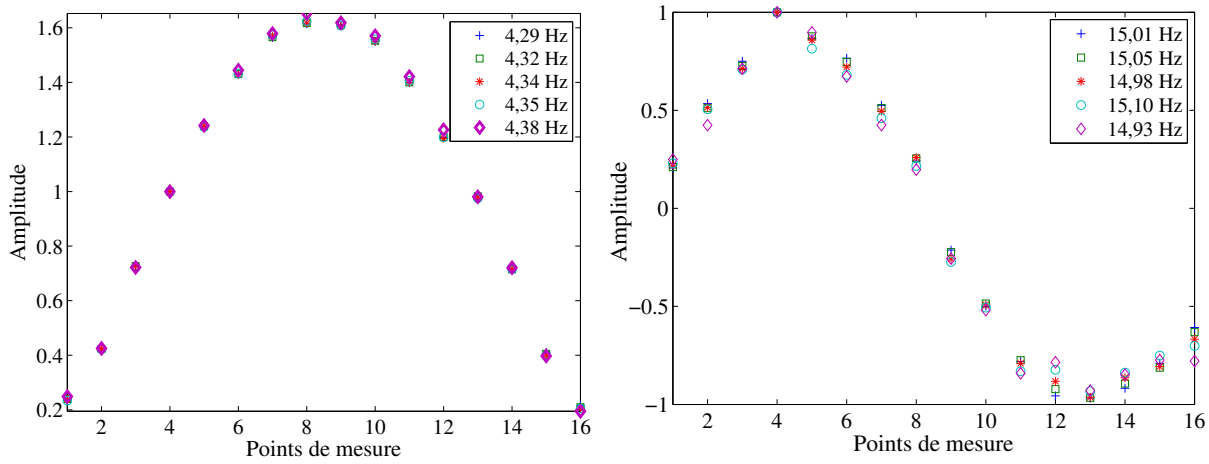
TABLEAU 2.4 – Description symbolique de l'essai **E** représenté par des capteurs.

Il existe d'autres manières de représenter des essais issus des campagnes de surveillance dynamique. Dans le cadre d'une identification modale, les essais peuvent être décrits par les paramètres modaux identifiés, comme les fréquences propres et les modes propres. Dans ce cas, les essais dynamiques sont toujours les **concepts**, mais les **variables** sont maintenant représentées soit par les fréquences propres, soit par les modes propres et non plus par les capteurs.

L'exemple de l'identification modale d'une poutre métallique sur deux appuis est maintenant considéré. Les deux premières fréquences propres et déformées modales sont calculées. Dans un certain nombre de cas, il ne suffit pas de trouver une valeur numérique convenable pour les paramètres modaux, mais il faut estimer également leur précision et les incertitudes liées à l'identification modale. En particulier, lorsque le nombre d'essais est insuffisant, l'estimation ponctuelle présente de grandes fluctuations et des erreurs importantes peuvent apparaître. Afin de traiter ce problème, il est possible de générer plusieurs estimations et d'en produire des statistiques à partir d'échantillons prélevés des réponses dynamiques. Ces estimations sont généralement appelées

1. Par exemple, pour une distribution continue de valeurs réelles entre 1 et 10, l'intervalle interquartile est [2,5 ; 7,5].

réalisations. La Figure 2.3 illustre les fréquences et déformées propres des cinq réalisations évaluées pour les deux premières paires de paramètres modaux identifiés. Cette procédure produit donc une distribution de valeurs pour chaque paramètre modal. À partir de cette distribution, les fréquences propres et modes propres sont transformés en données symboliques, comme détaillé dans les Tableaux 2.5 et 2.6, respectivement.



(a) Premier mode propre

(b) Deuxième mode propre

FIGURE 2.3 – Fréquences et modes propres identifiés pour 5 réalisations d'une identification modale.

Fréquence	Intervalle interquartile [Hz]	Histogramme
1	[4,32 ; 4,35]	[1(0,01), ..., 10(0,04), ..., 20(0,001)]
2	[14,98 ; 15,05]	[1(0,02), ..., 10(0,58), ..., 20(0,005)]

TABLEAU 2.5 – Description symbolique de l'essai **E** représenté par les fréquences propres.

Mode	Intervalle interquartile	Histogramme
1	[1,01 ; 1,22]	[1(0), ..., 10(0,40), ..., 20(0)]
2	[0,87 ; -0,88]	[1(0,2), ..., 10(0,12), ..., 20(0,04)]

TABLEAU 2.6 – Description symbolique de l'essai **E** représenté par les modes propres.

À travers ces exemples, il est possible de constater que plusieurs représentations symboliques sont possibles. Il faut noter cependant que chaque application entraîne l'utilisation d'un type de données symboliques approprié. Plusieurs simulations ont été réalisées préalablement en utilisant diverses représentations symboliques telles que des intervalles absolus et des histogrammes à 10 et à 40 catégories, par exemple. Or, dans le cadre de cette thèse, seuls des intervalles interquartiles et des histogrammes à 20 catégories sont utilisés. Il a été montré que de telles représentations sont généralement appropriées pour des applications étudiées (Cury, Crémona & Diday 2010). En réalité, il est important de signaler qu'il faut toujours trouver un compromis entre la représentation de la donnée symbolique avec la donnée classique. En d'autres termes, il est nécessaire de savoir compresser la donnée classique sans perdre ses propriétés originales (telles que la valeur moyenne, la variance, la médiane, etc.). Intuitivement, à mesure que le nombre de catégories de l'histogramme augmente, les propriétés de la donnée symbolique tendent à se rapprocher de celles de la donnée classique. Toutefois, ceci peut mener à la création des catégories « vides », qui n'apportent aucun poids aux calculs qui seront réalisés *a posteriori*. Le choix de la représen-

tation symbolique dépendra donc de l'application étudiée et il revient à l'utilisateur de choisir la représentation symbolique la plus adéquate.

2.3 Bilan

Ce chapitre introduit la procédure utilisée pour la transformation de données classiques en données symboliques. Dans un premier temps, un exemple simple avec 3 espèces d'oiseaux a été étudié. Cet exemple a montré que selon le type d'approche utilisée, des résultats différents peuvent être obtenus. Dans un second temps, l'ADS a été appliquée à des essais dynamiques. Il a été montré qu'un essai peut être représenté symboliquement soit à partir des capteurs soit à partir des paramètres modaux. Dans les deux cas, la représentation symbolique utilisée est basée sur la transformation des données originales (données classiques) en intervalles interquartiles et histogrammes (données symboliques).

Dans la suite, l'ADS est appliquée à différentes méthodes de classification de données où l'objectif est d'extraire des connaissances sur la santé structurale des ouvrages. L'un des apports majeurs de cette thèse est de coupler ces deux approches pour la discrimination de comportements structuraux anormaux.

Chapitre 3

Méthodes de classification des données et analyses de régression

Ce chapitre présente deux grandes familles de méthodes de classification des données : les méthodes dites non supervisées et les méthodes dites supervisées. La première consiste à grouper un ensemble d'observations inconnues (non étiquetées) dans des groupes significatifs (nommés *clusters*[1]). Des étiquettes sont associées aux clusters indirectement à partir des calculs réalisés par différents algorithmes. De cette façon, le but est d'identifier des groupes tels que les observations les plus similaires se retrouvent dans un groupe commun et les plus différentes soient séparées dans des groupes distincts. Pour la deuxième famille de méthode, une collection d'étiquettes (connues *a priori*) est fournie comme donnée d'entrée pour la détermination de différentes classes. Le problème consiste à donner une étiquette à une nouvelle observation sachant que les groupes sont déjà définis. Dans ces méthodes, deux phases sont nécessaires pour classer les données. Dans la première phase (dite d'entraînement), des étiquettes connues sont utilisées pour que l'algorithme de classification apprenne les descriptions relatives à chaque classe. Dans la seconde phase (dite de test), l'algorithme est utilisé pour classer une nouvelle observation. Il est important de signaler que, dans le cadre de cette thèse, les données d'entrée sont toujours des représentations symboliques d'essais dynamiques (évaluées à partir soit de signaux, soit de paramètres modaux).

Une fois les méthodes de classification présentées, deux types de techniques de régression des données sont introduits. L'objectif est d'étudier l'influence de l'environnement extérieur sur la détection d'endommagements structuraux. En particulier, l'influence de la température sur les fréquences propres et déformées modales est analysée. La première technique est basée sur la régression linéaire multiple et la seconde sur la régression non linéaire utilisant des réseaux de neurones. Dans un premier temps, la construction de ces deux modèles est réalisée afin de filtrer les effets thermiques sur les paramètres modaux. Dans un second temps, il est alors possible d'appliquer les méthodes de classification décrites auparavant aux fréquences et modes propres « corrigés » par les modèles de régression. L'idée est d'observer comment la classification d'essais dynamiques peut changer sans ou avec prise en compte des effets environnementaux. Par ailleurs, une étude statistique basée sur l'Analyse en Composantes Principales est proposée.

1. Le terme *cluster* est couramment utilisé dans le domaine scientifique pour designer un « groupe » ou une « classe », par exemple. Dans ce mémoire, ce terme est très utilisé en se référant toujours à la description citée.

3.1 Méthodes de classification non supervisée

Le groupement de données (encore appelé partitionnement de données ou *clustering*[2]) est une problématique considérée depuis de nombreuses années dans différents domaines tels que l'économie, les études statistiques, la médecine, etc. (Bock & Diday 2000). L'objectif, exprimé de manière très générale, est de séparer un ensemble d'objets en différents groupes en fonction d'une certaine notion de similarité (ou de dissimilarité). Les objets considérés comme similaires sont ainsi associés au même cluster alors que ceux considérés comme différents, sont associés à des clusters distincts. Intuitivement, l'objectif du clustering est de regrouper un ensemble de données de la manière la plus naturelle possible.

L'un des atouts des méthodes de clustering est lié à la capacité d'organiser de grands ensembles de données pour aider à simplifier l'information et généraliser l'analyse de différents comportements. Cette information simplifiée peut également être employée pour réaliser des prédictions et effectuer des hypothèses sur la base des clusters constitués.

Différents types de clustering sont proposés dans la littérature. Deux familles de méthodes sont distinguées : les méthodes *hiérarchiques* et les méthodes *par partition*. Dans le premier cas, une hiérarchie de clusters est formée pour caractériser la similarité (ou dissimilarité) entre les objets. Deux types de méthodes hiérarchiques se distinguent :

- les méthodes divisives ou descendantes. Elles démarrent avec un seul cluster qui regroupe l'ensemble des objets. Cet ensemble initial est alors divisé successivement en plusieurs clusters selon un certain critère jusqu'à ce que tous les objets soient classés dans des clusters différents (où chaque cluster ne contient qu'un seul objet) ;
- les méthodes agglomératives ou ascendantes. Elles démarrent avec autant de clusters que d'objets. Les clusters sont fusionnés successivement selon un certain critère jusqu'à ce que tous les objets soient finalement regroupés dans un seul cluster.

Dans le cas de clustering par partition, les clusters sont aussi définis par des critères de proximité, mais la notion de hiérarchie entre eux n'est pas nette.

Dans le cadre de cette thèse, la plupart des simulations de clustering ont été réalisées en utilisant le logiciel SODAS (Symbolic Official Data Analysis System (Diday & Noirhomme-Fraiture 2007)). Ce logiciel est gratuit et a été développé dans le cadre du projet ASSO (Analysis System of Symbolic Official data). Dans les paragraphes suivants, après avoir défini les mesures de distances pour caractériser les dissimilarités entre les objets, trois méthodes de partitionnement de données utilisées sont détaillées. Les deux premières sont basées sur le clustering hiérarchique et la troisième sur le regroupement par partition.

3.1.1 Calcul de mesures de dissimilarité et d'inertie

Avant d'introduire les méthodes de clustering, il est nécessaire de définir des mesures de distance utilisées pour déterminer des similarités/dissimilarités entre objets et créer des clusters. La dissimilarité entre objets est évaluée ici par une fonction qui estime la « distance » les séparant. Si la distance entre deux objets est petite, ils sont considérés comme proches et donc, similaires (Esposito et al. 2000). Au contraire, si la distance est grande, les objets sont « éloignés » et donc, différents. Une fois cette fonction de distance définie, la tâche de clustering consiste à minimiser la somme des distances entre membres d'un même cluster (défini comme l'inertie intra-cluster), et maximiser celle entre clusters (défini comme l'inertie inter-cluster). La procédure de clustering

2. Également au terme *cluster*, *clustering* est souvent utilisé dans ce mémoire. Ce terme signifie une procédure de partitionnement ou groupement de données.

doit finalement fournir un ensemble de clusters avec une grande similarité intra-cluster (groupes intrinsèques) et une grande dissimilarité inter-cluster (groupes isolés).

Le choix d'une mesure de distance constitue un pas très important dans le clustering. Manifestement, un objet A peut être considéré proche d'un objet B pour une mesure particulière, mais aussi distant d'après une autre mesure. Dans la littérature courante, plusieurs types de mesures sont répertoriés (distances Euclidienne, de Gowda-Diday, d'Ichino-Yaguchi, de De Carvalho, etc.) (Malerba et al. 2001).

Dans cette thèse, deux mesures de distance sont utilisées : la distance de Hausdorff (Billard & Diday 2006) et la distance catégorique (Billard & Diday 2006). La première est employée dans le cas où les signaux ou les paramètres modaux sont transformés en intervalles interquartiles. La deuxième est utilisée quand les données classiques sont transformées en histogrammes. Ces distances ont été choisies après comparaison avec d'autres distances introduites dans des études précédentes et disponibles dans la littérature (Cury, Crémona & Diday 2010).

Les essais dynamiques $\mathbf{E}_1$ et $\mathbf{E}_2$ sont considérés dans la suite et sont représentés par des intervalles interquartiles décrits par $[E_{1,\text{inf}}; E_{1,\text{sup}}]_q$ et $[E_{2,\text{inf}}; E_{2,\text{sup}}]_q$, respectivement, avec q étant la variable considérée (un capteur, si les essais sont décrits par des signaux ; une fréquence, si les essais sont représentés par des fréquences propres ; un mode, si les essais sont décrits par des déformées modales).

En définissant $E_{i,\text{inf}}$ et $E_{i,\text{sup}}$ comme les bornes inférieure et supérieure de l'intervalle interquartile de l'essai $\mathbf{E}_i$, respectivement, la distance de Hausdorff $\varphi_q(\mathbf{E}_1, \mathbf{E}_2)$ entre les essais $\mathbf{E}_1$ et $\mathbf{E}_2$ est calculée par

$$\varphi_q(\mathbf{E}_1, \mathbf{E}_2) = \text{Max}\left(|E_{1,\text{inf}} - E_{2,\text{inf}}|_q, |E_{1,\text{sup}} - E_{2,\text{sup}}|_q\right) \tag{3.1}$$

Lorsque toutes les variables et tous les essais sont considérées, l'Équation (3.1) peut être généralisée de la manière suivante

$$d^2(\mathbf{E}_i, \mathbf{E}_j) = \sum_{q=1}^{v} \left(\frac{\varphi_q(\mathbf{E}_i, \mathbf{E}_j)}{H_q}\right)^2 \quad i, j = 1, ..., n \tag{3.2}$$

où $d(.)$ représente les éléments de la matrice de distance normalisée, v le nombre de variables utilisées, n le nombre total d'essais et H_q le paramètre de normalisation calculé par

$$H_q^2 = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\varphi_q(\mathbf{E}_i, \mathbf{E}_j)\right]^2 \tag{3.3}$$

Dans les cas où les essais sont représentés par des histogrammes, la mesure de distance catégorique est utilisée. Considérant la même paire d'essais $\mathbf{E}_1$ et $\mathbf{E}_2$, la distance catégorique est

$$\varphi_q(\mathbf{E}_1, \mathbf{E}_2) = \sum_{k=1}^{m} \left(\frac{b_1^{q,k} - b_2^{q,k}}{b_1^{q,k} + b_2^{q,k}}\right) \tag{3.4}$$

où m est le nombre de catégories de l'histogramme et $b_1^{q,k}$ la proportion correspondante à la $k^{\text{ème}}$ catégorie du $q^{\text{ème}}$ capteur de l'essai $\mathbf{E}_1$.

De façon similaire, l'Équation (3.4) peut être généralisée en considérant toutes les variables par

$$d^2(\mathbf{E}_i, \mathbf{E}_j) = \sum_{q=1}^{v} \sum_{k=1}^{m} \left[\left(v \sum_{u=1}^{n} b_u^{q,k}\right)^{-1} \left(b_i^{q,k} - b_j^{q,k}\right)^2\right], \quad i, j = 1, \ldots, n \tag{3.5}$$

Afin d'optimiser le partitionnement des essais, il est nécessaire de minimiser la somme des distances entre les essais d'un cluster (l'inertie intra-cluster, dénotée par $I(.)$), ce qui est réalisé en minimisant l'inertie totale, nommée $W(.)$, de l'ensemble des clusters.

L'inertie intra-cluster d'un cluster $\mathbf{C}^k$ contenant n_k essais est définie par

$$I(\mathbf{C}^k) = \frac{1}{n \times n_k} \sum_{i=1}^{n_k} \sum_{j>i=1}^{n_k} d^2(\mathbf{E}_i, \mathbf{E}_j) \tag{3.6}$$

où $d(.)$ est calculé à partir soit de l'Équation (3.2) soit de l'Équation (3.5). Étant donné que la matrice de distances est symétrique, seuls les éléments de la matrice triangulaire supérieure sont pris en compte pour le calcul de l'inertie.

Considérant une partition $\mathbf{P}_r = (\mathbf{C}^1, \mathbf{C}^2, ..., \mathbf{C}^r)$, l'inertie totale intra-cluster $W(.)$ est évaluée comme la somme de toutes les inerties intra-cluster (Billard & Diday 2006)

$$W(\mathbf{P}_r) = \sum_{k=1}^{r} I(\mathbf{C}^k) \tag{3.7}$$

où r est le nombre de clusters considérés dans l'analyse.

Par ailleurs, l'inertie inter-clusters est définie comme suit

$$B(\mathbf{P}_r) = W(\mathbf{\Omega}) - W(\mathbf{P}_r) \tag{3.8}$$

où $\mathbf{\Omega}$ représente l'ensemble total d'essais. Il est évident qu'en minimisant l'inertie totale intra-cluster, l'inertie inter-cluster est par conséquent maximisée.

3.1.2 Méthode des divisions hiérarchiques

Dans le domaine de l'analyse de données, les méthodes de classification hiérarchique de type divisives ont été principalement développées dans les années 70. Son application aux données symboliques a été introduite par (Chavent 1997). Ces méthodes sont itératives et procèdent à chaque étape au partitionnement des données de façon descendante (souvent appelée « top-down »). Initialement, l'ensemble total des clusters est représenté par un seul cluster $\mathbf{C}$. La démarche consiste alors en une succession de divisions descendantes, comprenant autant de niveaux que nécessaires pour produire une hiérarchie $\mathbf{Y} = \{\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^r\}$. Théoriquement, il est possible d'arriver à un partitionnement avec n clusters où chaque cluster ne contient qu'un seul essai.

Dans la pratique, le processus de partitionnement peut s'arrêter plus tôt et un nombre de clusters r est généralement spécifié au préalable avec $r < n$. L'exemple d'un cluster $\mathbf{C} = \underset{k=1,\ldots,r}{\cup} \mathbf{C}^k$ contenant tous les essais dénotés par $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_n\}$ est considéré. Ce cluster est alors fractionné en deux clusters distincts $\mathbf{C}^1$ et $\mathbf{C}^2$. Ce processus est effectué en déterminant quels essais satisfont ou non un (des) critère(s) logique(s) $q_l(.)$. Pour un essai hypothétique $\mathbf{E}_i$,

$$\begin{aligned} \mathbf{C}^1 &= \{\mathbf{E}_i | q_l(\mathbf{E}_i) \text{ est vrai}\} \\ \mathbf{C}^2 &= \{\mathbf{E}_i | q_l(\mathbf{E}_i) \text{ est faux}\} \end{aligned} \tag{3.9}$$

la définition du (des) critère(s) logique(s) est basée sur l'évaluation d'une variable (capteur, fréquence ou mode propre) dite « discriminante » et de sa valeur associée nommée « valeur de coupure ». En effet, la paire variable discriminante/valeur de coupure est obtenue à partir de la minimisation des inerties intra-cluster, et par conséquent, la maximisation de l'inertie inter-clusters. L'exemple présenté dans le Tableau 3.1 est considéré où cinq essais $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4$ et $\mathbf{E}_5\}$ hypothétiques sont représentés par des signaux transformés en intervalles interquartiles. Ainsi, les

Essai	Capteur 1 [mm·s^{-2}]	Capteur 2 [mm·s^{-2}]
$\mathbf{E}_1$	[410 ; 460]	[165 ; 175]
$\mathbf{E}_2$	[390 ; 430]	[150 ; 170]
$\mathbf{E}_3$	[130 ; 190]	[107 ; 153]
$\mathbf{E}_4$	[410 ; 630]	[140 ; 170]
$\mathbf{E}_5$	[170 ; 450]	[155 ; 175]

TABLEAU 3.1 – Exemple d'essais représentés par des capteurs transformés en intervalles interquartiles.

variables utilisées sont le « Capteur 1 » et le « Capteur 2 ». Dans cet exemple, une partition avec trois clusters est souhaitée.

Le premier pas consiste à calculer les valeurs moyennes vm_i $(i = 1, \ldots, n)$ de chaque intervalle interquartile et à réordonner les essais de telle sorte que ces valeurs soient classées par ordre croissant. Les Tableaux 3.2 et 3.3 détaillent cette procédure (les unités ne sont pas données afin de simplifier les notations).

Essai	vm_i (Capteur 1)	vm_i (Capteur 2)
$\mathbf{E}_1$	435	170
$\mathbf{E}_2$	410	160
$\mathbf{E}_3$	160	130
$\mathbf{E}_4$	520	155
$\mathbf{E}_5$	310	165

TABLEAU 3.2 – Calcul des valeurs moyennes par capteur pour les essais dynamiques.

Capteur 1		Capteur 2	
Essai	vm_i	Essai	vm_i
$\mathbf{E}_3$	160	$\mathbf{E}_3$	130
$\mathbf{E}_5$	310	$\mathbf{E}_4$	155
$\mathbf{E}_2$	410	$\mathbf{E}_2$	160
$\mathbf{E}_1$	435	$\mathbf{E}_5$	165
$\mathbf{E}_4$	520	$\mathbf{E}_1$	170

TABLEAU 3.3 – Classement des valeurs moyennes par ordre croissant selon chaque capteur.

Les valeurs de coupure vc_i sont alors calculées en prenant la valeur moyenne de vm_i et vm_{i+1}

$$vc_i = \frac{vm_i + vm_{i+1}}{2} \quad i = 1, \ldots, n-1 \tag{3.10}$$

où n est toujours le nombre d'essais considérés.

Il est noté que chaque capteur est pris en compte séparément pour le calcul des ses possibles valeurs de coupures (cf. Tableau 3.4). Pour chaque paire capteur/valeur de coupure les inerties intra et inter-clusters sont calculées. La paire optimale est donc celle qui minimise l'inertie intra-cluster et maximise l'inertie inter-cluster. Par exemple, la première ligne du Tableau 3.4 illustre la première possibilité de partitionnement. Dans ce cas, le Capteur 1 est un candidat à variable discriminante et la valeur de coupure correspondante est 235 mm·s^{-2}. Pour ce critère, seulement l'essai $\mathbf{E}_3$ se trouve dans le cluster $\mathbf{C}^1$ (par conséquent, l'inertie intra-cluster est égale à zéro), tandis que tous les autres essais sont regroupés dans le cluster $\mathbf{C}^2$ (en indiquant une inertie intra-cluster égale à 11670). Par ailleurs, l'inertie totale intra-cluster et l'inertie inter-clusters sont égales à 11670 et 628930, respectivement. La totalité des calculs effectués et le choix de

la paire optimale sont détaillés dans le Tableau 3.4. Bien que les capteurs soient sélectionnés séparément, le calcul de la matrice des distances (défini par l'Équation (3.2)) prend toujours en compte la totalité des capteurs utilisés dans l'analyse. Cette remarque est importante, car le calcul des inerties dépend directement de la matrice de distances évaluée.

	Essai	vm [mm·s^{-2}]	vc [mm·s^{-2}]	$I(\mathbf{C}^1)$	$I(\mathbf{C}^2)$	$W(\mathbf{C}^1 \cup \mathbf{C}^2)$	$B(\mathbf{C}^1 \cup \mathbf{C}^2)$
	$\mathbf{E}_3$	160	235	0	11670	11670	628930
Coupure	$\mathbf{E}_5$	310	**360**	6760	4654	**11414**	**629186**
avec le	$\mathbf{E}_2$	410	422,5	12240	2890	15130	625470
Capteur 1	$\mathbf{E}_1$	435	477,5	16025	0	16025	624575
	$\mathbf{E}_4$	520	-	-	-	-	-
	$\mathbf{E}_3$	130	**142,5**	0	11670	**11670**	**628930**
Coupure	$\mathbf{E}_4$	155	157,5	19360	7127	26487	614113
avec le	$\mathbf{E}_2$	160	162,5	20080	5760	25840	614760
Capteur 2	$\mathbf{E}_5$	165	167,5	23740	0	23740	616860
	$\mathbf{E}_1$	170	-		-		

TABLEAU 3.4 – Division du cluster $\mathbf{C}$, calcul des inerties et choix de la paire optimale.

Dans le Tableau 3.4, $W(\mathbf{C}^1 \cup \mathbf{C}^2) = I(\mathbf{C}^1) + I(\mathbf{C}^2)$, et $B(\mathbf{C}^1 \cup \mathbf{C}^2)$ est calculé comme défini par l'Équation (3.8).

La paire variable discriminante/valeur de coupure optimale est celle correspondante au minimum des minima de la somme des inerties intra-cluster calculées. Dans l'exemple considéré, c'est la paire Capteur 1/360 qui satisfait cette condition ($W(\mathbf{C}^1 \cup \mathbf{C}^2) = 11414$). Par conséquent, la question logique à poser est la suivante : « Est-ce que Capteur 1 $\leq$ 360 mm· s^{-2} ? ». Si la réponse est oui, les essais qui satisfont ce critère sont classés dans le cluster $\mathbf{C}^1$. Si la réponse est non, ils sont classés dans le cluster $\mathbf{C}^2$. Plus précisément, ce sont les valeurs moyennes vm_i du Capteur 1 (de tous les essais) qui sont comparées à la valeur de coupure 360 mm· s^{-2}. Dans cet exemple, la division obtenue est

$$\mathbf{C}^1 = \{\mathbf{E}_3, \mathbf{E}_5\}$$
$$\mathbf{C}^2 = \{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_4\}$$

Finalement, une troisième division est réalisée où il est nécessaire de définir au préalable le cluster qui sera divisé. Une procédure similaire à celle introduite précédemment est utilisée. En ce sens, le cluster choisi est celui dont la somme des inerties inter-clusters correspond au maximum des maxima (i.e., le cluster le plus hétérogène). Intuitivement, ce raisonnement est logique puisque l'objectif est de toujours trouver des groupes les plus homogènes possibles. Le Tableau 3.5 contient les deux clusters obtenus lors du partitionnement précédent, aussi bien que les nouvelles valeurs de coupure et les inerties intra et inter-clusters calculées.

Dans le Tableau 3.5, $W(\mathbf{C}^{1'} \cup \mathbf{C}^{1''}) = I(\mathbf{C}^{1'}) + I(\mathbf{C}^{1''})$, et $B(\mathbf{C}^{1'} \cup \mathbf{C}^{1''})$ est calculé comme défini par l'Équation (3.8).

Les résultats montrés dans le Tableau 3.5 indiquent deux critères possibles, étant donné que deux minimums égaux sont obtenus. Dans ce cas, les questions logiques à poser sont : « Est-ce que Capteur 1 $\leq$ 235 mm· s^{-2} ? » ou « Est-ce que Capteur 2 $\leq$ 147,5 mm· s^{-2} ? »

Quelle que soit la question, le partitionnement final est le même

$$\mathbf{C}^{1'} = \{\mathbf{E}_3\}$$
$$\mathbf{C}^{1''} = \{\mathbf{E}_5\}$$
$$\mathbf{C}^2 = \{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_4\}$$

	Essai	vm [mm·s^{-2}]	vc [mm·s^{-2}]	$I(\mathbf{C}^{1'})$	$I(\mathbf{C}^{1''})$	$W(\mathbf{C}^{1'} \cup \mathbf{C}^{1''})$	$B(\mathbf{C}^{1'} \cup \mathbf{C}^{1''})$
Coupure avec le Capteur 1							
Cluster $\mathbf{C}^1$	$\mathbf{E}_3$	160	**235**	0	0	**0**	**6760**
	$\mathbf{E}_5$	310	-	-	-	0	-
Cluster $\mathbf{C}^2$	$\mathbf{E}_2$	410	422,5	0	2890	2890	1763
	$\mathbf{E}_1$	435	**477,5**	90	0	**90**	**4563**
	$\mathbf{E}_4$	520	-	-	-	-	-
Coupure avec le Capteur 2							
Cluster $\mathbf{C}^1$	$\mathbf{E}_3$	130	**147,5**	0	0	**0**	**6760**
	$\mathbf{E}_5$	165	-	-	-	0	-
Cluster $\mathbf{C}^2$	$\mathbf{E}_4$	155	**157,5**	0	90	**90**	**4563**
	$\mathbf{E}_2$	160	165	4000	0	4000	653
	$\mathbf{E}_1$	170	-	-	-	-	-

TABLEAU 3.5 – Second partitionnement.

Des critères plus rigoureux ont été développés afin de départager des situations comme celle de l'exemple précédent. Un critère souvent utilisé est basé sur le calcul des distances entre les objets du cluster qui a été divisé. Dans cet exemple, ces objets sont les essais $\mathbf{E}_3$ et $\mathbf{E}_5$. Les distances entre ces essais d'après les deux capteurs sont calculées comme suit

$$\begin{aligned} \text{Capteur 1} : d_1(\mathbf{E}_3, \mathbf{E}_5) &= \text{Max}(|130-170|, |190-450|) = 260 \text{ mm}\cdot \text{s}^{-2} \\ \text{Capteur 2} : d_2(\mathbf{E}_3, \mathbf{E}_5) &= \text{Max}(|107-155|, |153-175|) = 48 \text{ mm}\cdot \text{s}^{-2} \end{aligned}$$

Le critère définit que le capteur « ayant » la plus grande distance, est celui choisi comme variable discriminante. Étant donné que $d_1(\mathbf{E}_3, \mathbf{E}_5) > d_2(\mathbf{E}_3, \mathbf{E}_5)$, la question logique pour la division du cluster $\mathbf{C}^1$ est « Est-ce que le Capteur 1 $\leq$ 235 mm· s^{-2} ? ». Comme mentionné précédement, le processus du clustering peut continuer jusqu'au point où chaque cluster est représenté par un seul essai. La Figure 3.1 illustre une représentation graphique du clustering réalisé pour l'exemple introduit dans ce paragraphe.

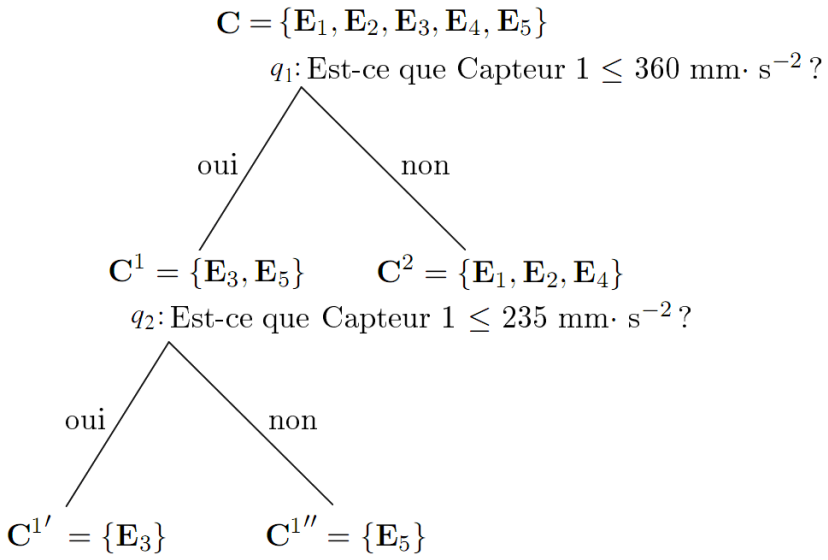


FIGURE 3.1 – Division hiérarchique obtenue à partir de l'exemple utilisé.

L'Algorithme 1 détaille la procédure de clustering exposée dans ce paragraphe. Il faut rappeler que v correspond au nombre de variables utilisées dans l'analyse. Par ailleurs, n_{vc}^i représente le

nombre de valeurs de coupure par variable.

Entrée : Cluster $\mathbf{C}(\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_n)$
Sortie : Clusters $\mathbf{C}'$ et $\mathbf{C}''$ avec W minimal

pour $i \leftarrow 1$ **a** v **faire**
  pour $j \leftarrow 1$ **a** n_{cv}^{i} **faire**
    si `Variable` $i \leq vc(j)$ **alors**
      $\mathbf{C}'_{i,j} \leftarrow$ **Essais** « Vrai » ;
    sinon
      $\mathbf{C}''_{i,j} \leftarrow$ **Essais** « Faux » ;
    fin
    $W_{i,j} \leftarrow I(\mathbf{C}'_{i,j}) + I(\mathbf{C}''_{i,j})$;
    $(i_0, j_0) \leftarrow \underset{(i,j)}{\texttt{arg min}}\{W_{i,j}\}$;
    $\mathbf{C}' \leftarrow \mathbf{C}'_{(i_0,j_0)}$;
    $\mathbf{C}'' \leftarrow \mathbf{C}''_{(i_0,j_0)}$;
  fin
fin
Répéter la procédure pour $\mathbf{C}'$ ou $\mathbf{C}''$.

Algorithme 1: Pseudo-algorithme de la méthode des divisions hiérarchiques.

Finalement, si les essais dynamiques sont décrits par des variables de type histogramme, un procédé similaire est également mis en œuvre. Dans ce cas, la valeur de coupure est définie par la moyenne des centres de gravité des distributions. Deux essais hypothétiques $\mathbf{E}_i$ et $\mathbf{E}_j$ sont maintenant considérés, chacun étant représenté par un seul capteur transformé en histogrammes avec 5 catégories

$$\begin{aligned}\mathbf{E}_i &= \{[0-20](0,10); [20-40](0,15); [40-60](0,25); [60-80](0,30); [80-100](0,20)\} \\ \mathbf{E}_j &= \{[0-20](0,05); [20-40](0,05); [40-60](0,40); [60-80](0,25); [80-100](0,25)\}\end{aligned} \tag{3.11}$$

Dans l'Équation (3.11), les valeurs entre crochets définissent les bornes inférieure et supérieure de chaque catégorie et les valeurs entre parenthèses représentent les proportions évaluées pour chaque catégorie. Le centres de gravité sont définis par

$$\begin{aligned}cg_i &= 10 \times (0,10) + 30 \times (0,15) + 50 \times (0,25) + 70 \times (0,30) + 90 \times (0,20) = 57 \text{ mm} \cdot \text{s}^{-2} \\ cg_j &= 10 \times (0,05) + 30 \times (0,05) + 50 \times (0,40) + 70 \times (0,25) + 90 \times (0,25) = 62 \text{ mm} \cdot \text{s}^{-2}\end{aligned} \tag{3.12}$$

Ainsi, la valeur de coupure est donnée par

$$vc = \frac{cg_i + cg_j}{2} = 59,5 \text{ mm} \cdot \text{s}^{-2} \tag{3.13}$$

La démarche utilisée pour la détermination des questions logiques est similaire à celle présentée en considérant des intervalles interquartiles. Ici, les valeurs des centres de gravité cg_k, $(k = 1, \ldots, n)$ de la variable la plus discriminante (de tous les essais) sont comparées à la valeur de coupure calculée vc.

3.1.3 Méthode des agglomérations hiérarchiques

La méthode des agglomérations hiérarchiques a été initialement proposée par (Bertrand 1986) pour une classification de données classiques. La généralisation à des données symboliques a été

considérée par (Brito & Diday 1990) et (Brito 1994) quelques années plus tard. Cette méthode consiste en un processus de partitionnement de type ascendant (« bottom-up ») et les principes pour la détermination des clusters, décrits pour la méthode de divisions hiérarchiques, sont appliqués de manière inversée. Initialement, chaque essai est représenté par un seul cluster $\mathbf{C}^k, k = 1, ..., n$. La démarche consiste alors en une succession de regroupements ascendants comprenant autant de niveaux que nécessaires pour produire un seul cluster $\mathbf{C}$ regroupant tous les essais.

Dans cette méthode, les clusters sont associés selon des critères de distance, en envisagent encore la minimisation des inerties intra-cluster et la maximisation des inerties inter-clusters. Le procédé de partitionnement commence par regrouper les clusters les plus proches selon une mesure appelée *degré de généralité* (Billard & Diday 2006). Afin d'illustrer cette procédure, le premier niveau d'agglomération contenant l'ensemble d'essais $\mathbf{\Omega} = \{\mathbf{E}_1, \ldots, \mathbf{E}_n\}$ représentés par v variables transformées en intervalles interquartiles est considéré. Le degré de généralité entre deux essais $\mathbf{E}_i$ et $\mathbf{E}_j$ s'écrit sous la forme (Diday & Noirhomme-Fraiture 2007)

$$dg(\mathbf{E}_i, \mathbf{E}_j) = \prod_{q=1}^{v} \frac{\text{Max}(|E_{i,\text{sup}}, E_{j,\text{sup}}|_q) - \text{Min}(|E_{i,\text{inf}}, E_{j,\text{inf}}|_q)}{\text{Max}(|E_{sup}|_q) - \text{Min}(|E_{inf}|_q)}, \quad i, j = 1, \ldots, n \tag{3.14}$$

où $\text{Max}(|E_{\text{sup}}|_q)$ et $\text{Min}(|E_{\text{inf}}|_q)$ représentent les valeurs absolues maximale et minimale pour chaque variable q, respectivement. Dans le Tableau 3.1, le degré de généralité entre les essais $\mathbf{E}_1$ et $\mathbf{E}_2$ est calculé par

$$dg(\mathbf{E}_1, \mathbf{E}_2) = \underbrace{\frac{(460 - 390)}{(630 - 130)}}_{\text{Capteur 1}} \times \underbrace{\frac{(175 - 150)}{(175 - 107)}}_{\text{Capteur 2}} = 0,05 \tag{3.15}$$

De façon similaire, les autres distances sont évaluées pour chaque paire d'essais procédant à la construction de la matrice des distances. Dans l'exemple considéré, la matrice des distances est

$$\mathbf{dg}_1 = \begin{array}{c} \begin{matrix} \mathbf{E}_1 & \mathbf{E}_2 & \mathbf{E}_3 & \mathbf{E}_4 & \mathbf{E}_5 \end{matrix} \\ \begin{pmatrix} 0 & \mathbf{0,05} & 0,66 & 0,23 & 0,17 \\ \mathbf{0,05} & 0 & 0,56 & 0,21 & 0,20 \\ 0,66 & 0,56 & 0 & 0,38 & 0,64 \\ 0,23 & 0,21 & 0,38 & 0 & 0,47 \\ 0,17 & 0,20 & 0,64 & 0,47 & 0 \end{pmatrix} \end{array} \begin{matrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \\ \mathbf{E}_5 \end{matrix} \tag{3.16}$$

La paire d'essais ayant la distance la plus petite est la première à être regroupée. Dans cet exemple, le premier cluster est $\mathbf{C}^1 = \{\mathbf{E}_1, \mathbf{E}_2\}$. Ensuite, la description symbolique de ce cluster correspond à l'union des descriptions des essais $\mathbf{E}_1$ et $\mathbf{E}_2$. En d'autres termes

$$\mathbf{C}^1 = [\text{Min}(|E_{1,\text{inf}}, E_{2,\text{inf}}|_q), \text{Max}(|E_{1,\text{sup}}, E_{2,\text{sup}}|_q)], \quad q = 1, \ldots, v \tag{3.17}$$

Dans l'exemple étudié, il vient

$$\mathbf{C}^1 = \{\underbrace{[390; 460]}_{\text{Capteur 1}}, \underbrace{[150; 175]}_{\text{Capteur 2}}\} \tag{3.18}$$

Le deuxième regroupement est réalisé en utilisant la même procédure que celle décrite précédemment en considérant maintenant le cluster $\mathbf{C}^1$ à la place des essais $\mathbf{E}_1$ et $\mathbf{E}_2$. La matrice

des distances est

$$\mathbf{dg}_2 = \begin{array}{c} \begin{matrix} \mathbf{C}^1 & \mathbf{E}_3 & \mathbf{E}_4 & \mathbf{E}_5 \end{matrix} \\ \begin{pmatrix} 0 & 0,66 & 0,25 & \mathbf{0,21} \\ 0,66 & 0 & 0,38 & 0,44 \\ 0,25 & 0,38 & 0 & 0,47 \\ \mathbf{0,21} & 0,44 & 0,47 & 0 \end{pmatrix} \end{array} \begin{matrix} \mathbf{C}^1 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \\ \mathbf{E}_5 \end{matrix} \tag{3.19}$$

Le deuxième niveau d'agglomération correspond au regroupement du cluster $\mathbf{C}^1$ avec l'essai $\mathbf{E}_5$. En ce sens, $\mathbf{C}^2 = \{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_5\}$ et ce cluster est décrit par $\mathbf{C}^2 = \{[170; 460], [150; 175]\}$. Pour le troisième regroupement, la matrice des distances est écrite de la manière suivante

$$\mathbf{dg}_3 = \begin{array}{c} \begin{matrix} \mathbf{C}^2 & \mathbf{E}_3 & \mathbf{E}_4 \end{matrix} \\ \begin{pmatrix} 0 & 0,66 & 0,47 \\ 0,66 & 0 & \mathbf{0,38} \\ 0,47 & \mathbf{0,38} & 0 \end{pmatrix} \end{array} \begin{matrix} \mathbf{C}^2 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \end{matrix} \tag{3.20}$$

Le troisième cluster est donc formé par les essais $\mathbf{E}_3$ et $\mathbf{E}_4$. Avec cette dernière agglomération, la hiérarchie est complètement définie. La Figure 3.2 illustre le regroupement final obtenu et les clusters assemblés. Il est possible de noter que cette méthode de partitionnement permet également de montrer une mesure de proximité parmi les clusters vis-à-vis de la « différence de hauteur » entre eux. Il est manifeste que les clusters $\mathbf{C}^1$ et $\mathbf{C}^2$ sont plus proches que les clusters $\mathbf{C}^1$ et $\mathbf{C}^3$ ou encore $\mathbf{C}^2$ et $\mathbf{C}^3$ (cf. Figure 3.2).

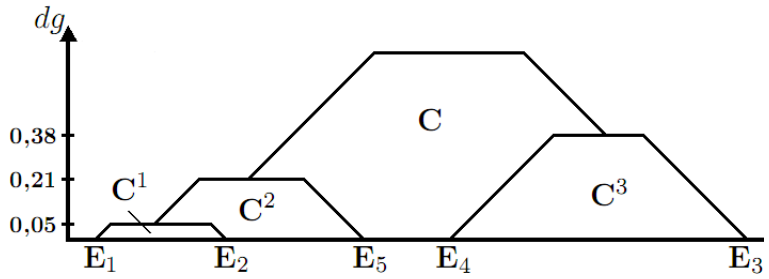


FIGURE 3.2 – Exemple d'une agglomération hiérarchique avec 3 clusters.

Finalement, si les essais dynamiques sont décrits par des variables de type histogramme, le degré de généralité est calculé différemment. Dans ce cas, l'expression est (Diday & Noirhomme-Fraiture 2007)

$$dg(\mathbf{E}_i, \mathbf{E}_j) = \prod_{q=1}^{v} \left(\frac{1}{\sqrt{m}} \sum_{k=1}^{m} \left(\sqrt{|b_i^{q,k} - b_j^{q,k}|} \right) \right), \quad q = 1, \ldots, v \tag{3.21}$$

où m est le nombre de catégories de l'histogramme, $b_i^{q,k}$ la proportion correspondante à la $k^{\text{ème}}$ catégorie du $q^{\text{ème}}$ capteur de l'essai $\mathbf{E}_i$. Ainsi, pour les essais présentés dans l'Équation (3.11), le degré de généralité est calculé par

$$\begin{aligned} dg(\mathbf{E}_i, \mathbf{E}_j) = \frac{1}{\sqrt{5}} \Big(& \sqrt{|0,10 - 0,05|} + \sqrt{|0,15 - 0,05|} + \sqrt{|0,25 - 0,40|} + \sqrt{|0,30 - 0,25|} + \\ & + \sqrt{0,20 - 0,25} \Big) = 0,61 \end{aligned} \tag{3.22}$$

3.1.4 Méthode des nuées dynamiques

Il a été montré aux paragraphes 3.1.2 et 3.1.3 que les méthodes de partitionnement hiérarchique organisent les objets de manière structurée. L'une des caractéristiques les plus importantes de ces techniques est liée à l'impact visuel des arbres et des « pyramides » générés. Ces sorties graphiques permettent à l'utilisateur non seulement de voir comment les objets sont divisés (ou fusionnés) dans les clusters, mais aussi de décider à quel niveau de regroupement s'arrêter pour que les résultats soient pertinents pour l'application étudiée.

Les méthodes de regroupement non hiérarchique sont connues sous le nom de méthodes de clustering par partition, dans le sens où elles produisent des séparations entre les objets afin d'assembler des groupes homogènes. La problématique pour ce type de technique est la suivante : Étant donnés n objets dans un espace v-dimensionnel (défini par le nombre de variables v - capteurs, fréquences propres ou déformées modales), comment déterminer r (défini au préalable) clusters de telle sorte que les objets dans un même groupe soient le plus similaires (proches) entre eux par rapport à ceux placés dans des groupes distincts ?

La solution à ce problème est, théoriquement, directe. L'idée est de trouver une partition de clusters dont l'inertie intra-cluster est minimale et l'inertie intra-cluster est maximale. Toutefois, afin de trouver telle partition il est nécessaire d'évaluer les inerties pour toutes les combinaisons possibles contenant r clusters. La principale difficulté rencontrée est liée au nombre de différentes combinaisons de partitionnement existantes, même si la quantité d'objets à regrouper n'est pas trop importante.

Afin de maîtriser le nombre de combinaisons, les inerties sont évaluées uniquement pour un sous-ensemble de partitions. Le problème revient donc à identifier quel sous-ensemble a une bonne possibilité d'avoir la partition optimale (i.e., la partition avec la somme des inerties intra-cluster minimale). L'approche la plus commune est basée sur un algorithme itératif. À partir d'une première partition, les objets sont déplacés d'un cluster à un autre de façon à minimiser les inerties. Ainsi, chaque partition successive est une perturbation de la précédente, ce qui permet de restreindre le nombre de combinaisons évaluées et d'arriver à la solution optimale. Les détails de l'algorithme sont discutés par la suite.

La méthode des nuées dynamiques est basée sur une généralisation de la méthode classique des clusters dynamiques (algorithme K-means) proposée par (MacQueen 1967). Cette méthode a été développée par (Lechevallier 1974) et consiste à minimiser le critère mesurant l'adéquation entre les clusters et ses prototypes respectifs (Diday (1971) et Celeux et al. (1989)). Un prototype est un modèle de description symbolique pour représenter un cluster ou, en d'autres termes, le noyau (l'essai « moyen ») d'un cluster. Il est utilisé comme une référence pour calculer les distances parmi les essais et pour définir chaque cluster.

L'algorithme est constitué de trois étapes (cf. Tableau 3.6) : la première, *initialisation*, consiste à choisir r prototypes $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_r$ générés aléatoirement qui représentent les r clusters $\mathbf{C}^1, \mathbf{C}^2, \ldots, \mathbf{C}^r$. La deuxième et troisième étapes sont appliquées de manière itérative : la deuxième, appelée *attribution*, consiste à évaluer les distances entre tous les objets et les prototypes initiaux. Une fois les distances calculées, chaque objet est attribué à un cluster où la distance entre l'objet et le prototype correspondant à ce cluster est minimale ; pour la troisième étape, appelée *représentation*, les prototypes sont actualisés selon les résultats de l'étape d'attribution. Cette étape est réalisée en calculant et en stockant la somme totale des distances entre le prototype et les objets du cluster. Le nouveau prototype est donc celui qui minimise cette valeur. Les distances sont calculées à partir de l'Équation (3.1) ou de l'Équation (3.4), par exemple. Ces deux étapes sont répétées jusqu'à la convergence (i.e., quand le critère d'adéquation atteint une valeur stationnaire) ou lorsque le nombre maximal d'itérations défini au préalable est atteint (afin d'éviter une convergence excessivement lente).

Entrées : nombre de clusters r, nombre d'itérations n_i et nombre de répétitions n_r **Sortie** : partitionnement optimal $\mathbf{P}_r = \{\mathbf{C}^1, \ldots, \mathbf{C}^r\}$	
Étape	**Procédure**
Pour $i = 1$ à n_r, répéter les étapes 1 à 3	
1. *Initialisation*	Choisir aléatoirement r essais pour représenter les prototypes $\mathbf{G} = \{\mathbf{G}_1, \ldots, \mathbf{G}_r\}$
Répéter jusqu'à la convergence ou après n_i itérations	
2. *Attribution*	Affecter chaque essai $\mathbf{E}_j (j = 1, \ldots, n)$ à un prototype $\mathbf{G}_{k^*}$ tel que $k^* = \arg\min_{k=1,\ldots,r} d(\mathbf{E}_j, \mathbf{G}_k)$ et définir la partition $\mathbf{P}_r = \{\mathbf{C}^1, \ldots, \mathbf{C}^r\}$
3. *Représentation*	Évaluer les prototypes actualisés $\mathbf{G}_k^*$ de façon à minimiser $\sum_{\mathbf{E}_j \in \mathbf{C}^k} d(\mathbf{E}_j, \mathbf{G}_k^*)$

TABLEAU 3.6 – Étapes de l'algorithme de la méthode des nuées dynamiques (adapté de Silva (2005)).

Généralement, l'algorithme des nuées dynamiques est assez robuste et dans certains cas converge en quelques itérations. Toutefois, il est possible qu'il converge vers un minimum local. Afin de diminuer l'occurrence de ce type de problème et d'améliorer la qualité du clustering, l'algorithme est exécuté plusieurs fois avec différentes combinaisons initiales de clusters. Ainsi, la meilleure configuration (celle dont la somme des inerties intra-clusters est le minimum des minima) est sélectionnée comme résultat final et le partitionnement obtenu est considéré optimal. La Figure 3.3 illustre un schéma simplifié de cette méthode avec la représentation de chaque étape de l'algorithme détaillé dans le Tableau 3.6.

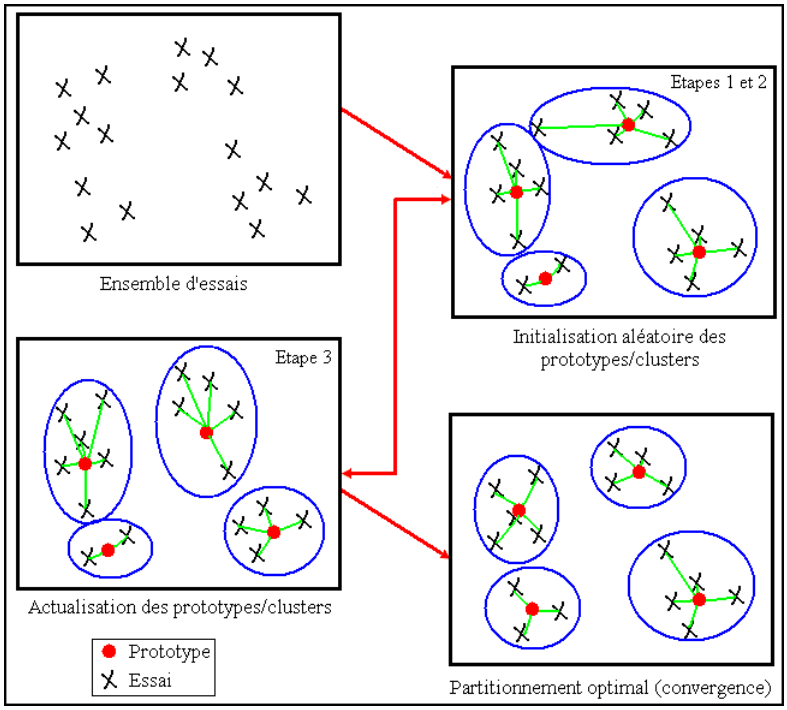


FIGURE 3.3 – Schéma simplifié de l'algorithme des nuées dynamiques.

3.1.5 Indices de partitionnement optimal

Dans les paragraphes 3.1.2, 3.1.3 et 3.1.4, deux des trois méthodes de clustering présentées (les méthodes des divisions hiérarchiques et des nuées dynamiques) nécessitent de fixer au préalable le nombre de clusters r souhaités. Néanmoins, il est délicat de connaître cette valeur *a priori*, d'une part du fait des incertitudes existantes dans l'analyse et d'autre part à cause du nombre, souvent important, de partitionnements possibles. Afin de traiter ce problème, plusieurs indices mathématiques ont été mis au point pour déterminer le partitionnement optimal. Différentes méthodes pour la détermination du nombre optimal de clusters existent aujourd'hui. Au début des années 80, Milligan & Cooper (1985) ont réalisé une étude comparative (à l'aide de simulations de Monte Carlo) détaillée et complète en regroupant les 30 indices les plus utilisés. Ces indices ont été analysés et classés après avoir étudié le point à partir duquel ils peuvent détecter le nombre correct de clusters pour un ensemble de données contenant une structure connue. Les trois meilleurs indices identifiés dans cette étude sont utilisés dans la thèse et présentés dans les paragraphes suivants.

La méthodologie générale utilisée est basée sur le calcul de chaque indice tenant compte d'une partition $\mathbf{P}_s = \{\mathbf{C}^1, \ldots, \mathbf{C}^s\}$ contenant différents nombres de clusters s. Cette valeur est choisie arbitrairement et est normalement plus grande que r (r étant le nombre de clusters considérés dans l'analyse). Il est important de signaler que tous les indices sont calculés pour des partitions avec un nombre de clusters égal ou supérieur à 2, étant donné que $s = 1$ représente l'ensemble des essais classés dans un seul cluster.

3.1.5.1 Indice de Calinski et Harabasz (CH)

L'introduction de l'indice CH (Calinski & Harabasz 1974) permet de comparer l'inertie du $s^{\text{ème}}$ cluster avec celles des autres $s-1$ clusters. Cet indice mesure donc la dissimilarité inter-clusters par rapport à celle intra-cluster. L'expression permettant de relier ces quantités est donnée par

$$CH(\mathbf{P}_j) = \frac{B(\mathbf{P}_j)}{W(\mathbf{P}_j)} \times \frac{(n-j)}{(j-1)}, \quad j = 2, \ldots, s \tag{3.23}$$

où $B(\mathbf{P}_j)$ est l'inertie inter-clusters (cf. Équation (3.8)), $W(\mathbf{P}_j)$ l'inertie totale intra-cluster (cf. Équation (3.7)), n le nombre total d'essais et j le nombre de clusters d'une partition $\mathbf{P}_j$.

D'après (Calinski & Harabasz 1974), la valeur **maximale** absolue de l'indice CH indique le partitionnement optimal, c'est-à-dire le nombre optimal de clusters. Toutefois, il est noté que ce n'est pas toujours la valeur absolue qui importe le plus, mais plutôt la tendance de l'ensemble des valeurs calculées.

3.1.5.2 Indice C^*

L'indice C^* défini par (Hubert & Levine 1976) prend en compte les valeurs des distances intra-cluster du cluster $\mathbf{C}^k$ et celles évaluées inter-clusters. L'équation permettant de le quantifier est

$$C^*(\mathbf{P}_j) = \frac{1}{n} \sum_{k=1}^{j} n_k \frac{(S^k - S^k_{\min})}{(S^k_{\max} - S^k_{\min})}, \quad j = 2, \ldots, s \tag{3.24}$$

où n_k est le nombre d'essais dans un cluster $\mathbf{C}^k$, S^k la somme des distances entre les k essais d'un cluster considéré, $S^k_{\min}$ la somme des k plus petites distances parmi tous les essais et $S^k_{\max}$ la somme des k plus grandes distances parmi tous les essais.

D'après (Hubert & Levine 1976), la valeur **minimale** de C^* indique le partitionnement optimal, sachant que le minimum absolu pour cet indice est égal à zéro.

3.1.5.3 Indice Γ

L'indice Γ (Baker & Hubert 1975) est calculé de la manière suivante

$$\Gamma(\mathbf{P}_j) = \frac{\Gamma_+(\mathbf{P}_j) - \Gamma_-(\mathbf{P}_j)}{\Gamma_+(\mathbf{P}_j) + \Gamma_-(\mathbf{P}_j)}, \quad j = 2, \ldots, s \tag{3.25}$$

où $\Gamma_+(\mathbf{P}_j)$ est le nombre de fois que les distances inter-clusters ont des valeurs plus grandes que les distances intra-cluster et $\Gamma_-(\mathbf{P}_j)$ le nombre de fois que les distances inter-clusters ont des valeurs plus petites que les distances intra-cluster.

Ainsi comme pour l'indice CH, la valeur **maximale** de Γ correspond au partitionnement optimal (Baker & Hubert 1975), sachant le maximum absolu est égal à 1.

3.1.6 Affectation de nouveaux essais

Les méthodes de clustering permettent de considérer un ensemble d'essais dynamiques et de les classer dans plusieurs clusters où chacun représente un comportement structural distinct. Un autre aspect essentiel à considérer est lié à l'affectation de nouveaux essais. En d'autres termes, il est pertinent de développer des approches permettant de classer un nouvel essai à des groupes de comportement connus. Supposons par exemple qu'une méthode de clustering appliquée aux essais acquis pendant une campagne de surveillance dynamique ait identifié deux groupes de comportement différents. Supposons encore que quelques années après, une nouvelle campagne de mesures soit mise en place afin d'observer l'évolution de la réponse dynamique de l'ouvrage, mais aussi celle des paramètres modaux identifiés. La question posée est alors de savoir comment identifier cette évolution. Intuitivement, l'affectation d'un nouvel essai peut être réalisée en comparant les distances entre cet essai et les deux groupes connus. Dans ce cas, il est affecté au groupe pour lequel la distance est la plus petite. Toutefois, il se peut que la distance entre le nouvel essai et le groupe 1 et la distance entre cet essai et le groupe 2 soient très importantes et que, finalement, cet essai n'appartient à aucun de ces deux groupes. Dans ce cas, cet essai est affecté à un nouveau groupe, représentant un troisième comportement structural.

Afin d'apporter des éléments complémentaires de réflexion sur cette problématique, deux approches appliquées en complément des méthodes de clustering sont proposées. Ces approches consistent essentiellement à comparer symboliquement des nouveaux essais à ceux regroupés dans différents clusters et à définir des seuils de distance pour leurs affectations. Ces démarches peuvent être appliquées aux essais représentés par des signaux ou des paramètres modaux.

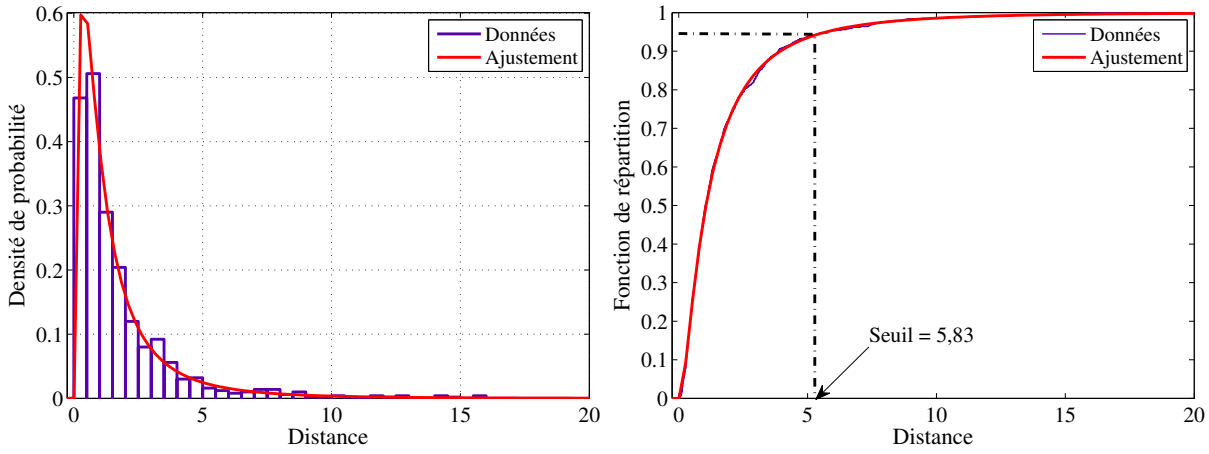
3.1.6.1 Détermination de seuils d'appartenance

L'exemple d'une partition $\mathbf{P}_2 = \{\mathbf{C}^1, \mathbf{C}^2\}$ où $\mathbf{C}^1 = \{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_4\}$ et $\mathbf{C}^2 = \{\mathbf{E}_3, \mathbf{E}_5\}$ est maintenant considéré. Dans un premier temps, l'objectif est d'affecter le nouvel essai $\mathbf{E}_6$ à l'un de ces deux groupes. Toutefois, cet essai peut ne pas appartenir véritablement au cluster auquel il a été affecté. En effet, il est possible que cet essai représente un nouveau comportement qui n'est décrit par aucun des clusters existants. Une approche originale est proposée pour aborder cette problématique.

Le cas où l'essai $\mathbf{E}_6$ est affecté au cluster $\mathbf{C}^1$, contenant les essais $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_4\}$, est considéré. La méthodologie proposée consiste en 4 étapes :

1. Évaluer le prototype (noyau), nommé $\mathbf{G}_1$;
2. Évaluer un vecteur de distances (Équation (3.1) ou Équation (3.4)) entre l'ensemble d'essais $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_4\}$ et le prototype $\mathbf{G}_1$;
3. Ajuster une loi de distribution de probabilités aux distances calculées (Figure 3.4(a)) ;
4. Déterminer un seuil, nommé ds_1, correspondant à 95% de la fonction de répartition (Figure 3.4(b)).

Les prototypes, comme décrit précédemment, peuvent être considérés comme une description de chaque groupe. Dans cette thèse, ils sont calculés en moyennant les représentations symboliques de tous les essais. Si la représentation en intervalles interquartiles est considérée, par exemple, les bornes inférieure et supérieure d'un prototype $\mathbf{G}_i$ sont égales à la moyenne de bornes inférieures et supérieures de tous les essais dans un cluster $\mathbf{C}^i$, respectivement. Dans le cas des histogrammes, le calcul des moyennes est réalisé en tenant compte de chaque catégorie des histogrammes.



(a) Loi de distribution de probabilités log-normale

(b) Fonction de répartition

FIGURE 3.4 – Exemple d'ajustement d'une distribution log-normale hypothétique et détermination du seuil des distances.

De cette façon, après que le nouvel essai soit affecté à un cluster ($\mathbf{C}^1$ dans cet exemple), une nouvelle règle de décision est créée : « Est-ce que $d(\mathbf{E}_6, \mathbf{G}_1) < ds_1$? ». Si la réponse est oui, cela signifie que l'essai $\mathbf{E}_6$ appartient effectivement au cluster $\mathbf{C}^1$ (étant donné que sa distance ne dépasse pas le seuil de « confiance » établi). Si la réponse est non, cela signifie que le nouvel essai représente un nouveau comportement, et il est donc affecté à un nouveau cluster $\mathbf{C}^3$. Il faut tenir compte que les étapes 1-4 peuvent également être appliquées si le nouvel essai est classé dans le cluster $\mathbf{C}^2$. Par ailleurs, cette démarche peut être appliquée pour quel que soit le nombre initial de clusters.

Afin d'ajuster une loi de distribution de probabilités (étape 3), plusieurs lois décrites dans la littérature sont prises en compte et chacune est évaluée en utilisant le test de Kolmogorov-Smirnov. Ce test est basé sur un test d'hypothèses et peut être utilisé pour déterminer si un échantillon de données suit une loi de probabilité donnée. L'hypothèse à valider est appelée *hypothèse nulle* (notée H_0). L'hypothèse pour laquelle les contraintes sont inexactes, est appelé *hypothèse alternative* (notée H_1). Il est noté que rejeter H_0 ne signifie pas nécessairement accepter H_1, puisque ce n'est pas H_1 qui est soumise au test. Dans ce cas, le test d'hypothèse est énoncé

par

$$\begin{cases} H_0 : & \text{L'échantillon de données suit la loi de probabilité donnée} \\ H_1 : & \text{L'échantillon de données ne suit pas la loi de probabilité donnée} \end{cases}$$

Pour chaque test d'hypothèse réalisé, il convient d'évaluer la valeur p^* du test et d'en définir un niveau d'acceptation. La valeur p^* indique avec quelle probabilité l'hypothèse nulle peut être acceptée et est calculée à partir d'un paramètre nommé « statistique de test ». Ce paramètre est défini par l'écart maximal entre la loi de probabilités ajustée et la distribution empirique des données. En d'autres termes, la distribution ayant la « statistique de test » la plus faible est celle la mieux ajustée aux données et donc, utilisée pour la détermination du seuil ds. Le niveau d'acceptation d'un test est la probabilité à partir de laquelle l'hypothèse nulle est rejetée lorsque celle-ci est vraie. Le niveau d'acceptation est choisi de façon plus ou moins arbitraire à une valeur très faible, de l'ordre de 0,1% à 10%. En général, si la valeur p^* est plus importante que le niveau d'acceptation du test, alors il n'y a aucune raison de rejeter l'hypothèse nulle H_0. Toutefois, si cette valeur est plus faible, il faut rejeter H_0. Il est spécifié que tous les tests d'hypothèses réalisés dans le cadre de cette thèse considèrent un niveau d'acceptation de 5% (valeur la plus répandue dans la pratique).

Cinq distributions sont considérées : normale, log-normale, logistique, log-logistique et Weibull. Ces lois ont présenté une bonne aptitude à s'ajuster aux données utilisées et ont été sélectionnées après des tests préliminaires considérant près d'une vingtaine de distributions.

La méthodologie décrite précédemment peut être couplée aux méthodes de clustering basées sur les divisions hiérarchiques et les nuées dynamiques. Ces deux approches sont décrites dans les paragraphes suivants.

3.1.6.2 Affectation basée sur la méthode des divisions hiérarchiques

Cette approche utilise l'arbre généré par la méthode des divisions hiérarchiques. En effet, une fois l'arbre construit, les paires « variable discriminante/valeur de coupure » sont utilisées pour affecter un nouvel essai à l'un des clusters obtenus.

Le partitionnement $\mathbf{P}_2 = \{\mathbf{C}^1, \mathbf{C}^2\}$ et le nouvel essai $\mathbf{E}_6$ sont maintenant considérés. Dans cet exemple, un arbre de décision est crée et la paire Capteur 2/0,035 m$\cdot$ s^{-2} est supposée être la variable la plus discriminante et la valeur de coupure, respectivement.

L'affectation du nouvel essai $\mathbf{E}_6$ est réalisée en comparant une valeur c (correspondante au Capteur 2) de cet essai avec la valeur de coupure évaluée. La valeur c est calculée d'après le type de représentation symbolique utilisée. Si des intervalles interquartiles sont employés, c est calculée par

$$c = \frac{(E_{6,\text{inf}} + E_{6,\text{sup}})}{2}. \tag{3.26}$$

où $E_{6,\text{inf}}$ et $E_{6,\text{sup}}$ sont les bornes inférieure et supérieure de l'intervalle interquartile, respectivement.

Dans le cas où la représentation symbolique est un histogramme, c est considérée comme étant égale à la valeur du centre de gravité de l'histogramme (Équation (3.12)). Ainsi, si la valeur comparative c est plus petite que la valeur de coupure, l'essai $\mathbf{E}_6$ est affecté au cluster $\mathbf{C}^1$. Dans le cas contraire, il est affecté au cluster $\mathbf{C}^2$.

Manifestement, ce type d'affectation oblige la classification d'un nouvel essai à l'un des groupes existants. Toutefois, ainsi comme il a été décrit précédemment, il se peut que cet essai n'appartienne pas au cluster auquel il a été affecté. Afin de traiter ce problème, la méthodologie en 4 étapes du paragraphe 3.1.6.1 doit être considérée.

3.1.6.3 Affectation basée sur la méthode des nuées dynamiques

La deuxième approche diffère de la première dans le sens où elle n'utilise pas des arbres de décision. Les essais sont affectés aux clusters d'après leurs distances par rapport à chaque prototype calculé. Dans le cas où la méthode des nuées dynamiques est appliquée à l'exemple précédent, les prototypes du partitionnement final (également notés $\mathbf{G}_1$ et $\mathbf{G}_2$) sont évalués et les distances par rapport au nouvel essai $\mathbf{E}_6$ sont dénotées par $\varphi(\mathbf{E}_6, \mathbf{G}_1)$ et $\varphi(\mathbf{E}_6, \mathbf{G}_2)$, respectivement.

Considérant les étapes 2-4 de la méthodologie proposée, les critères de décision pour affecter un nouvel essai à l'un de deux clusters sont

$$\text{Si} \begin{cases} \varphi(\mathbf{E}_6, \mathbf{G}_1) < \varphi(\mathbf{E}_6, \mathbf{G}_2) & \& & \varphi(\mathbf{E}_6, \mathbf{G}_1) < ds_1 & \Longrightarrow & \mathbf{E}_6 \in \mathbf{C}^1 \\ \varphi(\mathbf{E}_6, \mathbf{G}_1) < \varphi(\mathbf{E}_6, \mathbf{G}_2) & \& & \varphi(\mathbf{E}_6, \mathbf{G}_1) > ds_1 & \Longrightarrow & \mathbf{E}_6 \in \mathbf{C}^3 \\ \varphi(\mathbf{E}_6, \mathbf{G}_1) > \varphi(\mathbf{E}_6, \mathbf{G}_2) & \& & \varphi(\mathbf{E}_6, \mathbf{G}_2) < ds_2 & \Longrightarrow & \mathbf{E}_6 \in \mathbf{C}^2 \\ \varphi(\mathbf{E}_6, \mathbf{G}_1) > \varphi(\mathbf{E}_6, \mathbf{G}_2) & \& & \varphi(\mathbf{E}_6, \mathbf{G}_2) > ds_2 & \Longrightarrow & \mathbf{E}_6 \in \mathbf{C}^3 \end{cases}$$

où $\mathbf{C}^3$ représente un nouveau cluster et ds_1 et ds_2 sont les seuils correspondants aux clusters $\mathbf{C}^1$ et $\mathbf{C}^2$, respectivement. Par ailleurs, cette méthode peut être appliquée pour n'importe quel nombre de clusters existants.

3.1.7 Bilan

Dans les paragraphes précédents, trois méthodes de partitionnement de données ont été introduites. Les deux premières, basées sur le clustering hiérarchique, permettent à l'utilisateur de visualiser comment les relations entre clusters sont déterminées (à partir des arbres de décision ou des structures pyramidales). La troisième méthode, basée sur le clustering par partition ne donne pas une notion de hiérarchie entre groupes, mais détermine aussi des clusters regroupant les essais les plus similaires, comme dans le cas des deux premières méthodes. L'un des apports majeurs de cette thèse est d'utiliser de façon originale ces techniques pour la détection de modifications structurales. Il est important de noter qu'aucune méthode de clustering ne peut être considérée meilleure que toutes les autres sur l'ensemble de problèmes envisageables. Il faut considérer plutôt que certaines méthodes sont plus adaptées que d'autres dans certains cas. Par ailleurs, il est important d'analyser les résultats obtenus de façon complémentaire et d'observer les éventuelles différences dans les conclusions tirées.

Dans un deuxième temps, des indices mathématiques ont été présentés afin d'apporter des éléments de réponse sur la détermination d'un partitionnement optimal. Ces indices sont importants dans le sens où il n'est pas toujours possible de savoir *a priori* le nombre de comportements structuraux d'un ouvrage. En cela, les trois meilleurs indices existants dans la littérature ont été introduits et sont utilisés dans les applications présentées dans cette thèse.

Finalement, deux approches originales pour l'affectation de nouveaux essais ont été décrites. Ces démarches peuvent être appliquées aux cas de surveillances dynamiques continues d'un ouvrage. À partir de la connaissance des comportements de l'ouvrage, les approches énoncées sont utilisées afin de déterminer si un nouvel enregistrement suit ou non ces comportements.

3.2 Méthodes de classification supervisée

L'objectif des méthodes basées sur la classification supervisée est de construire un modèle (un classificateur) capable de relier une nouvelle observation à une classe connue après avoir observé un certain nombre d'observations. La première étape d'une classification supervisée consiste à

définir des étiquettes connues au préalable pour chaque classe contenant des ensembles d'observations différents. Dans cette étape, souvent nommée *entraînement*, il est nécessaire que l'algorithme de la méthode apprenne les caractéristiques de chaque classe. L'objectif est de permettre, dans une deuxième étape nommée *test*, la généralisation (voire même l'extrapolation) de ses connaissances et l'étiquetage correct de nouvelles observations.

Plusieurs méthodes de classification supervisée existent dans la littérature. Leur utilisation est très vaste dans les domaines du génie civil, de l'économie, de la biologie, etc. Ce fort intérêt est associé a un développement significatif d'algorithmes robustes pour réaliser des études de discrimination et prédiction de données. Dans cette thèse, trois méthodes de classification couplées à l'ADS sont utilisées ayant pour objectif la discrimination de différents comportements structuraux. Dans un premier temps, une méthode basée sur les arbres de décision Bayésiens est présentée. Dans un deuxième temps, une technique d'apprentissage utilisant des réseaux de neurones est détaillée. Finalement, une méthode basée sur les machines à vecteurs supports est introduite. Les simulations avec ces méthodes ont été réalisées sous MatLab®.

3.2.1 Arbres de décision Bayésiens

Les arbres de décision Bayésiens (ADB) consistent en une procédure de décision qui possède une structure similaire à celle décrite pour les divisions hiérarchiques. La méthode utilisée dans cette thèse est essentiellement basée sur celle détaillée par (Diday & Noirhomme-Fraiture 2007). L'extension de cette technique à des variables de type histogramme constitue un aspect original proposé dans le cadre de la thèse.

L'idée générale de cette méthode est de classer une observation inconnue dans l'une des classes définies *a priori* (i.e., dans l'étape d'entraînement). Considérons par exemple l'ensemble d'essais $\mathbf{\Omega} = \{\mathbf{E}_1, \ldots, \mathbf{E}_n\}$ et le paramètre M qui peut prendre les valeurs $\{1, \ldots, r\}$, où r définit le nombre de classes existantes. Cette méthode vise à prédire la valeur de M pour un essai $\widetilde{\mathbf{E}}$ inconnu à partir de ses v variables (capteurs, fréquences propres ou modes propres) et en considérant un sous-ensemble d'essais pour l'étape d'entraînement.

La première étape de cette procédure discriminatoire symbolique consiste à étiqueter les essais dans le groupe d'entraînement et puis, à représenter les classes sous la forme d'un arbre. Une fois l'arbre construit, une règle capable de classer un essai inconnu dans l'un des groupes existants est définie. La description des essais $\mathbf{E}_i,\ (i = 1, \ldots, n)$ est réalisée en considérant

- v variables représentées par des intervalles interquartiles ou des histogrammes ;
- un paramètre M qui spécifie la classe correspondante à chaque essai sous la forme d'une valeur entière (étiquette).

Le Tableau 3.7 détaille un exemple contenant cinq essais où $\mathbf{E}_1$ et $\mathbf{E}_2$ appartiennent à la classe 1 et $\mathbf{E}_3$ et $\mathbf{E}_4$ appartiennent à la classe 2. Ces essais font partie du sous-ensemble d'entraînement, tandis que $\widetilde{\mathbf{E}}$ est l'essai à classer. Dans cet exemple, les essais sont représentés par des signaux enregistrés par le Capteur 1 et le Capteur 2 transformés au préalable en intervalles interquartiles. Il est noté que la même analyse peut être menée avec des fréquences ou des modes propres ou encore en utilisant la transformation en histogrammes.

De façon similaire à celle présentée pour les divisions hiérarchiques, le procédé utilisé pour la construction de l'arbre reprend la notion de *variables discriminantes* et de *questions binaires*. Ces dernières sont formulées selon la loi de Bayes (Diday & Noirhomme-Fraiture 2007) définie par

$$Z = \{\widetilde{\mathbf{E}} | \rho_j f(\widetilde{\mathbf{E}}) \geq \rho_k f(\widetilde{\mathbf{E}})\}, \quad k > j = 1, \ldots, r \tag{3.27}$$

où ρ_j et ρ_k représentent les probabilités *a priori* de chaque classe et $f(.)$ les fonctions de densité de probabilité. Cette loi attribue l'essai $\widetilde{\mathbf{E}}$ à la classe dont la valeur de $\rho_i f(\widetilde{\mathbf{E}})$ est la plus grande.

Essai	Capteur 1 [mm·s^{-2}]	Capteur 2 [mm·s^{-2}]	M
$\mathbf{E}_1$	[410 ; 460]	[165 ; 175]	1
$\mathbf{E}_2$	[390 ; 430]	[150 ; 170]	1
$\mathbf{E}_3$	[130 ; 190]	[107 ; 153]	2
$\mathbf{E}_4$	[410 ; 630]	[140 ; 170]	2
$\widetilde{\mathbf{E}}$	[170 ; 450]	[155 ; 175]	?

TABLEAU 3.7 – Exemple de classification d'essais utilisant les ADB.

Les probabilités *a priori* peuvent être estimées

– selon le nombre de classes :

$$\rho_k = \frac{1}{r}, \quad k = 1, \ldots, r \tag{3.28}$$

– selon le nombre d'essais existants dans chaque classe :

$$\rho_k = \frac{n_k}{n}, \quad n = \sum_{k=1}^{r} n_k \tag{3.29}$$

où n_k est le nombre d'essais dans la classe k. Les fonctions de densité de probabilité sont généralement inconnues et doivent être estimées. Une façon de le faire est d'utiliser la méthode d'estimation par noyau (Silverman 1986) définie par

$$f(\widetilde{\mathbf{E}}) = \frac{1}{n_k} \sum_{i=1}^{n_k} \frac{1}{h} K\left(\frac{\widetilde{\mathbf{E}} - \mathbf{E}_i}{h}\right), \quad k = 1, \ldots, r \tag{3.30}$$

où $K(.)$ est une fonction noyau. Dans cette thèse, la fonction noyau choisie est une adaptation de la fonction normale (Bardos & Rasson 2001) qui s'écrit par

$$K\left(\frac{\widetilde{\mathbf{E}} - \mathbf{E}_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\varphi_q^2(\widetilde{\mathbf{E}}, \mathbf{E}_i)}{2h^2}\right), \quad i = 1, \ldots, n_k \tag{3.31}$$

où h est un paramètre de lissage défini selon la distribution des distances φ_q de la classe k. Une façon répandue d'obtenir la valeur de h est de supposer que l'échantillon des distances est distribué selon une loi paramétrique donnée. En cela, h peut être évalué comme la moyenne de l'intégrale des carrés des erreurs commises par la loi ajustée est la loi donnée. Si les variables sont décrites par des intervalles interquartiles, φ_q est évaluée par l'Équation (3.1). Dans le cas où des histogrammes sont utilisés, cette valeur est calculée à partir de l'Équation (3.4).

De manière générale, l'ADB n'utilise pas toutes les variables pour classer un essai. En effet, la classification d'un essai inconnu est réalisée en ne considérant que la variable la plus discriminante. Pour la déterminer, il est généralement nécessaire de réaliser une procédure itérative. L'exemple avec les essais détaillés dans le Tableau 3.7 est considéré. La variable la plus discriminante est celle qui évalue le mieux possible le fait que les essais $\mathbf{E}_1$ et $\mathbf{E}_2$ appartiennent à la classe 1 et que les essais $\mathbf{E}_3$ et $\mathbf{E}_4$ appartiennent à la classe 2.

L'essai $\mathbf{E}_1$ est pris comme exemple et le « Capteur 1 » est considéré comme un premier candidat à variable discriminante. La procédure commence par analyser les classifications par rapport à la classe 1 [3]. Dans un premier temps, $f(\mathbf{E}_1, \mathbf{E}_2)$ est comparée à $f(\mathbf{E}_1, \mathbf{E}_3)$ puis à $f(\mathbf{E}_1, \mathbf{E}_4)$. Si $f(\mathbf{E}_1, \mathbf{E}_2) > f(\mathbf{E}_1, \mathbf{E}_3)$ et $f(\mathbf{E}_1, \mathbf{E}_2) > f(\mathbf{E}_1, \mathbf{E}_4)$, la variable « Capteur 1 » classe correctement

3. Remarque : étant donné que les probabilités *a priori* sont identiques pour les deux classes, leur notation sera omise.

l'essai connu avec 100% de réussite (2/2). L'essai $\mathbf{E}_2$ est alors pris en compte et $f(\mathbf{E}_2, \mathbf{E}_1)$ est comparée à $f(\mathbf{E}_2, \mathbf{E}_3)$ et $f(\mathbf{E}_2, \mathbf{E}_4)$ en identifiant là aussi le pourcentage de classifications correctes. Inversement, la même démarche est appliquée pour la classe 2 où $f(\mathbf{E}_3, \mathbf{E}_4)$ est comparée à $f(\mathbf{E}_3, \mathbf{E}_1)$ et $f(\mathbf{E}_3, \mathbf{E}_2)$ aussi bien que $f(\mathbf{E}_4, \mathbf{E}_3)$ est comparée à $f(\mathbf{E}_4, \mathbf{E}_1)$ et $f(\mathbf{E}_4, \mathbf{E}_2)$. Il est important de noter que tous les essais sont comparés deux-à-deux et ne sont jamais comparés à eux mêmes.

La même procédure est également menée pour les autres variables. La variable capable de classer correctement les essais le plus grand nombre de fois, est considérée comme la variable discriminante et est utilisée pour la classification des essais inconnus. Toutefois, si le même nombre de classifications correctes est obtenu pour deux variables ou plus, celle qui a produit les plus grandes différences entres les fonctions $f(.)$, est utilisée.

Afin d'illustrer cette démarche, les données du Tableau 3.7 sont utilisées. Dans un premier temps, les distances de Hausdorff normalisées pour les capteurs 1 et 2 sont calculées par

$$\varphi_1 = \begin{array}{c} \begin{matrix} \mathbf{E}_1 & \mathbf{E}_2 & \mathbf{E}_3 & \mathbf{E}_4 \end{matrix} \\ \begin{pmatrix} 0 & 0,05 & 0,24 & 0,28 \\ 0,05 & 0 & 0,15 & 0,19 \\ 0,24 & 0,15 & 0 & 0,18 \\ 0,28 & 0,19 & 0,18 & 0 \end{pmatrix} \end{array} \begin{matrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \end{matrix} \qquad \varphi_2 = \begin{array}{c} \begin{matrix} \mathbf{E}_1 & \mathbf{E}_2 & \mathbf{E}_3 & \mathbf{E}_4 \end{matrix} \\ \begin{pmatrix} 0 & 0,26 & 0,33 & 0,51 \\ 0,26 & 0 & 0,44 & 0,38 \\ 0,33 & 0,44 & 0 & 0,21 \\ 0,51 & 0,38 & 0,21 & 0 \end{pmatrix} \end{array} \begin{matrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \end{matrix} \tag{3.32}$$

Dans un deuxième temps, les fonctions de densité sont évaluées à partir des Équations (3.30)-(3.31) pour chaque paire d'essais, comme détaillé dans les Tableaux 3.8 et 3.9.

Essai	$f(\mathbf{E}_1, \mathbf{E}_2)$	$f(\mathbf{E}_1, \mathbf{E}_3)$	$f(\mathbf{E}_1, \mathbf{E}_4)$	$f(\mathbf{E}_2, \mathbf{E}_3)$	$f(\mathbf{E}_2, \mathbf{E}_4)$	$f(\mathbf{E}_3, \mathbf{E}_4)$
$\mathbf{E}_1$	0,0324	0	0	-	-	-
$\mathbf{E}_2$	0,0324	-	-	0	0	-
$\mathbf{E}_3$	-	0	-	0	-	0,0165
$\mathbf{E}_4$	-	-	0	-	0	0,0165

TABLEAU 3.8 – Calcul de fonctions de densité considérant le Capteur 1.

Essai	$f(\mathbf{E}_1, \mathbf{E}_2)$	$f(\mathbf{E}_1, \mathbf{E}_3)$	$f(\mathbf{E}_1, \mathbf{E}_4)$	$f(\mathbf{E}_2, \mathbf{E}_3)$	$f(\mathbf{E}_2, \mathbf{E}_4)$	$f(\mathbf{E}_3, \mathbf{E}_4)$
$\mathbf{E}_1$	0,226	0,004	0,001	-	-	-
$\mathbf{E}_2$	0,226	-	-	0	0	-
$\mathbf{E}_3$	-	0,004	-	0	-	0,309
$\mathbf{E}_4$	-	-	0,001	-	0	0,309

TABLEAU 3.9 – Calcul de fonctions de densité considérant le Capteur 2.

Il est noté que les deux capteurs sont capables de classer tous les essais avec 100% de réussite. Considérant les probabilités $\rho_1 = \rho_2 = 0,5$, il suffit de regarder les différences entre les valeurs des fonctions de densité calculées afin de choisir le capteur le plus discriminant. Dans cet exemple, la variable la plus discriminante est le Capteur 2 (étant donné que les différences entre les fonctions f sont de l'ordre de 10^{-1}, tandis qu'elles ne le sont que de 10^{-2} pour le Capteur 1). La classification de l'essai $\widetilde{\mathbf{E}}$ est alors réalisée en utilisant uniquement ce capteur.

Bien que l'exemple considéré ne distingue que deux classes, le principe des ADB peut facilement être étendu à des problèmes plus complexes. En effet, afin de classer un essai dans l'une des r classes possibles, il est nécessaire d'évaluer $r - 1$ questions logiques comme illustré à la Figure 3.5.

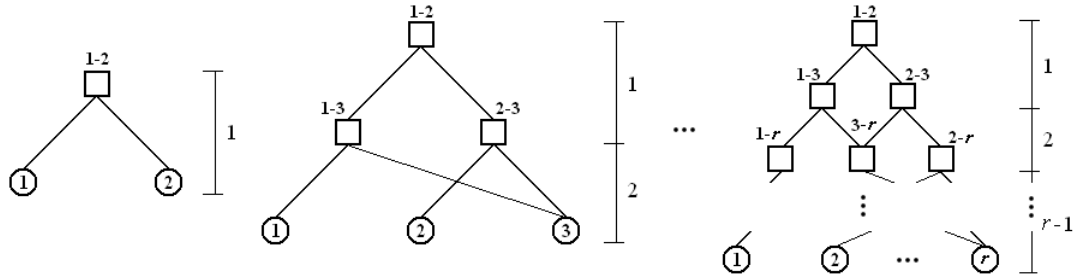


FIGURE 3.5 – Schéma simplifié des ADB avec 2, 3 et r classes.

3.2.2 Réseaux de neurones

De façon générale, des études sur les réseaux de neurones ont débuté en 1943 avec les travaux de McCulloch et Pitts (McCulloch & Pitts 1943). Cependant, la notion d'apprentissage n'est introduite qu'en 1958 par Rosenblatt (Rosenblatt 1958). Quelques années plus tard, Hopfield propose des réseaux associatifs à plusieurs couches qui ont montré être capables de résoudre des problèmes plus complexes (Hopfield 1982). La dernière amélioration a été apportée par (Rumelhart et al. 1986) qui ont proposé un algorithme basé sur la rétropropagation de l'erreur, permettant d'optimiser les paramètres d'un réseau de neurones. De nos jours, cette méthode est appliquée dans de nombreuses études dans le domaine scientifique et aussi industriel.

Un réseau de neurones est un modèle de calcul qui s'inspire du principe de fonctionnement du cerveau humain, en tenant compte de leurs liaisons synaptiques et de leurs règles de propagation du signal électrique. Ainsi, l'objectif des réseaux de neurones artificiels est de non seulement modéliser le fonctionnement des neurones réels, mais aussi de permettre l'apprentissage de nouvelles règles d'association, de décision, de prédiction, etc.

Cette technique a démontre sa capacité à résoudre des problèmes sur la classification linéaire et non linéaire. Généralement, les réseaux de neurones sont « ajustés » ou « entraînés » de sorte qu'une observation connue en entrée mène à un résultat spécifique en sortie. En effet, ceci est l'idée générale des méthodes supervisées en définissant des étiquettes pour chaque donnée d'entrée. Par ailleurs, l'entraînement du réseau est basé sur une comparaison itérative entre sa sortie et l'étiquette donnée jusqu'à ce que le réseau donne le résultat souhaité. Le but de l'entraînement est de permettre au réseau de neurones « d'apprendre » des règles à partir des observations initiales. Si l'entraînement est correctement réalisé, le réseau est capable de fournir des réponses en sortie très proches (voire même identiques) des étiquettes du jeu de données d'entraînement. Ce principe est illustré par la Fig 3.6.

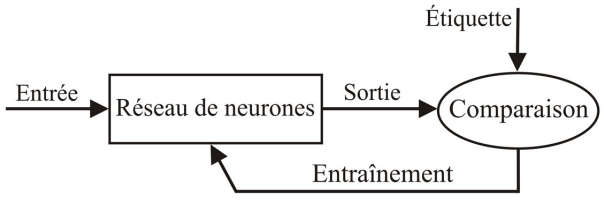


FIGURE 3.6 – Schéma d'entraînement d'un réseau de neurones.

3.2.2.1 Modèle mathématique d'un neurone

Le fonctionnement des neurones biologiques peut être décrit à partir de modèles mathématiques plus ou moins bien ajustés. Le plus connu a été proposé par (McCulloch & Pitts 1943).

Dans ce modèle, quatre éléments basiques sont définis : les données d'entrée, représentées par les essais $\mathbf{E}_i, (i = 1, \ldots, n)$, un biais β_j, les poids synaptiques w_{jg} et les sorties y_j où j indique le neurone étudié et g est l'indice qui relie une entrée à un poids. La première étape, définie par la *fonction de combinaison* Σ consiste à sommer les données d'entrée multipliées par les poids synaptiques. Cette étape est suivie d'une autre définie par une *fonction d'activation* ψ qui définit, à partir du résultat de l'étape précédente, la sortie du neurone. Cette fonction a pour but de vérifier si la somme pondérée des données d'entrée dépasse un certain seuil ou non. Si oui, la sortie du neurone n'est pas propagée dans le réseau. La Figure 3.7 illustre le modèle proposé par McCulloch et Pitts.

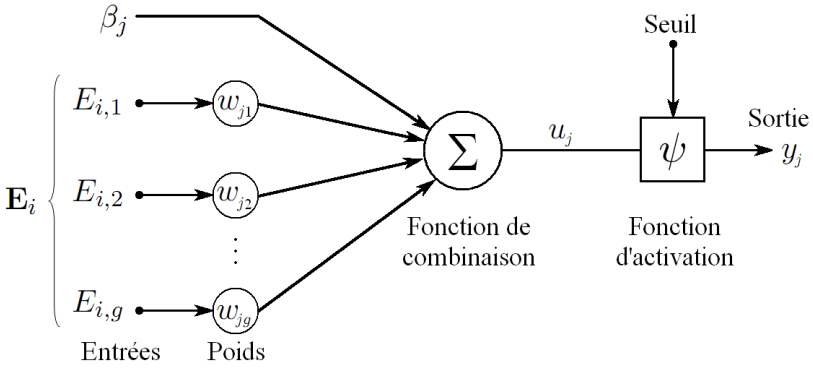


FIGURE 3.7 – Modèle mathématique d'un neurone artificiel.

Dans la Figure 3.7, $(E_{i,1}, E_{i,2}, \ldots, E_{i,g})$ dénotent les entrées d'un neurone, où g est le nombre de caractéristiques (*features*, en anglais) de chaque essai. Dans le cadre de cette thèse, les données d'entrée sont des représentations symboliques des essais dynamiques décrits par des signaux ou des paramètres modaux. Ces quantités sont transformées soit en intervalles interquartiles, soit en histogrammes à 20 catégories. Le nombre de caractéristiques g est différent selon la représentation symbolique utilisée.

Soient les essais $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_n\}$ transformés en intervalles interquartiles. Une matrice symbolique peut être définie par

$$\mathbf{E} = \begin{array}{c} \begin{matrix} \text{Variable 1} & & & \text{Variable } v & \end{matrix} \\ \begin{pmatrix} E^1_{1,\text{inf}} & E^1_{1,\text{sup}} & \cdots & E^v_{1,\text{inf}} & E^v_{1,\text{sup}} \\ E^1_{2,\text{inf}} & E^1_{2,\text{sup}} & \cdots & E^v_{2,\text{inf}} & E^v_{2,\text{sup}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ E^1_{n,\text{inf}} & E^1_{n,\text{sup}} & \cdots & E^v_{n,\text{inf}} & E^v_{n,\text{sup}} \end{pmatrix} \end{array} \begin{matrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_n \end{matrix} \tag{3.33}$$

où v est le nombre de variables (capteurs, fréquences propres ou modes propres), n le nombre d'essais et $(E^1_{1,\text{inf}}; E^1_{1,\text{sup}})$ les bornes inférieure et supérieure de l'intervalle interquartile pour la première variable du premier essai. Dans cette transformation, le nombre de caractéristiques $g = 2v$, puisque chaque variable de l'essai $\mathbf{E}_i$ est représenté par une paire de valeurs (bornes inférieure et supérieure des intervalles). En utilisant l'exemple du Tableau 3.1, la matrice $\mathbf{E}$ est donnée par

$$\mathbf{E} = \begin{array}{c} \begin{matrix} \text{Capteur 1} & & \vert & \text{Capteur 2} & \end{matrix} \\ \left(\begin{array}{cc|cc} 410 & 460 & 165 & 175 \\ 390 & 430 & 150 & 170 \\ 130 & 190 & 107 & 153 \\ 410 & 630 & 140 & 170 \\ 170 & 450 & 155 & 175 \end{array}\right) \end{array} \begin{matrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \\ \mathbf{E}_5 \end{matrix} \tag{3.34}$$

Dans cet exemple, il existe quatre caractéristiques. De façon similaire, la matrice symbolique utilisant des histogrammes est assemblée à partir des proportions $b_i^{q,k}$ évaluées pour chaque catégorie d'une variable correspondante à un essai

$$\mathbf{E} = \begin{array}{c} \quad\text{Variable 1} \qquad\qquad\qquad\qquad\qquad\qquad\text{Variable } v \\ \begin{pmatrix} b_1^{1,1} & b_1^{1,2} & \dots & b_1^{1,m} & \dots & b_1^{v,1} & b_1^{v,2} & \dots & b_1^{v,m} \\ b_2^{1,1} & b_2^{1,2} & \dots & b_2^{1,m} & \dots & b_2^{v,1} & b_2^{v,2} & \dots & b_2^{v,m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ b_n^{1,1} & b_n^{1,2} & \dots & b_n^{1,m} & \dots & b_n^{v,1} & b_n^{v,2} & \dots & b_n^{v,m} \end{pmatrix} \begin{array}{c} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_n \end{array} \end{array} \tag{3.35}$$

où m est le nombre de catégories de l'histogramme. Dans cette transformation, le nombre de caractéristiques $g = 20v$, puisque chaque variable de l'essai $\mathbf{E}_i$ est représenté par vingt valeurs (catégories de l'histogramme). Par exemple, si deux essais sont décrits par

$$\begin{aligned} \mathbf{E}_1 &= \{\overbrace{[1(0,01),\dots,10(0,04),\dots,20(0,001)]}^{\text{Fréquence 1}};\overbrace{[1(0,1),\dots,10(0,40),\dots,20(0,09)]}^{\text{Fréquence 2}}\} \\ \mathbf{E}_2 &= \{[1(0,02),\dots,10(0,58),\dots,20(0,005)];[1(0,2),\dots,10(0,12),\dots,20(0,04)]\} \end{aligned} \tag{3.36}$$

La matrice symbolique est construite de la manière suivante

$$\mathbf{E} = \begin{array}{c} \quad\text{Fréquence 1} \qquad\qquad\qquad\quad\text{Fréquence 2} \\ \left(\begin{array}{cccccc|cccccc} 0,01 & \dots & 0,04 & \dots & 0,001 & | & 0,1 & \dots & 0,40 & \dots & 0,09 \\ 0,02 & \dots & 0,58 & \dots & 0,005 & | & 0,2 & \dots & 0,12 & \dots & 0,04 \end{array}\right) \begin{array}{c} \mathbf{E}_1 \\ \mathbf{E}_2 \end{array} \end{array} \tag{3.37}$$

Mathématiquement, la sortie d'un neurone j peut être représenté par

$$u_j = \sum_{k=1}^{g} w_{jk} \cdot E_{i,k} + \beta_j, \quad i = 1,\dots,n \tag{3.38}$$

où g est le nombre de catégories et n le nombre d'essais. La sortie y_j est obtenue en appliquant la fonction d'activation ψ sur u_j

$$y_j = \psi(u_j) \tag{3.39}$$

En général, les fonctions d'activation prennent les valeurs calculées par la fonction de combinaison et leur attribuent des valeurs binaires (0 et 1 ou même -1 et 1) ou des valeurs continues. Il en existe plusieurs types, mais les plus utilisées sont (Bishop 1995)

- la fonction Heaviside, définie par

$$\psi(x) = \begin{cases} 0 & \text{si} \quad x < 0, \quad \forall x \in \Re \\ 1 & \text{si} \quad x \geq 0, \quad \forall x \in \Re \end{cases} \tag{3.40}$$

  qui est discontinue et prend la valeur 0 quand la somme pondérée est négative et la valeur 1 partout ailleurs. Généralement, cette fonction est utilisée avec des réseaux plus simples et/ou appliquée à des problèmes de classification.
- la fonction sigmoïde écrite par

$$\psi(x) = \frac{1}{1+e^{-\lambda x}}; \quad \psi(x) \in [0,1], \quad \forall x \in \Re \tag{3.41}$$

  où λ est un paramètre qui indique l'inclination de la fonction. Cette fonction est très utilisée, car elle est infiniment dérivable et non polynomiale. Cela permet une meilleure adaptation à des problèmes plus complexes (régression non linéaire, par exemple).

– la fonction tangente hyperbolique donnée par

$$\psi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}; \quad \psi(x) \in [-1, 1], \quad \forall x \in \Re \tag{3.42}$$

est aussi largement utilisée dans des problèmes de régression, étant donné qu'elle est aussi dérivable et non polynomiale.

Dans cette thèse, les trois fonctions d'activation sont utilisées, étant donné leur bonne adaptation aux problèmes traités (Cury & Crémona 2010*c*). La fonction d'activation Heaviside est plus souvent utilisée dans le cadre des problèmes de classification. Les autres deux fonctions sont généralement employées dans des problèmes de régression (non linéaire, par exemple). Bien que les valeurs de sortie soient comprises entre des intervalles précis (0 et 1 ou -1 et 1), il est toujours possible d'utiliser des manipulations matricielles permettant de les remettre aux échelles originales. Cette manipulation est expliquée dans le paragraphe 3.4.

3.2.2.2 Réseau de neurones artificiel

L'assemblage des neurones énoncés précédemment forme le réseau de neurones. Schématiquement il est représenté par un graphe orienté où les neurones et les synapses (les connexions) sont symbolisés par des cercles et des flèches, respectivement (Fig 3.8). Le réseau utilisé a une architecture composée de plusieurs neurones et de différentes couches. Ce type de réseau, souvent nommé MLP (*Multilayer Perceptron*) est constitué d'une couche d'entrées, d'une couche de sorties et d'un nombre variable de couches cachées (qui dépend de la complexité du problème). En effet, c'est l'existence des couches cachées qui permet au réseau de neurones de résoudre des problèmes non linéaires.

Par ailleurs, il y a trois topologies de réseaux MLP qui déterminent comment l'information est transmise d'un neurone à un autre. La première, utilisée dans cette thèse, est la propagation vers l'avant (*feed-forward propagation*). Dans ce cas, chaque neurone d'une couche se connecte avec tous les autres de la couche suivante. Toutefois, il n'existe pas de connections entre neurones d'une même couche. L'information provient de la couche d'entrées et ensuite est transmise aux couches cachées, qui finalement la transmettent à la couche de sorties. La Fig 3.8 illustre cette classe de MLP avec deux couches cachées.

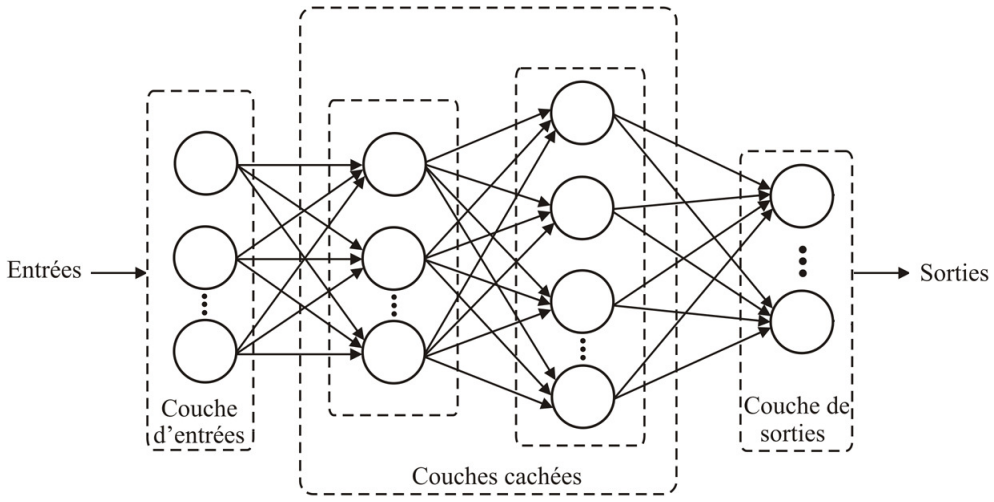


FIGURE 3.8 – Réseau de neurones du type MLP.

Les deux autres topologies de MLP considèrent une propagation vers l'arrière (*feed-backward propagation*) où les neurones se communiquent entre eux dans tous les sens, et une propagation

concurrente, où les meilleurs neurones d'une couche sont choisis (d'après un critère pré-défini) pour transmettre l'information. Un exemple de cette dernière topologie est le réseau de Kohonen, aussi connu sous le nom de cartes auto-organisatrices.

L'apprentissage d'un réseau de neurones MLP est un procédé itératif, souvent appelé « apprentissage par expérience ». Le principal algorithme utilisé pour l'entraînement d'un réseau est basé sur la technique de rétropropagation du gradient ou propagation de l'erreur (*backpropagation*). Cette technique a pour objectif de trouver les poids des connexions qui minimisent l'erreur quadratique moyenne (EQM) commise par le réseau sur l'ensemble de l'apprentissage. De façon itérative, les valeurs de sortie du réseau sont comparées aux étiquettes définies au préalable et les différences sont évaluées par une fonction erreur (comme l'EQM, par exemple). Ensuite, l'erreur est rétropropagée dans le réseau et l'algorithme ajuste les poids de chaque connexion inter-neurones afin de diminuer les différences entre les valeurs de sortie et les étiquettes. Après avoir répété cette procédure un certain nombre de fois, le réseau généralement converge vers un certain état où l'erreur des calculs est négligeable. Dans ce cas, le réseau est considéré « entraîné ».

Afin de calibrer les poids, une méthode d'optimisation non linéaire nommée « règle de delta modifiée » est utilisée. Cette méthode consiste à évaluer la dérivée de la fonction erreur par rapport aux poids du réseau. Les poids sont donc actualisés de sorte que l'erreur diminue (représentée par la descente dans la surface d'erreur). Cette règle est une modification de la règle de delta et a pour caractéristique l'ajout d'un terme à expression originale qui permet de ne pas converger vers un minimum local. Son efficacité dépend, en effet, de paramètres qui doivent être fixés au préalable, comme le pas du gradient et le nombre d'itérations, par exemple.

Les étapes de la procédure de rétropropagation sont détaillées dans l'Algorithme 2 (Principe et al. 2000). Chacune de ces étapes est discutée dans la suite.

Entrées : Données d'entrées, étiquettes, nombre d'itérations, critère d'arrêt
Sorties : Poids synaptiques optimaux du réseau entraîné
1 Initialisation aléatoire des poids synaptiques
2 **répéter**
3 Calcul des valeurs de sortie des neurones de la couche cachée ;
4 Calcul des valeurs de sortie des neurones de la couche de sortie (réponse réelle du réseau) ;
5 Calcul de l'erreur (différence entre la réponse du réseau et celle espérée) ;
6 Actualisation des poids ;
7 **jusqu'à** *critère d'arrêt atteint*;

Algorithme 2: Pseudo-algorithme de la technique de rétropropagation.

Ligne 1 : L'initialisation des poids synaptiques est généralement réalisée de manière aléatoire. Normalement, les valeurs ne sont pas trop importantes et varient entre 0 et 1.

Lignes 3,4 : Les valeurs de sortie sont obtenues à partir de l'application de la somme u_j à une fonction d'activation $\psi(.)$. Mathématiquement, pour un neurone j dans une couche l et à la $t^{\text{ème}}$ itération

$$u_j^{(l)}(t) = \sum_{k=1}^{g+1} w_{jk}^{(l)}(t) z_k^{(l-1)}(t) \tag{3.43}$$

où $z_k^{(l-1)}(t)$ est la de sortie du neurone k de la couche précédente $(l-1)$ à la $t^{\text{ème}}$ itération et $w_{jk}^{(l)}$ le poids du neurone j de la couche (l) (qui est alimenté par le neurone k de la couche $(l-1)$). Il est noté que $g+1$ est utilisé au lieu de g dans l'Équation (3.43). Cela indique que le biais β_j est déjà pris en compte dans le calcul de la somme. Considérant que $\psi(u_j)$

est utilisé pour dénoter la fonction d'activation d'un neurone j, la sortie d'une couche (l) est

$$y_j^{(l)} = \psi(u_j(t)) \tag{3.44}$$

Ligne 5 : L'erreur de la $n^{\text{ème}}$ sortie du réseau à la $t^{\text{ème}}$ itération est

$$e_n(t) = c_n(t) - y_n(t) \tag{3.45}$$

où c_n est la $n^{\text{ème}}$ sortie souhaitée [4] et y_n la sortie du réseau.

Ligne 6 : L'actualisation des poids est faite entre la(les) couche(s) cachée(s) et la couche d'entrée. L'expression qui permet de relier ces quantités est

$$w_{jk}^{(l)}(t) = w_{jk}^{(l)}(t-1) + \Delta w_{jk}^{(l)}(t) \tag{3.46}$$

où $\Delta w_{jk}^{(l)}(t)$ est la correction appliquée aux poids. Comme énoncé auparavant, cette correction est déterminée par la règle de delta modifiée, définie par

$$\Delta w_{jk}^{(l)}(t+1) = \eta \delta_j^{(l)} y_k^{(l-1)} + \mu \Delta w_{jk}^{(l)}(t) \tag{3.47}$$

où η et μ sont des paramètres nommés *taux d'apprentissage* et *constante de moment*, respectivement ; $\delta_j^{(l)}$ est le gradient local, et $y_k^{(l-1)}$ la sortie du neurone k dans la couche $l-1$. Le gradient local est calculé de manière distincte pour les neurones de la couche cachée et de sortie. En considérant L la couche de sortie, le gradient d'un neurone j à la $t^{\text{ème}}$ itération est

$$\delta_j^{(L)}(t) = e_j^{(L)}(t)\psi'(u_j^{(L)}) \tag{3.48}$$

où $\psi'(.)$ est la dérivée de la fonction d'activation. Pour les neurones d'une couche cachée, l'expression du gradient est

$$\delta_j^{(l)}(t) = \psi'(u_j^{(l)}) \left(\sum_{i=k}^{g+1} \delta_k^{(l+1)} w_{jk}^{(l+1)} \right) \tag{3.49}$$

Le taux d'apprentissage η définit la vitesse de descente du gradient et est initialement choisi de façon arbitraire. La constante de moment μ est utilisée d'une part pour éviter de grandes oscillations pour des valeurs de η trop importantes et d'autre part, pour éviter une convergence très lente pour des valeurs de η trop faibles.

Ligne 7 : Le critère d'arrêt est généralement défini par un nombre maximal d'itérations ou par une tolérance d'erreur maximale ou encore basé sur les résultats de la validation croisée des données d'entrée. Si cette dernière est utilisée, l'ensemble de données étiquetées est divisé en deux parties : une pour l'entraînement du réseau et l'autre pour la validation. Le but est d'observer la performance du réseau à classer des observations connues avant de l'utiliser pour la classification de données inconnues. Le critère d'arrêt est défini par l'itération à partir de laquelle le réseau commence à perdre sa capacité à valider les données, comme illustré à la Figure 3.9.

L'un des grands avantages des réseaux de neurones est lié à leur capacité à s'adapter à des problèmes de nature complexe. La principale difficulté est de savoir choisir les paramètres nécessaires comme le taux d'apprentissage, la constante de moment, le nombre de couches cachées, etc.

4. Dans les problèmes de classification, les sorties souhaitées sont les étiquettes de chaque classe. Dans le cas des problèmes de régression, ces sorties sont les observations d'une quantité donnée (e.g., fréquences propres, composantes modales, etc.).

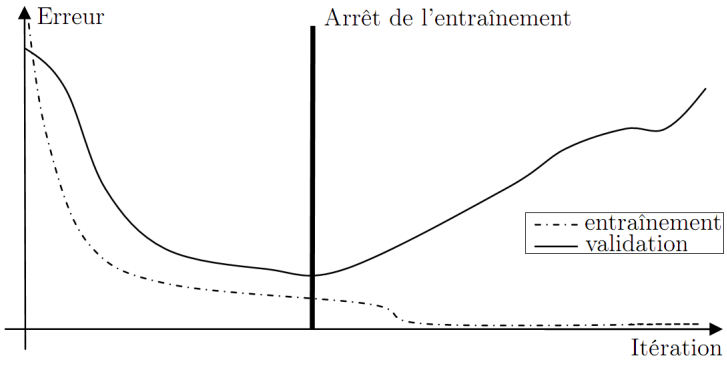


FIGURE 3.9 – Choix du critère d'arrêt pour l'entraînement du réseau de neurones.

Comme décrit précédemment, les entrées du réseau de neurones sont les matrices symboliques des essais et les sorties, les étiquettes correspondantes à chaque classe. Dans cette thèse, les simulations réalisées avec cette technique sont menées à l'aide de la boîte à outils gratuite Netlab développée par Bishop (Bishop 1995). Les paramètres utilisés sont différents pour chaque application et sont détaillés dans les chapitres d'applications numériques et expérimentales.

3.2.3 Machines à vecteurs support

Une machine à vecteurs supports (Support Vector Machine ou SVM) est une méthode de classification supervisée, introduite par V. Vapnik (Vapnik 1995). Elle repose sur l'existence d'un classificateur dans un espace approprié, le but étant de séparer des données en deux classes distinctes. Pour cela, il est nécessaire d'utiliser un jeu de données d'entraînement pour apprendre les paramètres du modèle. Dans cette section, le principe de fonctionnement général des SVM est mis en évidence, suivi d'une étude de cas où les données d'entrées à classifier ne sont pas linéairement séparables (cas le plus répandu dans la pratique).

3.2.3.1 Principe des SVM

Considérons un ensemble de données d'entraînement représentées dans un espace d'une certaine dimension et distribuées selon deux classes. Le but des SVM est de trouver un séparateur linéaire ou non entre ces deux classes de données qui maximise la distance entre ces dernières. Dans le cas des SVM, ce séparateur est souvent appelé *hyperplan*. Un hyperplan est une généralisation à v dimensions (v étant le nombre de variables) d'une ligne (1 dimension) séparant un espace à 2 dimensions, ou d'un plan (2 dimensions) séparant un espace à 3 dimensions. Une fois cet hyperplan déterminé, il est utilisé pour déterminer la classe à laquelle appartient une donnée différente de celles de l'ensemble d'entraînement. Les points des deux classes les plus proches de l'hyperplan les séparant sont appelés *vecteurs supports*, comme illustré à la Figure 3.10.

Il est manifeste qu'il existe une infinité d'hyperplans valides séparant deux classes, mais l'objectif des SVM est de déterminer l'hyperplan optimal. Cela revient à chercher un hyperplan valide qui passe entre les points des deux classes de données et qui maximise la distance minimale entre ces dernières. La distance entre la donnée la plus proche et l'hyperplan est appelée *marge*. L'hyperplan optimal est donc celui qui maximise la marge.

Le fait d'avoir une marge la plus large possible correspond à plus de « sécurité » et moins de risque lorsque l'objectif est classer une donnée inconnue. Cette remarque est illustré à la Figure 3.11. La Figure 3.11(a) montre un hyperplan qui possède une faible marge de séparation, tandis que la Figure 3.11(b) illustre un hyperplan optimal. La classification d'une nouvelle

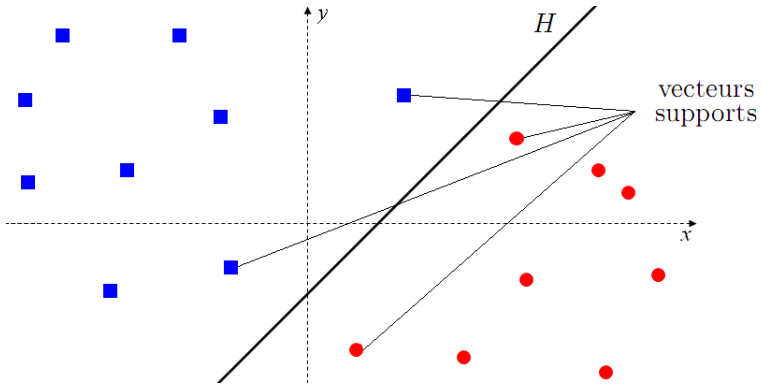
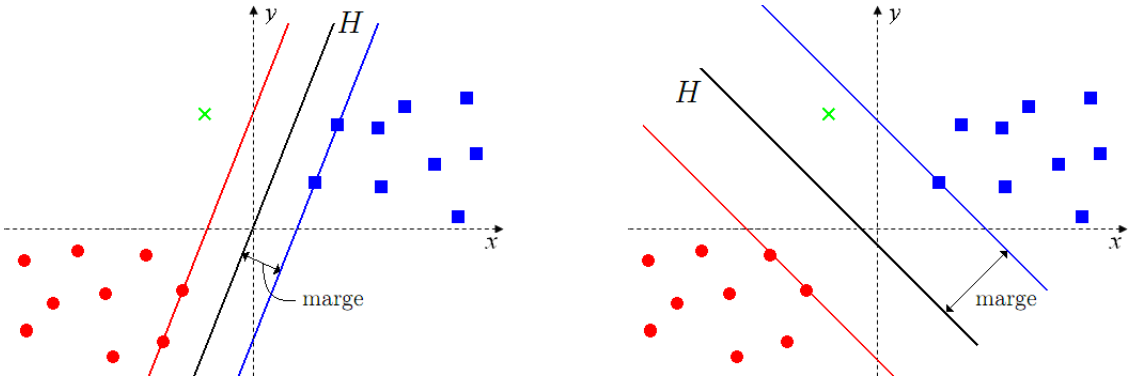


FIGURE 3.10 – Séparation de deux classes de données par un hyperplan H.

observation se fait suivant sa position par rapport à l'hyperplan optimal. Dans cet exemple, l'observation représentée par une croix est classé incorrectement dans la catégorie des « ronds », si l'hyperplan avec faible marge est considéré. Toutefois, si l'hyperplan optimal est pris en compte, la nouvelle observation est classée correctement dans la catégorie des « carrés ».



(a) Hyperplan avec faible marge.

(b) Hyperplan optimal.

FIGURE 3.11 – Définition de l'hyperplan optimal.

3.2.3.2 Classifications linéaire et non linéaire

Il existe deux types de problèmes de classification utilisant des SVM. Le premier est celui où les données d'entraînement sont linéairement séparables dans leur espace de représentation et le second où elles ne le sont pas. Le premier cas est relativement simple à résoudre, mais dans la plupart des problèmes réels, les données ne sont pas linéairement séparables et il n'est donc pas possible de déterminer un hyperplan pour les classifier.

Afin de contourner cette limitation, l'idée de l'algorithme des SVM est de changer l'espace original de représentation des données. Ce nouvel espace, de dimension supérieure à celui d'origine, est appelé *espace de re-description* et est obtenu à partir d'une fonction de transformation non linéaire appelé *fonction noyau*.

Cas linéaire

Un hyperplan, étant un séparateur linéaire, peut être formalisé mathématiquement par

$$H(\hat{\mathbf{E}}) = \hat{\mathbf{E}}\mathbf{w} + \beta \tag{3.50}$$

où $\hat{\mathbf{E}}$ est un essai hypothétique (un vecteur représenté par des intervalles interquartiles ou des histogrammes), $\mathbf{w}$ le vecteur des poids qui définissent l'hyperplan, et β un biais qui permet d'augmenter ou diminuer la marge. Le biais a pour objectif d'étendre la recherche de l'hyperplan optimal. Sans ce terme, les hyperplans contiennent nécessairement l'origine de l'espace de description, ce qui restreint la solution au problème.

Un ensemble de données d'entraînement

$$\{(\mathbf{E}_1, c_1), (\mathbf{E}_2, c_2), ..., (\mathbf{E}_n, c_n)\} \tag{3.51}$$

et d'étiquettes c_i prenant les valeurs 1 ou -1, dénotant les classes aux lesquelles appartiennent les essais $\mathbf{E}_i$ sont maintenant considérés. Supposons que ces essais soient linéairement séparables dans leur espace d'origine. Dans ce cas, il est possible d'évaluer une paire $(\mathbf{w}, \beta)$ de telle sorte que

$$\begin{cases} \mathbf{E}_i\mathbf{w} + \beta \geq 1, & \forall \mathbf{E}_i | c_i = 1 \\ \mathbf{E}_i\mathbf{w} + \beta \leq -1, & \forall \mathbf{E}_i | c_i = -1 \end{cases} \tag{3.52}$$

Dans la recherche de l'hyperplan optimal, deux hyperplans parallèles sont construits contenant chacun un certain nombre d'essais d'une des deux classes (les vecteurs supports). Il est démontré que la distance entre ces hyperplans est égale à $\frac{2}{||\mathbf{w}||}$ (Burbridge & Buxton 2002). L'objectif est de maximiser cette distance et donc la marge. De plus, il est nécessaire de s'assurer qu'aucun essai ne se trouve entre les hyperplans, ce qui revient à minimiser $||\mathbf{w}||^2$ sous certaines contraintes. Le problème d'optimisation portant sur les paramètres $\mathbf{w}$ et β est défini par (Cristianini & Shawe-Taylor 2000)

$$\begin{cases} \underset{\mathbf{w},\beta}{\text{Min}} & \left(\frac{1}{2}||\mathbf{w}||^2\right) \\ \text{sous les contraintes} & c_i(\mathbf{E}_i\mathbf{w} + \beta) \geq 1, \ 1 \leq i \leq n \end{cases} \tag{3.53}$$

où n est le nombre d'essais.

L'Équation (3.53) est appelée *formulation primale* et implique la détermination de $v + 1$ paramètres, où v est la dimension de l'espace des données d'entraînement. Ce problème peut être aisément résolu pour des valeurs de dimensions petites, mais devient beaucoup plus complexe quand la valeur de v dépasse quelques centaines. En effet, il existe une transformation de ce problème nommée *formulation duale* donnée par

$$\begin{cases} \underset{\alpha}{\text{Max}} & (\tilde{L}(\alpha)) \\ \text{avec} & \tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j c_i c_j \langle \mathbf{E}_i \cdot \mathbf{E}_j \rangle \\ \text{sous les contraintes} & \alpha_i \geq 0, 1 \leq i \leq n \quad \text{et} \quad \sum_{i=1}^{n} \alpha_i c_i = 0 \end{cases} \tag{3.54}$$

où $\langle \mathbf{E}_i \cdot \mathbf{E}_j \rangle$ représente le produit scalaire entre les essais $\mathbf{E}_i$ et $\mathbf{E}_j$, par variable. Cette formulation consiste à résoudre un problème quadratique où le Lagragian $\tilde{L}$ est maximisé par rapport aux coefficients α_i (appelés *multiplicateurs de Lagrange*) et minimisé par rapport à $\mathbf{w}$ et β. Dans ce cas, l'hyperplan solution correspondant est

$$H(\mathbf{E}_i, \mathbf{E}_j) = \sum_{i,j=1}^{n} \alpha_i^* c_i . \langle \mathbf{E}_i \cdot \mathbf{E}_j \rangle + \beta^* \tag{3.55}$$

où les α_i^* sont les solutions optimales du problème dual, $\mathbf{w}^*$ et β^* sont les poids et biais optimaux, respectivement. Ces derniers sont calculés par

$$\begin{aligned} \mathbf{w}^* &= \sum_{i=1}^{n} \alpha_i^* c_i \mathbf{E}_i \\ \beta^* &= c_i - \mathbf{E}_i\mathbf{w}^*, \quad \forall \mathbf{E}_i \end{aligned} \tag{3.56}$$

Il est important de noter que les multiplicateurs de Lagrange sont égaux à zéro pour les essais qui se trouvent sur la marge, ce qui les identifie comme des vecteurs supports. La Figure 3.12 illustre les éléments discutés auparavant.

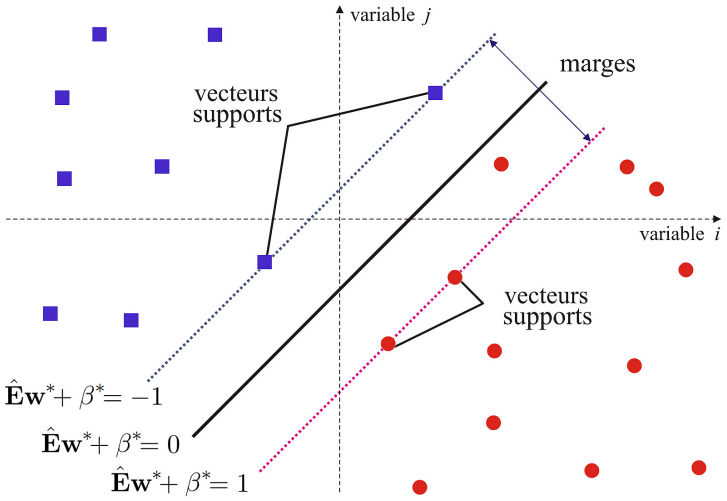


FIGURE 3.12 – Hyperplan, marges et vecteurs supports pour la classification linéaire.

Cas non linéaire

Un classificateur linéaire peut ne pas être l'approche la plus appropriée pour la classification de données. Les SVM peuvent être employés lorsque des fonctions de décision non linéaires sont considérées. Cela est réalisé à partir de manipulations mathématiques qui transforment l'espace original de données dans un autre de dimension plus élevé. Dans cet espace, un hyperplan optimal est construit afin de séparer des classes distinctes. Étant donné Φ une transformation non linéaire de l'espace d'entrées $\mathbf{E}$ en un espace de re-description $\Phi(\mathbf{E})$, il vient (Burbridge & Buxton 2002) :

$$\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2, ..., \mathbf{E}_n) \xrightarrow{\Phi} \Phi(\mathbf{E}) = (\Phi_1(\mathbf{E}_1), \Phi_2(\mathbf{E}_2), ..., \Phi_d(\mathbf{E}_n)) \tag{3.57}$$

Le problème d'optimisation se traduit dans ce cas par

$$\begin{cases} \underset{\alpha}{\text{Max}} & (\tilde{L}(\alpha)) \\ \text{avec} & \tilde{L}(\alpha) = \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j c_i c_j \langle \Phi(\mathbf{E}_i), \Phi(\mathbf{E}_j) \rangle \right\} \\ \text{sous les contraintes} & \alpha_i \geq 0, 1 \leq i \leq n \quad \text{et} \quad \sum_{i=1}^{n} \alpha_i c_i = 0 \end{cases} \tag{3.58}$$

L'équation de l'hyperplan séparateur dans cet espace est

$$H(\mathbf{E}_i, \mathbf{E}_j) = \sum_{i,j=1}^{n} \alpha_i^* c_i \langle \Phi(\mathbf{E}_i), \Phi(\mathbf{E}_j) \rangle + \beta^* \tag{3.59}$$

où les coefficients α_i^*, $\mathbf{w}^*$ β^* sont obtenus comme décrit dans l'Équation 3.56. La Figure 3.13 illustre la transformation entre espaces.

Le produit scalaire $\langle \Phi(\mathbf{E}_i), \Phi(\mathbf{E}_j) \rangle$ est généralement très difficile à calculer quand la dimension de $\Phi(\mathbf{E})$ augmente. En cela, il existe des fonctions bi-linéaires symétriques positives $K(\mathbf{E}_i, \mathbf{E}_j)$ déjà introduites, appelées *fonctions noyau*. Ces fonctions sont plus faciles à calculer et il est possible de démontrer qu'elles correspondent à un produit scalaire $\langle \Phi(\mathbf{E}_i), \Phi(\mathbf{E}_j) \rangle$ dans un espace de grande dimension (Vapnik 1995). Lorsqu'une telle correspondance est exploitable, le problème

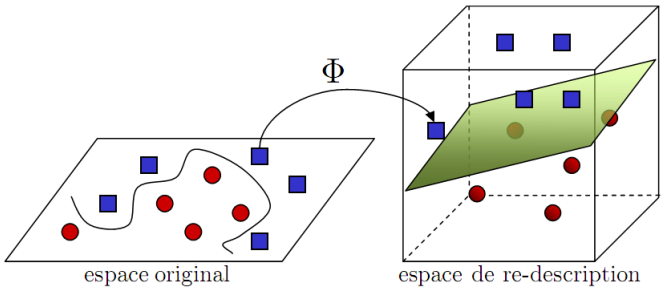


FIGURE 3.13 – Représentation de la transformation d'espaces (adapté de (Kaundal 2005)).

d'optimisation précédent est équivalent au problème suivant

$$\begin{cases} \max\limits_{\alpha} & (\tilde{L}(\alpha)) \\ \text{avec } \tilde{L}(\alpha) & = \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j c_i c_j K(\mathbf{E}_i, \mathbf{E}_j) \right\} \\ \text{sous les contraintes} & \alpha_i \geq 0, 1 \leq i \leq n \quad \text{et} \quad \sum_{i=1}^{n} \alpha_i c_i = 0 \end{cases} \tag{3.60}$$

L'équation de l'hyperplan séparateur recherché devient par conséquent

$$H(\mathbf{E}_i, \mathbf{E}_j) = \sum_{i,j=1}^{n} \alpha_i^* c_i K(\mathbf{E}_i, \mathbf{E}_j)\rangle + \beta^* \tag{3.61}$$

Dans la littérature, plusieurs *fonctions noyau* sont répertoriées. Les plus utilisées sont

- linéaires : $K(\mathbf{E}_i, \mathbf{E}_j) = \mathbf{E}_i.^*\mathbf{E}_j$.
- polynomiales : $K(\mathbf{E}_i, \mathbf{E}_j) = (\mathbf{E}_i.^*\mathbf{E}_j)^n$.
- gaussiennes (RBF) : $K(\mathbf{E}_i, \mathbf{E}_j) = e^{-\frac{||\mathbf{E}_i - \mathbf{E}_j||^2}{2\sigma^2}}$, σ étant un paramètre spécifique.
- sigmoïdes : $K(\mathbf{E}_i, \mathbf{E}_j) = \tanh(a(\mathbf{E}_i.^*\mathbf{E}_j - b))$, a et b étant des paramètres spécifiques.

Dans le cadre de cette thèse, les simulations avec les SVM ont été réalisées à l'aide de la boîte à outils gratuite LibSVM développée par (Chang & Lin 2001). Plusieurs méthodes ont été proposées dans la littérature pour étendre le développement mathématique énoncé auparavant au cas où plus de deux classes sont à séparer. La plus connue et celle utilisée dans cette thèse est basée sur l'approche « un-contre-tous » (*one-against-all*). Cette technique consiste à construire r classificateurs binaires en attribuant l'étiquette « +1 » aux observations de l'une des classes et l'étiquette « -1 » à toutes les autres. Cette procédure est répétée jusqu'à ce qu'un nombre de classes souhaitées soit obtenu (Hsu & Lin 2002).

Finalement, il est important de remarquer que de façon similaire à celle détaillée pour la méthode des réseaux de neurones, les simulations avec les SVM utilisent toujours des représentations symboliques des essais dynamiques comme données d'entrée.

3.3 Bilan

Dans les sections précédentes, trois méthodes de classification supervisée de données ont été présentées. La première, basée sur les arbres de décision Bayésiens (ADB) a une structure similaire aux arbres générés par la méthode des divisions hiérarchiques. La principale différence est que les ADB sont construits en tenant compte de groupes (classes) connus au préalable. En cela, les variables discriminantes sont calculées et des nouveaux essais (inconnus) sont classés

selon les critères définis. La deuxième méthode de classification utilise des réseaux de neurones (RN). Pour cela, une architecture d'un réseau MLP avec une couche cachée est employée dans les applications de cette thèse. Finalement, la troisième méthode, basée sur les SVM, est mise en œuvre pour la classification d'essais. L'un des apports majeurs de cette thèse est d'avoir couplé de façon originale l'ADS à ces techniques pour la classification de comportements structuraux. Les méthodes de classification supervisée peuvent aussi être utilisées pour l'affectation de nouveaux essais. Le principe est d'entraîner les algorithmes de classification à partir d'une base de données connue et, dans un deuxième temps, fournir des nouveaux essais pour que ces méthodes indiquent où ils peuvent être classés. L'ADS permettant la construction des tableaux de données plus compacts, rend possible l'utilisation de ces techniques de manière robuste et efficace.

3.4 Analyse des techniques de régression

Les méthodes de régression sont utilisées depuis le XIX$^{\text{ème}}$ siècle et ont été largement étendues dans les années 50 (Fisher 1954). L'un des objectifs est d'analyser la variation des observations dites « dépendantes » par rapport à un ensemble d'observations dites « indépendantes ». De plus, les méthodes de régression rendent possible la construction des fonctions (linéaires ou non) qui relient ces observations et permettent d'analyser de façon plus ou moins approfondie les relations entre ces observations.

L'un des enjeux de cette thèse est d'étudier les effets environnementaux, surtout des variations thermiques, sur les fréquences et modes propres d'une structure. Afin de répondre à cet enjeu, une étude des techniques de régression est proposée dans les paragraphes suivants. Dans un premier temps, la technique de régression linéaire multiple est introduite. Dans un deuxième temps, une procédure de régression non linéaire est mise en œuvre en utilisant des réseaux de neurones. Ces deux approches ont pour objectif la détermination de fonctions qui expliquent la variation des paramètres modaux par rapport aux températures mesurées. Une comparaison entre ces démarches est proposée afin de vérifier quel type de technique est la plus appropriée pour prendre en compte les effets thermiques sur la détection de modifications structurales.

3.4.1 Régression Linéaire Multiple

La régression linéaire multiple est une généralisation, à v variables explicatives, de la régression linéaire simple. Le modèle vise à exprimer la relation entre une série de températures mesurées notée par $\mathbf{T}_i = [T_i^1\ T_i^2\ \dots\ T_i^v]$ et une fréquence propre f_i (ou une composante modale [5]), où v est le nombre de sondes et $i = 1, \dots, n$, n étant le nombre d'enregistrements (observations). Mathématiquement, le modèle est représenté par

$$f_i = w_0 + \mathbf{T}_i\mathbf{w} + \epsilon_i, \quad i = 1, \dots, n \tag{3.62}$$

Si toutes les n observations des v sondes sont considérées, il vient

$$\mathbf{f} = \mathbf{w}_0 + \mathbf{T}\mathbf{w} + \boldsymbol{\epsilon} \tag{3.63}$$

où $\mathbf{f}$ est le vecteur des fréquences propres, $\mathbf{w}_0$ le vecteur des biais, $\mathbf{T}$ la matrice des températures regroupant n observations enregistrées par v sondes, $\mathbf{w} = [w_1\ w_2\ \dots\ w_v]^T$ [6] le vecteur des poids (les coefficients qui pondèrent les enregistrements de chaque sonde de température) et $\boldsymbol{\epsilon}$ le vecteur

5. Dans le paragraphes suivants, pour des raisons de clarté du texte, le développement mathématique est réalisé en tenant compte des fréquences propres. Toutefois, la même procédure peut être menée en considérant les composantes modales. Dans ce cas, les analyses de régression sont réalisées pour chaque composante séparément.

6. L'opérateur transposé est représenté par l'exposant $(.)^T$.

erreur du modèle. L'Équation (3.63) peut être réécrite de façon à assembler le terme $\mathbf{w}_0$ dans le vecteur de poids $\mathbf{w}$. Cela est réalisé en le redéfinissant de sorte que

$$\mathbf{f} = \mathbf{T}\mathbf{w} + \boldsymbol{\epsilon} \tag{3.64}$$

où $\mathbf{T}_i = [1\ T_i^1\ T_i^2\ \ldots\ T_i^v]$, (avec $i = 1, \ldots, n$) et $\mathbf{w} = [w_0\ w_1\ \ldots\ w_v]^T$. En exprimant cette relation sous une forme matricielle, il vient

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}_{(n\times 1)} = \begin{pmatrix} 1 & T_1^1 & T_1^2 & \ldots & T_1^v \\ 1 & T_2^1 & T_2^1 & \ldots & T_2^v \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & T_n^1 & T_n^1 & \ldots & T_n^v \end{pmatrix}_{(n\times v+1)} \times \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_v \end{bmatrix}_{(v+1\times 1)} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{(n\times 1)} \tag{3.65}$$

Afin d'estimer le vecteur de poids $\mathbf{w}$, la démarche la plus utilisée est celle basée sur la méthode des moindres carrés. Cette méthode vise la minimisation de l'espérance de l'erreur au carré donnée par

$$\min_{\mathbf{w}} \mathbb{E}[\epsilon_i^2], \quad i = 1, \ldots, n \tag{3.66}$$

L'espérance $\mathbb{E}$ est calculée comme la moyenne des erreurs des n observations (Sohn et al. 1999)

$$\begin{aligned} \mathbb{E}[\epsilon_i^2] &= \mathbb{E}[(f_i - \mathbf{T}_i\mathbf{w})^2], \quad i = 1, \ldots, n \\ &= \mathbb{E}[f_i^2 + \mathbf{w}^T\mathbf{T}_i^T\mathbf{T}_i\mathbf{w} - 2f_i\mathbf{w}^T\mathbf{T}_i^T] \\ &= \mathbb{E}[f_i^2] + \mathbf{w}^T\mathbb{E}[\mathbf{T}_i^T\mathbf{T}_i]\mathbf{w} - 2\mathbf{w}^T\mathbb{E}[\mathbf{T}_i^T f] \\ &= \mathbb{E}[f_i^2] + \mathbf{w}^T\mathbf{R}\mathbf{w} - 2\mathbf{w}^T\mathbf{p} \end{aligned} \tag{3.67}$$

où $\mathbf{R} = (\mathbf{T}_i^T\mathbf{T}_i)$ est l'auto-corrélation pour chaque ligne de la matrice de températures $\mathbf{T}$ et $\mathbf{p} = (\mathbf{T}_i^T f_i)$ est la corrélation croisée entre une fréquence et une ligne de la matrice de températures. Il est possible de noter que $\mathbb{E}[\epsilon_i^2]$ est quadratique par rapport à $\mathbf{w}$. Les coefficients estimés $\mathbf{w}^*$ sont donc calculés à partir de la minimisation de l'espérance $\mathbb{E}$. Cela est réalisé en résolvant sa dérivée

$$\frac{\partial \mathbb{E}[\epsilon_i^2]}{\partial \mathbf{w}} = 2(\mathbf{R}\mathbf{w}^* - \mathbf{p}) = 0 \tag{3.68}$$

En isolant $\mathbf{w}$,

$$\mathbf{w}^* = \mathbf{R}^{-1}\mathbf{p} \tag{3.69}$$

L'Équation (3.69) est connue sous le nom d'Équation de Wiener-Hopf et est la plus souvent utilisée pour résoudre des problèmes de régression linéaire. Il est important de noter que la résolution de l'Équation (3.69) est dépendante du nombre d'observations n et du nombre de variables v utilisées. Pour les applications étudiées dans le cadre de cette thèse, le nombre d'observations est significativement supérieur au nombre de variables ($n \gg v$). Toutefois, si le modèle de régression est « sous-spécifié » (i.e., $n < v$), la solution du système produira un certain nombre de poids $\mathbf{w}^*$ égaux à zéro, ce qui mènera à des erreurs aussi égales à zéro pour les observations associées.

L'un des avantages de la régression linéaire est que de nombreux phénomènes en sciences et en ingénierie sont relativement bien décrits par des modèles simples. Cela est essentiellement lié au fait que sur de courts intervalles, la plupart des analyses peuvent être menée par des modèles linéaires. À l'inverse, quand il s'agit de grands ensembles de données, le modèle de régression linéaire peut ne pas être le plus adapté. Cette technique est souvent très sensible à la présence d'observations « anormales », c'est-à-dire, celles que se démarquent des autres données. Afin de traiter ces éventuelles limitations, un modèle de régression non linéaire est présenté dans la suite.

3.4.2 Régression Non Linéaire

Il existe plusieurs méthodes dans la littérature pour ajuster un modèle de régression non linéaire. Dans cette thèse, une approche basée sur les réseaux de neurones est proposée. L'un des objectifs est de la comparer avec des modèles de régression linéaire et d'observer comment l'ajustement peut être amélioré ou non en insérant des non linéarités.

Une démarche similaire à celle expliquée dans le paragraphe 3.2.2 est proposée. Auparavant, des représentations symboliques d'essais étaient utilisées comme données d'entrée et des étiquettes comme données de sortie, le but étant de classer des essais inconnus. Ici, les données d'entrée sont les n enregistrements $T_l^q, (l = 1, \ldots, n; q = 1, \ldots, v)$ effectués par les v sondes de températures et les sorties, les fréquences propres ou les composantes des déformées modales identifiées. L'Équation (3.70) exprime cette relation.

$$f_i = \Upsilon \left[\psi^{(\mathrm{CS})} \left(\sum_{j=1}^{D+1} w_{ij}^{(\mathrm{CC})} \psi^{(\mathrm{CC})} \left(\sum_{l=1}^{n+1} \sum_{q=1}^{v} w_{ql}^{(\mathrm{CE})} T_l^q \right) \right) \right], \; i = 1, \ldots, m \tag{3.70}$$

où CE, CC et CS représentent la couche d'entrées, la couche cachée et la couche de sortie, respectivement, m le nombre de sorties, $w(.)$ les poids synaptiques, $\psi(.)$ les fonctions d'activation de chaque couche, D le nombre de neurones de la couche cachée et Υ l'opérateur d'échelle. Il est noté que $n + 1$ et $D + 1$ sont utilisés au lieu de n et D, respectivement. Cela indique que les biais sont déjà pris en compte dans le calcul des sommes. Comme décrit précédemment, les fonctions d'activation tangente hyperbolique (Équation (3.42)) et sigmoïde (Équation (3.41)) sont généralement employées dans des problèmes de régression non linéaire. Bien que les valeurs de sortie soient comprises entre des intervalles précis, il est toujours possible d'utiliser des manipulations vectorielles permettant de les remettre aux échelles originales. Soient $\mathbf{y}$ et $\mathbf{f}$ les vecteurs des sorties du réseau et des fréquences propres mesurées (valeurs réelles). L'opérateur Υ s'écrit par

$$\mathbf{f} = \Upsilon(\mathbf{y}) = \left(\frac{f_{\max} - f_{\min}}{y_{\max} - y_{\min}} \right) (\mathbf{y} - y_{\min} \cdot \mathbf{1}) + f_{\min} \cdot \mathbf{1} \tag{3.71}$$

où $y_{\min}$ et $y_{\max}$ sont, respectivement, les valeurs minimales et maximales du vecteur des sorties $\mathbf{y}$ (normalement égales à 0 et 1 ou -1 et 1) ; $f_{\min}$ et $f_{\max}$ les valeurs minimales et maximales du vecteur des fréquences propres $\mathbf{f}$, respectivement et $\mathbf{1}$ dénote le vecteur unité de même dimension du vecteur $\mathbf{y}$.

Contrairement au cas de la régression linéaire, les méthodes non linéaires basées sur les réseaux de neurones permettent en général de trouver un meilleur ajustement pour des grands ensembles de données. Toutefois, cela peut devenir un inconvénient étant donné que ces techniques peuvent forcer une relation entre les données d'entrée et de sortie qui n'existe pas dans la réalité. Cela est produit par un sur-apprentissage du réseau où il « mémorise » simplement le jeu de données et, par conséquent, peine à généraliser au cas des nouvelles observations. Ainsi comme il est proposé pour la classification, les données utilisées pour la régression sont divisées en deux groupes : l'un pour l'entraînement et l'autre pour la validation. Une fois le réseau calibré et ajusté par rapport au groupe de validation, il est testé en considérant des nouveaux enregistrements. Dans cette thèse, les simulations sur la régression non linéaire sont réalisées à l'aide de la boîte à outils Netlab développé par Bishop (Bishop 1995).

3.4.3 Applications

Dans cette thèse, les modèles de régression sont utilisés dans deux cadres distincts. Dans une premier temps, le but est de prédire des valeurs des fréquences propres pour des nouveaux

enregistrements de température. Ce type d'étude permet d'analyser comment les fréquences varient et détecter des modifications anormales. Le premier pas consiste à construire un modèle de régression avec des enregistrements de température et fréquences réalisés à un instant t. Ensuite, le modèle est alimenté avec des nouveaux enregistrements de température, obtenus à un instant $t + \Delta t$ et les fréquence prédites sont comparées avec celles mesurées.

Il est pratiquement impossible d'avoir des fréquences prédites identiques à celles mesurées en raison des incertitudes liées aux modèles de régression, aux erreurs numériques, etc. Le plus important, cependant, est d'évaluer des intervalles de confiance autour des valeurs calculées. Les bornes inférieure et supérieure d'un intervalle à $100(1-\alpha)\%$ de confiance sont données par

$$\hat{f}_i^{(t+\Delta t)} \mp t_{\alpha/2,n-v+1}\sqrt{\sigma^2\left(1+\widehat{\mathbf{T}}_i^T\left(\mathbf{T}^T\mathbf{T}\right)^{-1}\widehat{\mathbf{T}}_i\right)},\ \text{avec}\ \sigma^2=\sum_{i=1}^{n}\frac{(\hat{f}_i^{(t)}-f_i^{(t)})^2}{n-v},\ n\gg v \quad (3.72)$$

où $\mathbf{T}$, $f_i^{(t)}$ et $\hat{f}_i^{(t)}$ sont, respectivement, la matrice des températures, la fréquence propre mesurée et la fréquence propre prédite à l'instant t par le modèle de régression. $\widehat{\mathbf{T}}_i = [1\ \widehat{T}_i^1\ \widehat{T}_i^2\ \ldots\ \widehat{T}_i^v]$ et $\hat{f}_i^{(t+\Delta t)}$ représentent, respectivement, le vecteur des nouveaux enregistrements de température et la fréquence propre prédite à l'instant $(t+\Delta t)$ par le modèle de régression. $t_{\alpha/2,n-v+1}$ est trouvé dans le tableau de distributions de t-Student avec $n-v+1$ degrés de liberté. L'Équation (3.72) peut être utilisée pour des modèles de régression linéaire (Sohn et al. 1999) et non linéaire, sans perte de généralité (Seber & Wild 2003).

L'objectif de cette étude est de séparer les effets d'une modification liée aux variations de température de celle liée à d'autres phénomènes, comme des changements structuraux, par exemple. En effet, en considérant que le modèle construit représente les données de la régression avec une bonne précision, il est possible de détecter toute nouvelle observation qui se trouverait dehors de l'intervalle de confiance et représenterait potentiellement une anomalie.

Toutefois, cette analyse n'apporte pas suffisamment d'éléments pour conclure si l'observation anormale est causée par d'autres effets environnementaux comme l'humidité, le vent, etc. ou par des modifications structurales. Pour cela, l'approche proposée est couplée à des méthodes d'affectation d'essais décrites dans le paragraphe 2.3. L'idée est d'utiliser l'information donnée par les méthodes de régression (l'identification des essais « anormaux ») et de l'appliquer afin de vérifier si, effectivement, ces essais appartiennent ou non à un groupe de comportement normal.

Le deuxième cadre d'application consiste à « corriger » les fréquences propres et les déformées modales en enlevant les effets thermiques. Pour les premières, cela est réalisé à partir de la normalisation de toutes les fréquences par rapport à une fréquence mesurée à une température donnée

$$\bar{\mathbf{r}} = \frac{\mathbf{f}}{f_{\text{ref}}} \quad (3.73)$$

où $\mathbf{f}$ est le vecteur de fréquences propres mesurées, f_{ref} la fréquence de référence mesurée à une température T_{ref} (choisie au préalable) et $\bar{\mathbf{r}}$ le vecteur des ratios des fréquences.

Les méthodes de régression linéaire et non linéaire sont appliquées en utilisant les enregistrements de températures comme données d'entrée et les ratios $\bar{\mathbf{r}}$ comme sorties

$$\bar{\mathbf{r}} = \Theta(\mathbf{T}, \mathbf{w}) \quad (3.74)$$

La résolution de l'Équation (3.74) permet de déterminer les coefficients $\mathbf{w}^*_{\text{corr}}$ corrigés. Dans le cas de la régression linéaire, les ratios des fréquences propres corrigées sont évalués en multipliant la matrice de températures $\mathbf{T}$ par les poids corrigés

$$\bar{\mathbf{r}}^*_{\text{corr}} = \mathbf{T}\mathbf{w}^*_{\text{corr}} \quad (3.75)$$

Dans le cas de la régression non linéaire, cette multiplication est réalisée de manière indirecte. En effet, une fois le réseau de neurones entraîné, la matrice des températures y est réintroduite et multipliée par les poids synaptiques maintenant calibrés. La sortie du réseau est alors le vecteur des ratios $\bar{\mathbf{r}}^*_{\text{corr}}$.

Finalement, les fréquences corrigées sont calculées par le rapport des fréquences mesurées avec les ratios corrigés

$$\mathbf{f}_{\text{corr}} = \mathbf{f}./\bar{\mathbf{r}}^*_{\text{corr}} \tag{3.76}$$

où ./ représente l'opérateur de division élément par élément.

Il est à noter que les poids corrigés $\mathbf{w}^*_{\text{corr}}$ obtenus à partir de la résolution de l'Équation (3.74) peuvent être également utilisés pour la correction des paramètres modaux identifiés à l'instant $(t + \Delta t)$. Il est manifeste que l'adéquation du modèle de régression construit à l'instant t doit être vérifiée au préalable. En d'autres termes, il faut vérifier que les corrections produites par les modèles construits aux instants t et $(t+\Delta t)$ soient comprises dans une même bande de confiance. Dans un tel cas, cela revient à dire que la correction parvenue par le modèle « $(t + \Delta t)$ » n'est pas significativement différente de celle du modèle « t ». Dans ce cas, les poids corrigés obtenus à partir de l'Équation (3.74) sont utilisés pour la correction des nouvelles mesures. Ainsi, les Équations (3.75) et (3.76) sont réécrites comme

$$\hat{\mathbf{r}}^*_{\text{corr}} = \widehat{\mathbf{T}}\mathbf{w}^*_{\text{corr}} \tag{3.77}$$

et

$$\hat{\mathbf{f}}_{\text{corr}} = \hat{\mathbf{f}}./\hat{\mathbf{r}}^*_{\text{corr}} \tag{3.78}$$

Dans le cas contraire, un nouveau modèle est crée afin d'évaluer des nouveaux poids corrigés. Dans ce cas, l'Équation (3.74) est réécrite en utilisant la matrice de températures $\widehat{\mathbf{T}}$ et le vecteur des ratios $\hat{\mathbf{r}}$ enregistrés à l'instant $(t + \Delta t)$.

Si les modes propres sont considérées, la même démarche peut être appliquée en faisant la correction des composantes modales séparément. Étant donné que cette manipulation est valable pour les deux types de régression, la démarche de correction proposée est aussi couplée aux méthodes de clustering. En effet, l'idée est de vérifier si les partitionnements obtenus avec des fréquences et modes propres dits non corrigés sont différents de ceux trouvés en appliquant cette approche.

3.4.4 Bilan

Les méthodes de régression proposées dans cette section ont pour but d'étudier les effets thermiques sur les paramètres modaux structuraux. En cela, deux approches ont été présentées. La première, basée sur la régression linéaire multiple et la deuxième, en utilisant des réseaux de neurones. La démarche consiste à créer des modèles permettant d'une part, de détecter des variations anormales des fréquences (i.e., qui ne sont pas liées aux changements de température) et d'autre part, de filtrer les effets thermiques et évaluer des nouvelles valeurs corrigées pour les paramètres modaux. Dans ce dernier cas, les fréquences et modes propres corrigés sont utilisés dans les procédures de clustering afin de vérifier si les variations de température sont ou non négligeables. Ceci est vérifié en comparant les clusters obtenus sans prise en compte de température à ceux déterminés suite à la correction des paramètres modaux.

3.5 Analyse en Composantes Principales

L'analyse en composantes principales (ACP ou PCA en anglais) est une technique statistique multi-variée, aussi connue sous le nom de transformation Karhunen-Loève ou encore décomposition orthogonale (Krzanowski 2000). En général, son utilisation a pour but de simplifier des problèmes complexes par leur taille. De nos jours, il est courant de devoir manipuler des ensembles de données possédant des dimensions importantes. Manipuler ces grands ensembles de données peut alors être rendu impossible ou approximatif par le nombre de données trop importantes par rapport au problème. Il devient alors difficile d'obtenir les informations souhaitées alors qu'un simple traitement pourrait rendre le processus relativement plus abordable.

L'idée de l'ACP est de réduire la dimension des ensembles de données en ne gardant que les informations « importantes ». Pour cela, l'ensemble de données subit des modifications mathématiques permettant de mettre en valeur leurs ressemblances et différences. L'ACP est notamment avantageuse dans le cas de grandes dimensions puisqu'elle rend une interprétation plus aisée de l'ensemble de données.

L'ACP a été appliquée dans le domaine de la dynamique des structures pour la construction des modèles réduits (Azeez & Vakakis (2001) et Kerschen et al. (2003)), pour l'analyse modale (Han & Feeny 2003) et pour l'actualisation des modèles linéaires et non linéaires (Lenaerts et al. 2003). Cette technique a été également utilisée pour la détection d'endommagements structuraux (De Boe & Golinval 2003) et pour la validation du positionnement de capteurs (Friswell & Inman 2000).

Dans le cadre de cette thèse, l'ACP est utilisée comme une approche alternative à celle basée sur des modèles de régression. Le but est de détecter des variations anormales des fréquences propres sous différentes conditions environnementales, notamment de température. En général, cette technique peut être employée quand il n'existe pas des sondes thermiques instrumentées sur l'ouvrage. En réalité, les effets environnementaux sont traités comme variables inhérentes aux variations fréquentielles. L'idée de cette approche est liée au fait que les changements sur les fréquences propres dus aux variations thermiques peuvent être séparés de ceux causés par un endommagement. Ainsi comme il a été montré dans les paragraphes précédents, les erreurs obtenues lors des procédures de prédiction des fréquences propres par le modèle ACP peuvent servir comme un indicateur d'endommagement.

3.5.1 Méthodologie

L'objectif des étapes suivantes consiste à changer la base de représentation des données. Cela peut se résumer à une rotation, mais généralement ceci est suivi d'une diminution du nombre de dimensions à prendre en compte (Figure 3.14). En considérant le vecteur de fréquences propres $\mathbf{f}_k$ $(k = 1, \ldots, v)$, il est possible de construire la matrice des fréquences propres $\mathbf{F} \in \Re^{n \times v}$, où v représente le nombre de modes considérés et n le nombre d'observations. L'ACP fournit donc une projection linéaire des données de dimension v vers une dimension réduite q. Cette transformation est représentée par

$$\mathbf{Z} = \mathbf{F}\Psi \tag{3.79}$$

où $\mathbf{Z} \in \Re^{n \times q}$ est la matrice des fréquences propres transformées et $\Psi \in \Re^{v \times q}$ la matrice de transformation.

La matrice de transformation Ψ peut être calculée à partir de l'extraction des q vecteurs propres de la matrice de covariance de $\mathbf{F}$, ici notée $\mathbf{X}$. La première étape du procédé est alors de trouver le degré de dépendance entre chacune des dimensions. En prenant les dimensions deux à deux, il est possible d'indiquer la façon dont leurs variations sont liées : si la covariance est

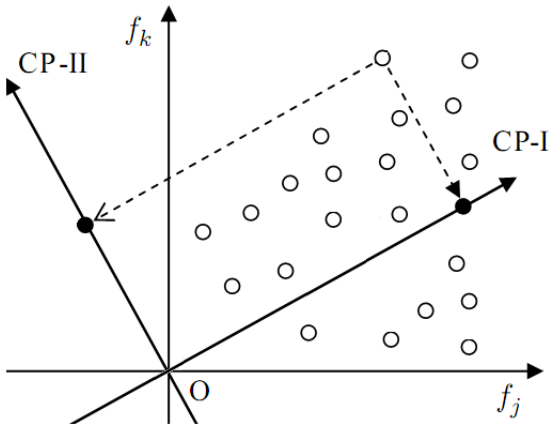


FIGURE 3.14 – Exemple hypothétique d'une transformation 2D en composantes principales.

nulle alors les deux dimensions n'ont aucun lien entre elles. L'expression permettant de calculer la covariance entre les vecteurs des fréquences propres $\mathbf{f}_j$ et $\mathbf{f}_k$ est donnée par

$$cov(\mathbf{f}_j, \mathbf{f}_k) = \frac{1}{n}\left(\mathbf{A}^T\mathbf{A}\right), \quad \text{avec} \quad \mathbf{A} = [\mathbf{f}_j - \bar{f}_j \cdot \mathbf{1}, \mathbf{f}_k - \bar{f}_k \cdot \mathbf{1}] \tag{3.80}$$

où $\bar{f}_j$ et $\bar{f}_k$ sont les valeurs moyennes des vecteurs $\mathbf{f}_j$ et $\mathbf{f}_k$, respectivement et $\mathbf{1}$ dénote le vecteur unité.

Afin d'avoir un estimateur non biaisé, le dénominateur de l'Équation (3.80) peut être défini comme égal à $n-1$ si l'étude porte sur un échantillon des données et non sur l'ensemble des données.

La matrice de covariance $\mathbf{X}$ s'écrit alors par

$$\mathbf{X} = \begin{pmatrix} var(\mathbf{f}_1) & cov(\mathbf{f}_1, \mathbf{f}_2) & \dots & cov(\mathbf{f}_1, \mathbf{f}_v) \\ cov(\mathbf{f}_2, \mathbf{f}_1) & var(\mathbf{f}_2) & \dots & cov(\mathbf{f}_2, \mathbf{f}_v) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\mathbf{f}_v, \mathbf{f}_1) & cov(\mathbf{f}_v, \mathbf{f}_2) & \dots & var(\mathbf{f}_v) \end{pmatrix}$$

où $var(.)$ indique la variance et $cov(.)$ la covariance entre vecteurs.

La deuxième étape consiste à extraire les vecteurs et valeurs propres de la matrice $\mathbf{X}$. Le problème aux valeurs propres peut être énoncé par : si le produit entre un vecteur $\boldsymbol{\phi}$ et la matrice $\mathbf{X}$ est égal à lui-même, à un coefficient multiplicateur près, $\boldsymbol{\phi}$ est donc un vecteur propre de la matrice $\mathbf{X}$. En d'autres termes

$$\mathbf{X}.\boldsymbol{\phi} = \lambda.\boldsymbol{\phi} \leftrightarrow (\mathbf{X} - \lambda.\mathbf{I}).\boldsymbol{\phi} = 0 \tag{3.81}$$

où λ est la valeur propre associée au vecteur propre $\boldsymbol{\phi}$ et $\mathbf{I}$ la matrice identité.

Résoudre un problème aux valeurs propres peut être délicat à mesure que la dimension de la matrice de covariance augmente. En cela, il existe d'autres techniques pour calculer les valeurs et vecteurs propres. La solution la plus connue est la décomposition en valeurs singulières (SVD, en anglais). La SVD permet de décomposer la matrice $\mathbf{X}$ de la manière suivante

$$\mathbf{X} = \mathbf{U}\Lambda\mathbf{V}^T \tag{3.82}$$

où $\mathbf{U}$ est une matrice orthonormal où les colonnes sont les vecteurs qui définissent les composantes principales dans l'espace de transformation, $\mathbf{V}$ la matrice contenant les vecteurs de l'espace

d'origine, Λ la matrice des valeurs singulières ou valeurs propres, lorsque l'espace Hilbertien est considéré (ce qui est une généralisation d'un espace Euclidien à une dimension quelconque).

Une fois que les valeurs et vecteurs propres ont été extraits de la matrice de covariance des dimensions, les valeurs propres sont triées selon leur amplitude ou encore leur « énergie ». Les vecteurs propres correspondants à ces valeurs propres sont les plus pertinents, c'est-à-dire, ceux qui permettent de garder le plus d'informations possible sur l'ensemble original de données. Étant donné que chaque vecteur propre correspond, d'une certaine façon, à une composante principale, il est intuitif d'utiliser ceux qui ont la meilleur « représentativité ».

La matrice de transformation est donc construite en n'utilisant que q vecteurs propres (colonnes) de la matrice $\mathbf{U}$. Le nombre de composantes principales choisi dépend fortement de type de problème traité. Il est alors important d'observer les pourcentages d'explication fournis par les valeurs propres. Dans certains cas, il est possible qu'une seule composante soit capable de représenter les données originales sans grandes pertes d'information.

La perte d'information en faisant cette projection peut être évaluée en re-projetant les données vers l'espace d'origine. Manifestement, si tous les vecteurs propres sont utilisés, l'erreur de re-projection est égale à zéro. La matrice des fréquences propres re-projetées $\hat{\mathbf{F}}$ est calculée par

$$\hat{\mathbf{F}} = \mathbf{Z}\Psi^T = \mathbf{F}\left(\Psi\Psi^T\right) \tag{3.83}$$

La matrice d'erreur résiduelle peut être estimée par

$$\mathbf{\Xi} = \mathbf{F} - \hat{\mathbf{F}} \tag{3.84}$$

En cela, l'indice d'endommagement κ est défini à partir de la norme de l'erreur évaluée pour les observations de chacune des v fréquences propres considérées, notée par $\mathbf{\Xi}_i \in \Re^{1\times v}$ $(i = 1, \ldots, n)$. Dans le cadre de cette thèse, deux normes sont utilisées (Worden et al. 2000) : la norme Euclidienne et la norme de Mahalanobis. La première s'écrit par

$$\kappa_i = ||\mathbf{\Xi}_i||, \quad i = 1, \ldots, n \tag{3.85}$$

et la deuxième, par

$$\kappa_i = \sqrt{\mathbf{\Xi}_i \mathbf{X}^{-1} \mathbf{\Xi}_i^T}, \quad i = 1, \ldots, n \tag{3.86}$$

En considérant que l'indice κ est distribué selon une loi normale, une analyse statistique peut être réalisée en évaluant sa moyenne et respectif écart-type. En ce sens, il est possible de construire des intervalles de confiance pour la moyenne et, de façon similaire à celle présentée pour les méthodes de régression, d'identifier des variations anormales des fréquences propres. Les bornes inférieure et supérieure de l'intervalle sont calculées par

$$[\mathrm{IC}_{\mathrm{inf}}; \mathrm{IC}_{\mathrm{sup}}] = \overline{\kappa} \mp \alpha\sigma_\kappa \tag{3.87}$$

où $\overline{\kappa}$ et σ_κ sont la moyenne et écart-type de l'indice d'endommagement et α définit le degré de confiance (par exemple, pour un intervalle de confiance à 95%, $\alpha = 1,96$ en utilisant la table de distribution normale).

3.5.2 Bilan

Dans les paragraphes précédents, une méthodologie basée sur l'Analyse en Composantes Principales (ACP) a été proposée. L'objectif de cette démarche est d'apporter nouveaux éléments de réponse pour la détection de variations anormales des fréquences propres. En réalité, l'idée est

de pouvoir séparer les déviations fréquentielles liées aux changements de températures de celles liées à la présence d'endommagements structuraux. Particulièrement, cette approche n'utilise pas les mesures enregistrées par les sondes de températures. Contrairement à ce qui a été proposée par les méthodes de régression, les effets thermiques sont ici considérés comme inhérents aux valeurs des fréquences propres identifiées. En faisant la projection de la base de données vers un autre espace, de dimension réduite, il est alors possible de discriminer un comportement normal d'un anormal.

Chapitre 4

Validation numérique et étude de sensibilité

L'enjeu majeur de ce chapitre est de présenter, à partir d'une étude numérique, l'application des méthodes de classification de données énoncées dans les paragraphes précédents. L'idée est d'évaluer la robustesse de l'approche basée sur le couplage de l'ADS aux méthodes permettant de discriminer des comportements structuraux différents. Pour cela, diverses simulations d'endommagement et de bruit sont réalisées en utilisant des signaux et paramètres modaux identifiés, transformés en données symboliques (Cury et al. 2009*a*,*b*, Cury & Crémona 2010*c*).

4.1 Présentation du cas d'étude

Les études numériques représentent généralement un premier pas pour analyser, à l'aide de simulations, la performance des techniques de détection d'endommagements structuraux. Une vaste étude statistique concernant les indicateurs d'endommagement dits « évolués » (cf. paragraphe 1.2.2) a été menée par Alvandi & Crémona (2006). Les simulations numériques permettent de tester un nombre significatif de conditions structurales (e.g., différents niveaux de dommages, de bruit, de conditions aux limites, de matériaux, etc.). Il apparaît donc pertinent d'analyser la robustesse des approches proposées dans cette thèse à l'aide de simulations numériques. Ce chapitre a pour objectif majeur d'analyser la sensibilité des méthodes de classification de données à partir de diverses simulations d'endommagement et de bruit. En particulier, le but est d'observer si ces méthodes sont capables de détecter un dommage structural même si d'importants niveaux de bruit sont ajoutés aux mesures dynamiques.

Dans le cadre de cette thèse, un modèle numérique d'une poutre bi-appuyée est étudié. La poutre est modélisée à partir d'une discrétisation par éléments finis (Zienkiewicz & Taylor 2000) sous MatLab® à l'aide d'un programme développé par l'auteur. Le modèle est discretisé en 200 éléments de portique (deux nœuds par élément avec trois degrés de liberté par nœud). La poutre, de section de type I, a une longueur de 6 m et le matériau choisi est l'acier. La source d'excitation est une force aléatoire placée au nœud 178 (à 0,69 m de l'appui situé à droite) dont l'amplitude et la fréquence suivent un processus aléatoire. Il est à noter que cette position permet d'exciter les cinq premières fréquences propres (et modes propres associés) de la structure.

Les réponses dynamiques sont mesurées en onze points équidistants de la poutre pendant 50 secondes. La fréquence d'échantillonnage est de 1/1000 s, ce qui correspond à 50000 instants de mesure par « capteur ». La Figure 4.1 illustre la représentation schématique de la poutre

étudiée, la position de la source d'excitation et les points de mesure. Les propriétés relatives à la géométrie et à la nature du matériau de la poutre considérée sont les suivantes :

- **module d'Young :** 210 GPa,
- **masse volumique :** 7850 kg.m^{-3},
- **section droite :** 2,81 $\times$ 10^{-3} m^{-2},
- **moment d'inertie :** 1,845 $\times$ 10^{-8} m^{-4}.

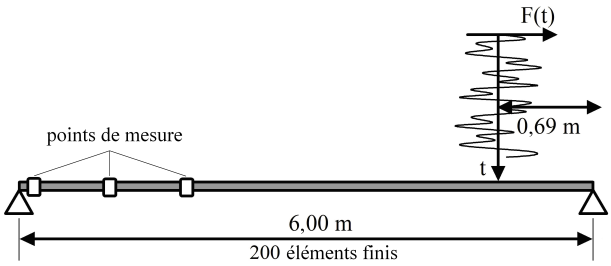


FIGURE 4.1 – Schéma de la poutre étudiée.

Afin d'analyser la robustesse des méthodes de classification vis-à-vis des niveaux d'endommagement, de leurs différentes localisations et d'occurrences simultanées, quatre scénarios sont proposés. Dans toutes les simulations réalisées, il est supposé que les niveaux d'endommagement correspondent à une réduction du module d'Young de(s) l'élément(s) endommagé(s). L'état initial de la structure, nommé D0, représente la poutre intègre (i.e., sans endommagement). Le premier scénario d'endommagement, nommé D1, consiste en une réduction de 5% du module d'Young aux éléments 100, 101 et 102 (ce qui correspond au milieu de la poutre). Pour le deuxième scénario, D2, une réduction de 5% aux éléments 1, 2 et 3 (à coté de l'appui gauche) est considérée. Finalement, pour le troisième scénario d'endommagement (D3), les éléments 150, 151 et 152 (situés au quart de la poutre) ont leur module d'Young également réduit de 5%. Par ailleurs, le module d'Young des éléments au milieu de la poutre (éléments 100, 101 et 102) est maintenant fixé à 10%. La Figure 4.2 détaille les trois scénarios d'endommagement imposés à la structure.

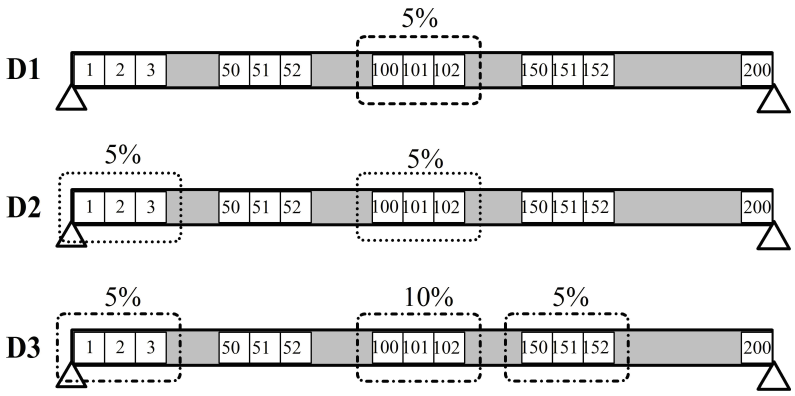


FIGURE 4.2 – Scénarios d'endommagement simulés.

Trois niveaux de bruit ont été considérés dans les simulations. Considérant que n0 indique les réponses dynamiques sans bruit, le premier niveau, noté n1, consiste à ajouter 1% de bruit aux réponses dynamiques de la poutre. Le deuxième niveau, n2, ajoute 2% de bruit et le troisième (n3), 5%. Les bruits sont générés en multipliant le niveau (le pourcentage) de bruit par l'écart-type de chaque point de mesure (« capteur ») et par un vecteur de valeurs aléatoires issues d'une distribution gaussienne. Ceci caractérise un bruit *blanc*. La réponse dynamique bruitée est donc

obtenue en ajoutant la réponse non bruitée au vecteur de bruit calculé de la façon suivante :

$$X_{i,\text{bruitée}} = X_i + n_{\text{bruit}} \cdot \sigma_{X_i} \cdot V \sim N(0,1) \tag{4.1}$$

où X_i et $X_{i,\text{bruitée}}$ représentent, respectivement, les réponses dynamiques de la structure non bruitée et bruitée mesurées par le capteur i ; n_{bruit} le niveau de bruit (en pourcentage) ajouté, σ_{X_i} l'écart-type de X_i et $V \sim N(0,1)$ un vecteur gaussien de moyenne nulle et d'écart-type unitaire.

Pour chaque scénario d'endommagement et chaque niveau de bruit, 10 essais différents ont été simulés. Cela produit un total de 160 essais (cf. Tableau 4.1). Sachant que pour chaque essai, 11 capteurs mesurent 50000 points chacun, il existe $160 \times 11 \times 50000 = 88000000$ (quatre-vingt huit millions) de valeurs à traiter. Ce premier calcul donne un aperçu de la taille conséquente des bases de données à traiter et du besoin d'utiliser des outils pertinents tels que l'ADS pour une manipulation adéquate de ces données.

		Scénario d'endommagement			
		D0	D1	D2	D3
Niveau de bruit	n0	10	10	10	10
	n1	10	10	10	10
	n2	10	10	10	10
	n3	10	10	10	10

TABLEAU 4.1 – Résumé des essais simulés.

Les paramètres modaux de la poutre sont identifiés à partir de l'outil Sys-Ident développé au LCPC en 2001 (Barbosa & Crémona 2001). Comme décrit dans le paragraphe 2.2, il ne suffit généralement pas de trouver une valeur numérique convenable pour les paramètres modaux, mais il faut également estimer leur précision et les incertitudes liées à l'identification modale. L'identification modale est réalisée en utilisant la méthode du Décrément Aléatoire (Cole 1968) couplée à la méthode d'Ibrahim (Ibrahim 1977, 1987). La méthode du Décrément Aléatoire est une technique non paramétrique temporelle, simple à mettre en œuvre. Son principe est d'estimer des fonctions dites de décrément aléatoire à partir de la moyennisation des segments de mesures, vérifiant des conditions précises dites de déclenchement. Ces fonctions sont des fonctions amorties, ce qui permet d'y appliquer des algorithmes classiques d'identification modale, telle que la méthode d'Ibrahim. Ibrahim (1977) a proposé d'utiliser deux matrices déduites de la réponse transitoire de la structure (estimée par la méthode du Décrément Aléatoire). L'une de ces matrices est la matrice de réponse et l'autre est celle décalée par rapport au temps. La méthode d'Ibrahim (1977) avait l'inconvénient de nécessiter un nombre de voies de mesure supérieur au nombre de paramètres modaux estimés. Ibrahim (1987) propose alors le concept de « pseudo-mesures ». L'idée est de créer artificiellement des voies de mesure supplémentaires en construisant des réponses dynamiques décalées dans le temps. Cela permet d'estimer plusieurs *réalisations* des caractéristiques modales à partir de ces réponses. Par l'étude d'un problème aux valeurs propres, les paramètres modaux de la structure peuvent être estimés (Alvandi 2004). Le développement mathématique de ces méthodes est fourni dans l'Annexe D.

Dans cette étude, 50 réalisations sont évaluées pour chaque essai menant à 50 paires de fréquences et modes propres évaluées. À partir de ces réalisations, les paramètres modaux sont transformés en données symboliques, soit en intervalles interquartiles soit en histogrammes à 20 catégories 2.2. Les simulations réalisées dans ce chapitre ne considèrent que les cinq premières fréquences et modes propres identifiés à partir des réponses dynamiques non bruitées (n0) et bruitées (n1, n2 et n3).

4.2 Analyse classique

Préalablement à l'utilisation de l'ADS, une étude classique simple, utilisant les fréquences et modes propres, est proposée. Le Tableau A.1 (fourni dans l'Annexe A) présente les valeurs moyennes ($\overline{f}$) et écarts-types (σ_f) des réalisations pour chacune des fréquences propres identifiées, par scénario d'endommagement, pour le niveau de bruit n0 (sans bruit). Par ailleurs, les différences en pourcentage entre les fréquences de la structure non endommagée (D0) et endommagée (D3) sont calculées. Il est noté que la première, troisième et cinquième fréquences sont les plus affectées par les endommagements imposés à la structure. Par ailleurs, à mesure que le niveau de dommage augmente, les modifications sont plus significatives sur les fréquences élevées.

Le Tableau A.2 regroupe les valeurs moyennes et écarts-types des fréquences identifiées à partir des signaux bruités (niveau n1). Dans ce cas, les valeurs des fréquences sont légèrement différentes de celles identifiées sans bruit. Toutefois, les variations entre les niveaux d'endommagement sont toujours du même ordre de celles montrées au Tableau A.1. Par ailleurs, les Tableaux A.3 et A.4 détaillent les résultats obtenus en considérant les niveaux de bruit n2 et n3, respectivement. Des variations similaires à celles déjà mentionnées sont encore remarquées. Il est noté qu'une simple analyse des valeurs moyennes et écarts-types ne permet pas de distinguer clairement les différents niveaux d'endommagement.

La Figure 4.3 montre l'évolution des cinq premières fréquences propres pour les quatre scénarios d'endommagement (D0 à D3) et les quatre niveaux de bruit (n0 à n3). Il est noté que dans certains cas (e.g., la première et quatrième fréquences propres), le niveau de bruit altère le sens de variation des fréquences. Cet effet est accentué lorsqu'une analyse par intervalles de confiance pour la valeur moyenne est réalisée. La Figure 4.4 présente les variations observées pour les cinq fréquences propres, regroupées par niveau de bruit. Les intervalles de confiance (IC= $[\overline{f} - \sigma_f; \overline{f} + \sigma_f]$) sont construits en considérant les écarts-types calculés. Il est observé que dans la plupart des simulations réalisées, les intervalles de confiance se superposent. Ceci est observé même dans les cas où le niveau de bruit est relativement faible. Par conséquent, il est très difficile de conclure si une modification structurale est ou non conséquence d'un dommage structural. Cette remarque s'applique en particulier aux cas des scénarios D1 et D2 pour lesquels le bruit affecte l'identification dans le sens contraire de l'endommagement imposé.

Un autre outil statistique permettant de comparer différents échantillons de données est le test d'hypothèse. Comme il a déjà été mentionné, un test d'hypothèse est une démarche qui consiste à accepter ou non une hypothèse statistique, appelée *hypothèse nulle*, en fonction d'un jeu de données. Dans cette application, l'hypothèse nulle (H_0) est que les fréquences propres identifiées pour chaque niveau d'endommagement sont issues d'une même distribution. L'hypothèse alternative (H_1) est le contraire (i.e., les fréquences propres sont issues de distributions différentes). Par conséquent, si l'hypothèse nulle n'est pas acceptée, il est possible de dire que le test d'hypothèse peut indiquer une modification structurale. Les tests d'hypothèse sont réalisés en considérant chaque fréquence propre, pour un niveau de bruit donné. Dans les simulations suivantes, le test de Kolmogorov-Smirnov est utilisé afin de comparer deux échantillons indépendants de données. L'avantage de ce test est qu'aucune hypothèse sur les distributions n'est faite *a priori*.

Le test de Kolmogorov-Smirnov est essentiellement basé sur le calcul des fonctions de répartition. Ces fonctions sont évaluées à partir des fonctions de densité de probabilité, déterminées pour chaque échantillon. Les échantillons utilisés dans cette étude sont les réalisations des fréquences propres identifiées pour chaque niveau d'endommagement (les variables aléatoires associées sont donc discrètes). Pour deux échantillons $E_{c,1}$ et $E_{c,2}$, les fonctions de répartition F_1 et F_2 sont déterminées. Il est alors possible de calculer la valeur « statistique du test » notée K_s définie comme l'écart maximal entre F_1 et F_2. Lorsque K_s est proche de zéro, les distributions des fré-

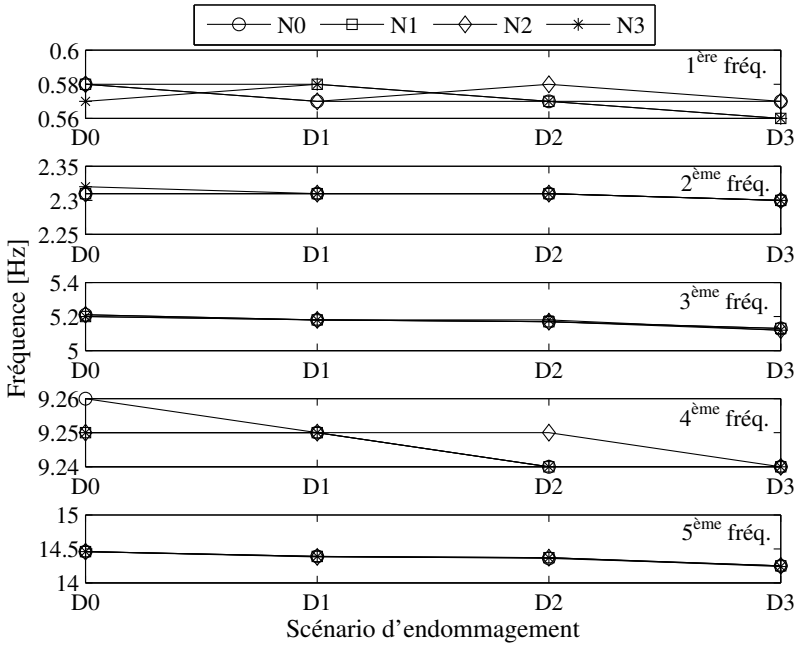


FIGURE 4.3 – Évolution des fréquences par rapport aux scénarios d'endommagement et niveaux de bruit.

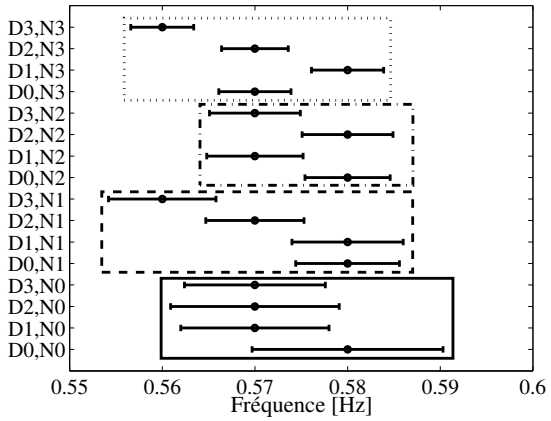
quences propres sont pratiquement égales. Dans les simulations suivantes, les résultats des tests d'hypothèse sont présentés à chaque fois en donnant la valeur de H (H= 0 lorsque l'hypothèse nulle est acceptée et H= 1 lorsqu'elle est rejetée) et de p^*, définie comme ci-dessous (Massey 1951)

$$p^* = 2\sum_{k=1}^{\infty}(-1)^{(k-1)}e^{-2k^2\lambda^2}, \text{ avec } \lambda = \sqrt{\frac{N_1N_2}{N_1+N_2}}K_s \tag{4.2}$$

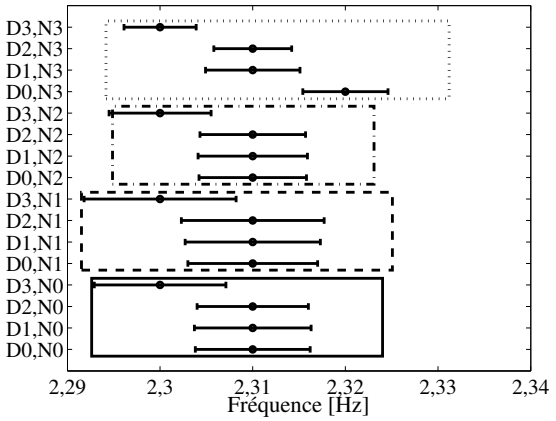
où N_1 et N_2 sont, respectivement, les tailles des échantillons $E_{c,1}$ et $E_{c,2}$. Si la valeur de p^* est supérieure au niveau d'acceptation du test, l'hypothèse nulle H_0 est acceptée. Dans le cas contraire, l'hypothèse nulle H_0 est rejetée. Le niveau d'acceptation des tests est fixé à 5%.

Le Tableau 4.2 regroupe les résultats obtenus en comparant les fréquences identifiées pour l'état intègre de la poutre (D0) aux états le moins dégradé (D1) et le plus dégradé (D3), sans bruit (n0). Contrairement à ce qui a été observé en utilisant des intervalles de confiance, les résultats ici obtenus montrent que la détection de modifications structurales à partir des distributions de certaines fréquences propres est possible. En effet, en comparant les scénarios D0 et D1, le test a échoué seulement pour les troisième et cinquième fréquences (en signifiant donc que les distributions de ces fréquences sont différentes). Toutefois, il est noté que les valeurs de p^* sont proches du niveau de signification du test, ce qui pourrait indiquer une « hésitation » à non accepter l'hypothèse nulle. Dans les cas de la première, deuxième et quatrième fréquences propres, l'hypothèse nulle est toujours acceptée. D'ailleurs, les valeurs p^* sont significativement élevées. Ceci n'est pas observé en comparant les états D0 et D3. Dans ce cas, les tests des distributions de la première, troisième et cinquième fréquences propres indiquent la non acceptation de l'hypothèse nulle. Par ailleurs, les valeurs de p^* sont plus proches de zéro, en indiquant une forte évidence de que ces scénarios sont différents.

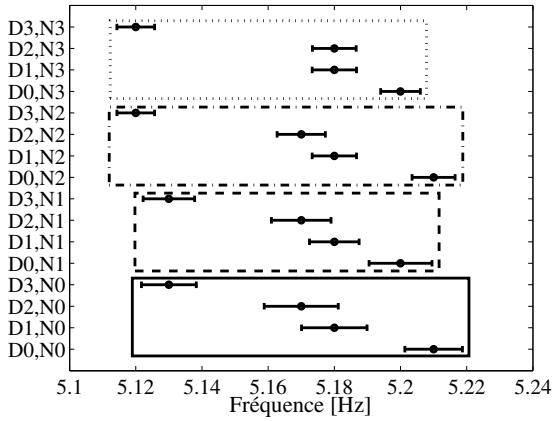
L'ajout de bruit (n3) aux mesures mène à des résultats intéressants (cf. Tableau 4.3). En comparant le scénario D0 au D3, il est noté que le test pour la troisième fréquence propre indique l'acceptation de l'hypothèse nulle (contrairement à ce qui a été observé dans le cas sans



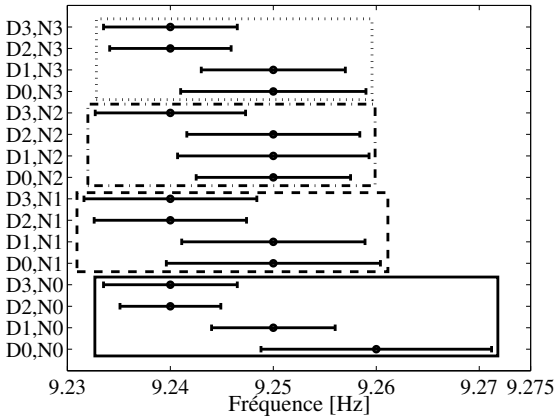
(a) Première fréquence



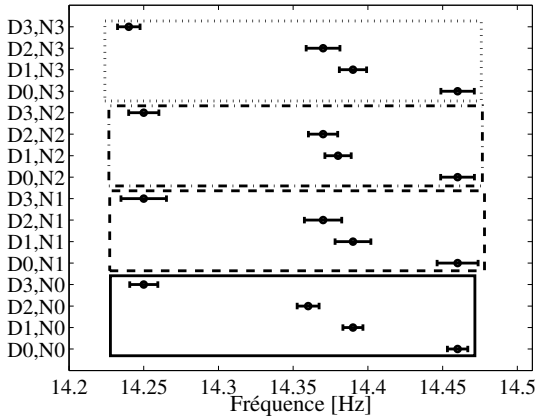
(b) Deuxième fréquence



(c) Troisième fréquence



(d) Quatrième fréquence



(e) Cinquième fréquence

FIGURE 4.4 – Intervalles de confiance (± 1 écart-type) pour les fréquences identifiées.

bruit). Inversement, en considérant les états D0 et D3, le comparatif de la deuxième fréquence propre échoue au test d'hypothèse. De manière générale, l'utilisation des tests d'hypothèse ne permet pas de tirer une conclusion globale sur l'existence d'endommagements sur la poutre.

Niveau de bruit n0										
	1ère fréq.		2ème fréq.		3ème fréq.		4ème fréq.		5ème fréq.	
	H	p^* (%)	H	p^* (%)	H	p^* (%)	H	p^* (%)	H	p^*(%)
D0-D1	0	26,8	0	12,7	1	4,6	0	8,9	1	3,5
D0-D3	1	1,3	0	5,9	1	0	0	7,7	1	0,5

TABLEAU 4.2 – Résultats des tests d'hypothèse de Kolmogorv-Smirnov pour les scénarios d'endommagement D0-D1 et D0-D3 (niveau de bruit n0).

Niveau de bruit n3										
	1ère fréq.		2ème fréq.		3ème fréq.		4ème fréq.		5ème fréq.	
	H	p^* (%)	H	p^* (%)	H	p^* (%)	H	p^* (%)	H	p^* (%)
D0-D1	0	17,9	0	5,4	0	6,1	0	6,8	1	2,0
D0-D3	1	1,2	1	4,7	1	0	0	10,3	1	0,2

TABLEAU 4.3 – Résultats des tests d'hypothèse de Kolmogorv-Smirnov pour les scénarios d'endommagement D0-D1 et D0-D3 (niveau de bruit n3).

En ce qui concerne les déformées propres, la Figure 4.5 montre une comparaison entre les moyennes des cinq premiers modes propres identifiés, en ne considérant que les cas D0 et D3, sans bruit. Il est noté que les modes 1, 3 et 5 sont les plus affectés par l'endommagement, ce qui a été signalé dans le Tableau A.1. Toutefois, quand 5% de bruit (n3) est ajouté aux signaux, les différences entre les modes propres sont relativement plus petites comme illustré à la Figure A.1. En effet, bien que les modes 1, 3 et 5 soient toujours les plus sensibles à l'endommagement, cet effet est moins visible en présence de bruit. Par ailleurs, les coefficients MAC sont calculés pour chaque paire de déformée modale. L'expression permettant de calculer le MAC pour deux déformées modales est donnée par (Allemang & Brown 1982)

$$\mathrm{MAC}_{ij} = \frac{\left|\sum_{k=1}^{n} [\boldsymbol{\phi}_A]_k^i \, [\boldsymbol{\phi}_B]_k^j\right|^2}{\sum_{k=1}^{n} \left([\boldsymbol{\phi}_A]_k^i\right)^2 \sum_{k=1}^{n} \left([\boldsymbol{\phi}_B]_k^j\right)^2} \tag{4.3}$$

où $[\boldsymbol{\phi}_A]_k^i$ et $[\boldsymbol{\phi}_B]_k^j$ sont, respectivement, les $k^{\text{ème}}$ composantes des modes $[\boldsymbol{\phi}_A]^i$ et $[\boldsymbol{\phi}_B]^j$.

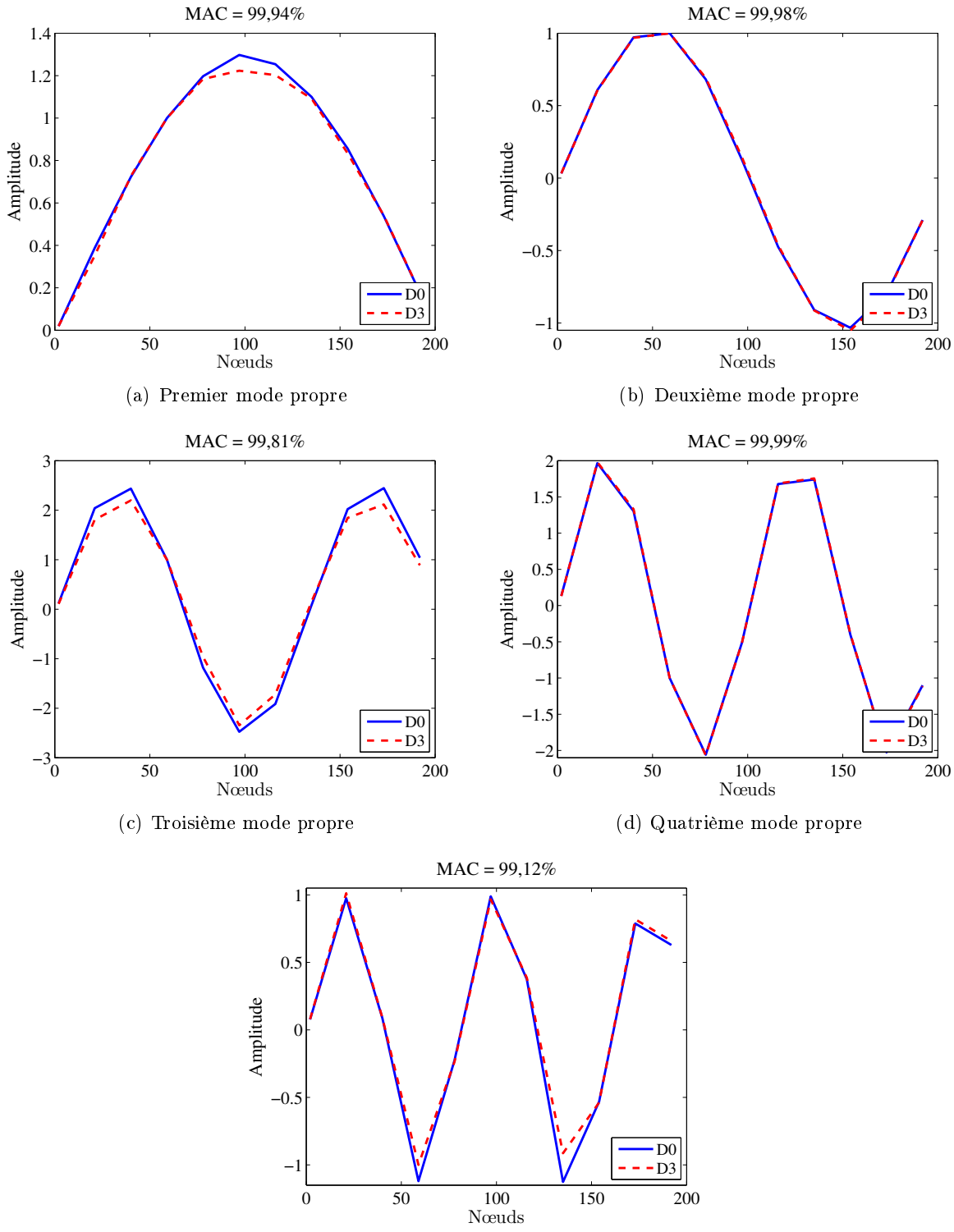
Il est noté que ce coefficient varie entre 99% et 100% pour tous les cas, ce qui ne permet pas de conclure clairement s'il existe une modification structurale ou non.

De manière générale, les analyses dites classiques montrent qu'elles ne sont pas suffisamment sensibles pour discriminer les différents états d'endommagements imposés à la structure. Par ailleurs, quand le bruit est ajouté aux mesures, il est pratiquement impossible de conclure si les différences entre les paramètres modaux sont liées aux dommages existants. Il est donc essentiel d'employer de nouvelles analyses, plus pertinentes pour la détection et la discrimination de comportements structuraux.

4.3 ADS couplée aux méthodes de clustering

4.3.1 Analyse de la sensibilité au bruit

Les résultats obtenus à partir du couplage entre l'ADS et les méthodes de classification non supervisée sont présentés dans ce paragraphe. Ces méthodes sont appliquées successivement aux représentations symboliques des signaux, des fréquences propres et finalement, des modes



(a) Premier mode propre

(b) Deuxième mode propre

(c) Troisième mode propre

(d) Quatrième mode propre

(e) Cinquième mode propre

FIGURE 4.5 – Comparaison des modes propres non endommagés et endommagés, sans bruit.

propres pour discriminer les différents états structuraux. Dans ce chapitre, deux représentations symboliques sont utilisées. La première, basée sur des intervalles interquartiles et la deuxième, sur des histogrammes à 20 catégories. Dans les analyses suivantes, seuls les résultats obtenus à

partir de la transformation en histogrammes sont donnés. Dans le cas des intervalles, les résultats sont montrés à l'Annexe A de ce mémoire.

4.3.1.1 Signaux

Les premiers résultats correspondent à l'application de la méthode des divisions hiérarchiques aux signaux, transformés en histogrammes. La Figure 4.6 illustre l'arbre de décision obtenu pour les simulations avec les signaux non bruités. L'ensemble initial comporte 40 essais, où 10 essais correspondent au cas de la poutre intègre (D0), 10 essais à celui de la poutre endommagée (D1), 10 essais au scénario D2 et finalement 10 essais au scénario D3. Il est manifeste que le partitionnement idéal est celui capable de discriminer les quatre sous-ensembles d'essais de manière homogène (i.e., obtenir quatre groupes ne contenant que des essais correspondant à un seul scénario d'endommagement). La Figure 4.6 détaille l'arbre construit en considérant les signaux non bruités (n0). Les taux de classification correcte sont adéquats : 80% pour D0, 80% pour D1, 70% pour D2 et 90% pour D3. Il est noté que les états intermédiaires ont un nombre d'essais mélangés plus élevé. Cela peut être expliqué par le fait que les niveaux d'endommagement imposés dans les scénarios D1 et D2 sont, dans un certain sens, similaires. Pour cette classification, les variables les plus discriminantes sont les capteurs 4 et 5, situés au quart de la poutre (et non les capteurs 5 et 9 comme dans le cas des signaux transformés en intervalles interquartiles ; cf. Figure A.2). Il faut toujours tenir compte que la détermination des variables les plus discriminantes est strictement basée sur la minimisation d'inerties intra-clusters. Ainsi, le type de transformation symbolique utilisée peut conduire à des résultats différents.

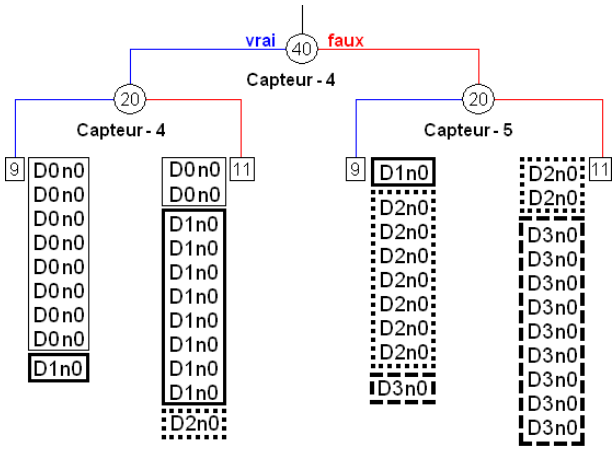


FIGURE 4.6 – Classification des signaux sans bruit (n0) utilisant les divisions hiérarchiques (histogrammes).

La même procédure de clustering est maintenant appliquée aux signaux bruités à 1% (n1). Les Figures 4.7 et A.3 (cf. Annexe A) montrent les partitionnements obtenus pour les deux types de transformation des signaux en données symboliques. Dans les deux cas, une diminution relative des taux de classification correcte et donc un nombre plus important d'essais mélangés est observé. Dans le cas d'une transformation des signaux en histogrammes, les résultats sont montrés à la Figure 4.7. Les taux de classification correcte sont 80% pour D0, 70% pour D1, 70% pour D2 et 80% pour D3. Il est également noté que les capteurs 2, 4 et 6 sont maintenant les variables les plus discriminantes.

L'analyse des résultats avec des signaux bruités à 2% est maintenant considérée. La Figure 4.8 montre l'arbre construit à partir des histogrammes. Les résultats sont relativement similaires à ceux obtenus par des intervalles interquartiles (cf. Figure A.4). Les taux sont relativement plus

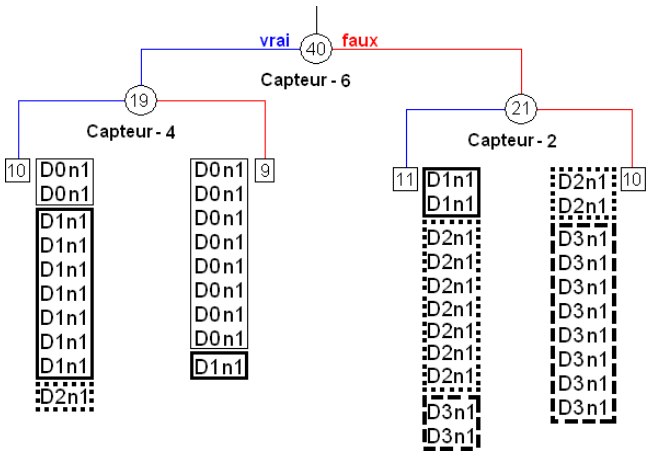


FIGURE 4.7 – Classification des signaux avec bruit (n1) utilisant les divisions hiérarchiques (histogrammes).

faibles par rapport aux analyses précédentes : 70% pour D0, 60% pour D1, 60% pour D2 et 80% pour D3. De plus, le cluster contenant essentiellement les essais D2 est très hétérogène, puisqu'il contient également des essais D0, D1 et D3. La transformation des signaux en histogrammes, appliqués aux divisions hiérarchiques, semble être plus sensible à la présence de bruit que la transformation en intervalles interquartiles.

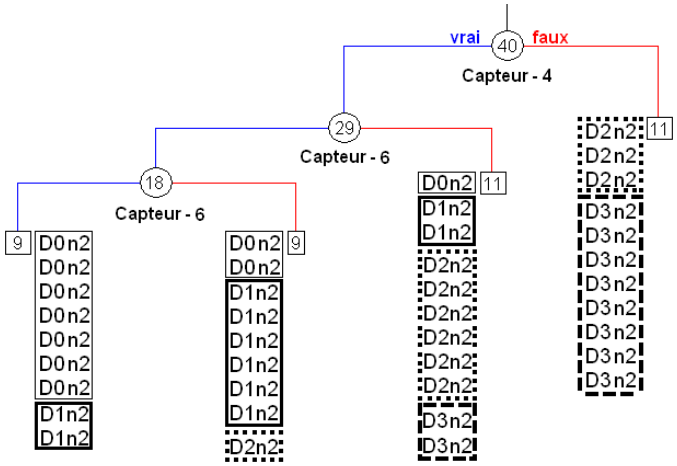


FIGURE 4.8 – Classification des signaux avec bruit (n2) utilisant les divisions hiérarchiques (histogrammes).

La Figure 4.9 montre la classification obtenue lorsque 5% (n3) de bruit sont ajoutés aux signaux. Il est intéressant de noter que le bruit a influencé la classification des essais des états D0 et D3. En effet, les taux de classification correcte correspondant à ces conditions structurales a varié significativement avec l'ajout de bruit. Dans le cas des intervalles interquartiles, les résultats restent cohérents même si la performance vis-à-vis de la classification est affectée par le bruit (cf. Figure A.5). Pour les histogrammes, toutefois, un grand nombre d'essais est mélangé. Les taux de classification correcte sont également plus faibles : 70% pour D0, 60% pour D1, 50% pour D2 et 70% pour D3.

Les premiers résultats en considérant l'ADS montrent que les taux de classification correcte tendent à diminuer lorsque le bruit dans les mesures augmente. Cette tendance est observée en particulier dans le cas où les signaux sont transformés en histogrammes. Par ailleurs, les états

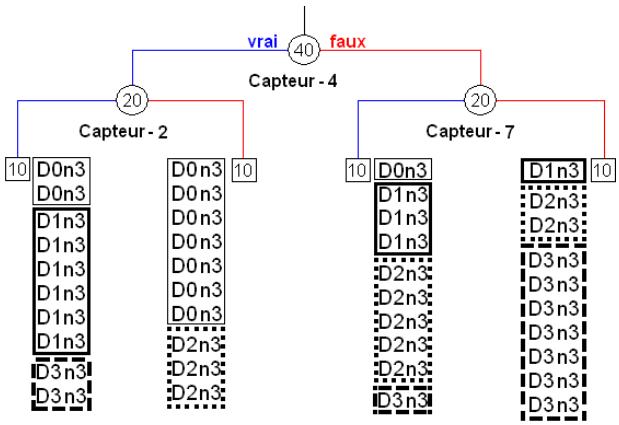


FIGURE 4.9 – Classification des signaux avec bruit (n3) utilisant les divisions hiérarchiques (histogrammes).

structuraux les plus différents (i.e., D0 et D3) ont été relativement bien séparés dans l'ensemble des simulations. Enfin, la transformation en intervalles interquartiles permet de manière générale une meilleure discrimination des essais.

La deuxième méthode de clustering utilisée dans cette application est celle basée sur les agglomérations hiérarchiques. Cette méthode se différencie des autres dans le sens où le nombre de clusters souhaité n'est pas défini au préalable. Ainsi, il est possible d'observer comment les groupes s'assemblent, ce qui permet d'analyser leur « proximité ». Dans le premier cas, sans bruit (n0) et en considérant les signaux transformés en histogrammes, les taux de classification correcte sont adéquats pour les scénarios D0 et D3 (90%), mais relativement faibles pour les états D1 et D2 : 50% et 60%, respectivement (Figure 4.10). Les essais sont donc plus mélangés que dans le cas où le partitionnement est effectué avec la méthode des divisions hiérarchiques. Il se peut que le critère de distance utilisé ne soit pas suffisamment sensible pour discriminer les différents états structuraux. Il est cependant important de remarquer que pour le scénario D3, tous les essais sont classés dans le même groupe.

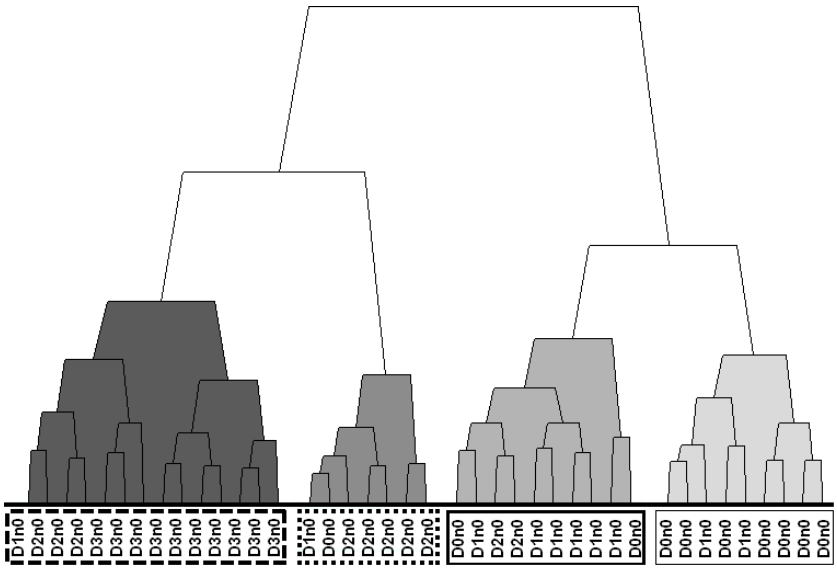


FIGURE 4.10 – Classification des signaux sans bruit (n0) utilisant les agglomérations hiérarchiques (intervalles).

Comme dans le cas de la méthode des divisions hiérarchiques, la procédure de partitionnement est réalisée en considérant les signaux bruités à 1% (n1). Le partitionnement obtenu en transformant les signaux en histogrammes est montré à la Figure 4.11. Dans ce cas, une diminution des pourcentages est observée pratiquement pour tous les clusters, sauf dans le cas du cluster des essais D2, pour lequel le taux de classification correcte augmente de 50% a 60%. L'agglomération est essentiellement la même que celle obtenue en utilisant des intervalles interquartiles (cf. Figure A.7), mais la quantité d'essais mélangés est plus importante (notamment pour les groupes D1 et D3).

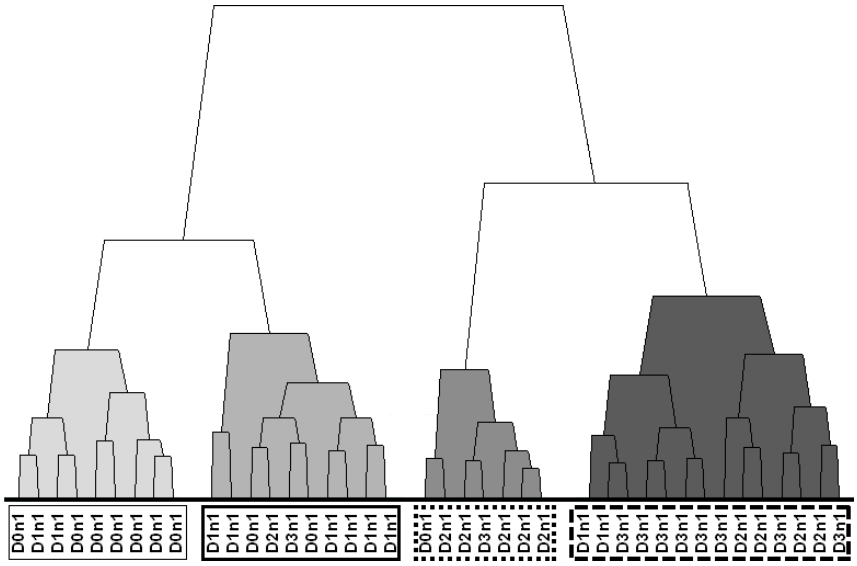


FIGURE 4.11 – Classification des signaux avec bruit (n1) utilisant les agglomérations hiérarchiques (histogrammes).

La simulation suivante consiste à appliquer 2% de bruit (n2) aux signaux. Dans ce cas, les taux de classification correcte ont diminué pour tous les scénarios. Un partitionnement moins adéquat est obtenu en utilisant des histogrammes. Les pourcentages de classification correcte sont de 60% pour D0 et D2 et 70% pour D1 et D3 (Figure 4.12).

Le dernier niveau de bruit est maintenant considéré (n3). Il est noté la méthode basée sur les agglomérations hiérarchiques est plus sensible au bruit que celle basée sur des divisions hiérarchiques. Les pourcentages de classification correcte sont plus faibles pour les deux types de transformation en données symboliques. Lorsque des histogrammes sont considérés, les résultats sont légèrement meilleurs : 60% de classification correcte pour les états D0, D1 et D3, et 50% pour D2. Cependant, de nombreux essais sont mélangés dans les groupes D1 et D2 (Figure 4.13). Deux essais D3 apparaissent dans le cluster D0. Cela montre que le bruit peut effectivement mener à une classification incorrecte.

De manière générale, les résultats obtenus en utilisant les agglomérations hiérarchiques ont montré que cette méthode est plus sensible au bruit que les divisions hiérarchiques. Bien que cette démarche réussisse à séparer relativement bien les états structuraux, les classifications obtenues ne sont jamais parfaites, même dans les cas où il n'y a aucun bruit. Chaque méthode possède ses particularités et sensibilités différentes (notamment pour le calcul des distances). Il est donc raisonnable de toujours considérer l'ensemble des résultats obtenus et d'analyser les similarités (et différences) entre eux.

Afin de compléter l'analyse des méthodes de clustering appliquées à l'exemple considéré, la méthode des nuées dynamiques est maintenant considérée. Comme décrit précédemment, pour chaque niveau de bruit, l'ensemble des 40 essais est séparé en 4 groupes où chacun représente

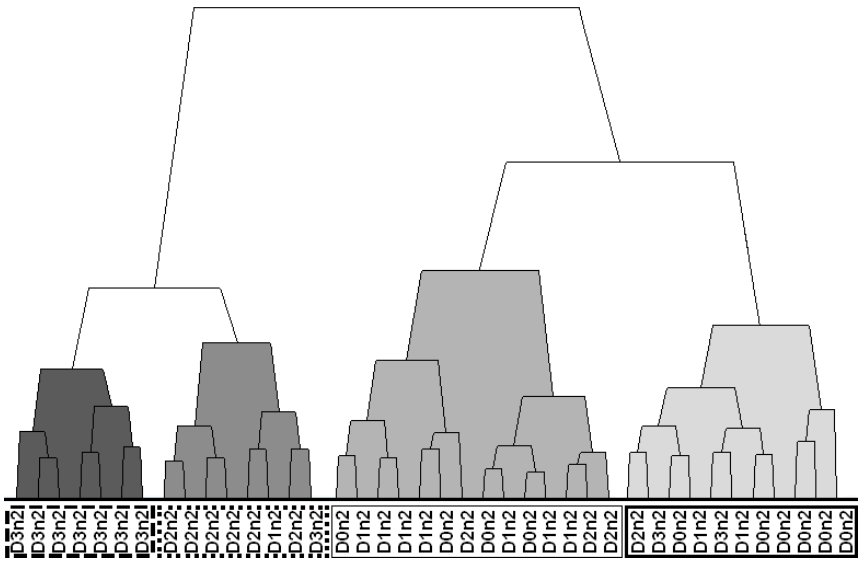


FIGURE 4.12 – Classification des signaux avec bruit (n2) utilisant les agglomérations hiérarchiques (histogrammes).

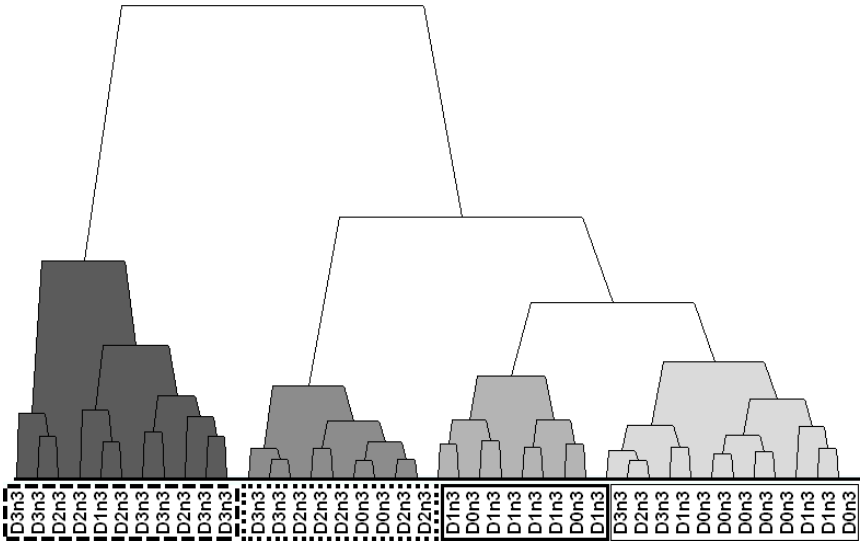


FIGURE 4.13 – Classification des signaux avec bruit (n3) utilisant les agglomérations hiérarchiques (histogrammes).

un scénario d'endommagement. Le Tableau 4.4 regroupe les taux de classification obtenus pour toutes les simulations de bruit réalisées, utilisant des histogrammes. Dans ce tableau, chaque ligne correspond à un niveau de bruit, lorsque les colonnes sont les scénarios d'endommagement.

La méthode des nuées dynamiques permet d'obtenir des meilleurs résultats que ceux obtenus en utilisant les divisions hiérarchiques ou les agglomérations hiérarchiques. Dans cet exemple, la méthode des nuées dynamiques est plus performante et moins sensible à la présence de bruit. Il est également intéressant d'observer que les taux de classification obtenus en utilisant des histogrammes sont maintenant plus élevés que ceux obtenus par des intervalles interquartiles (cf. Tableau A.5). Il est manifeste qu'à mesure que le niveau de bruit augmente, les taux de classification correcte diminuent. Pour les scénarios d'endommagement D0 et D3 (colonnes 1 et 4 du Tableau 4.4) la sensibilité au bruit semble être cependant moins flagrante. Comme mentionné précédemment, « l'écart » entre les états D0 et D3 permet de les séparer plus facilement, même

	Histogrammes			
	D0	D1	D2	D3
n0	100	90	90	80
n1	80	80	80	80
n2	90	80	70	70
n3	80	60	70	70

TABLEAU 4.4 – Taux de classification correcte obtenus en utilisant les signaux appliqués aux nuées dynamiques (valeurs en %).

en présence de bruit. Inversement, les états D1 et D2, ayant un état de dégradation plus proches, sont relativement plus difficiles à séparer lorsque le niveau de bruit augmente.

4.3.1.2 Fréquences propres

Le couplage entre l'ADS et les méthodes de clustering est maintenant étudié lorsque les fréquences propres, préalablement identifiées, sont considérées. Cette étude consiste donc à transformer les cinq premières fréquences propres en données symboliques et de les utiliser comme données d'entrée dans les méthodes de clustering. La méthode des divisions hiérarchiques est considérée en premier. Pour toutes les simulations de bruit, 100% de classification correcte est obtenue pour les quatre scénarios d'endommagement. Ce résultat est important car il montre que les fréquences propres transformées en données symboliques peuvent être utilisées comme des indicateurs adéquats pour détecter des modifications structurales. La Figure 4.14 montre les arbres construits pour le dernier niveau de bruit, n3 (les arbres obtenus pour les autres niveaux de bruit ne sont pas montrés, car identiques à ceux obtenus pour le niveau de bruit n3). Bien que la seule différence entre les arbres soient les variables discriminantes, le partitionnement obtenu est toujours le même. Il est également important de noter que les états D0 et D3 sont séparés dans des clusters différents dès le premier critère de décision, alors que les clusters D1 et D2 sont les derniers à être séparés.

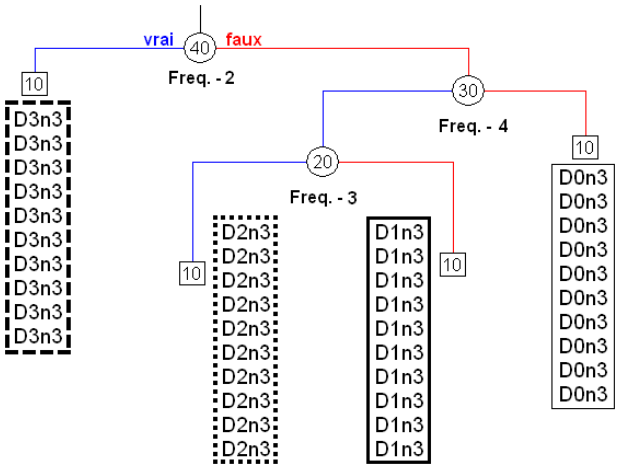


FIGURE 4.14 – Classification des fréquences propres avec bruit (n3) utilisant les divisions hiérarchiques (histogrammes).

La méthode des agglomérations hiérarchiques est maintenant considérée. Comme dans le cas des divisions hiérarchiques, 100% des fréquences propres sont classées correctement, quel que soit le niveau de bruit ou le type de transformation en données symboliques utilisé. Les sorties graphiques de cette méthode ne sont pas montrées, par soucis de clarté du texte.

Finalement, la procédure de clustering est réalisée en utilisant la méthode des nuées dynamiques. Comme il a été observé pour les autres méthodes, pour tous les niveaux de bruit, la classification obtenue est parfaite (100% de classification correcte). Toutefois, ceci n'est pas observé dans le cas des intervalles interquartiles où la classification est parfaite uniquement dans le cas où il n'y a aucun bruit (cf. Tableau A.6).

4.3.1.3 Modes propres

La troisième partie de cette étude considère les cinq premiers modes propres transformés en données symboliques. Dans un premier temps, ces quantités transformées en histogrammes sont appliquées à la méthode des divisions hiérarchiques. Une classification adéquate est obtenue : les états D0 et D3 permettent d'atteindre 100% de classification correcte (Figure 4.15) alors que les états D1 et D2 tendent encore à se mélanger. Il est intéressant de noter que les variables les plus discriminantes sont les modes 2 et 4 (ceux qui ont des coefficients MAC les plus élevés. Ceci montre que les critères de séparation des essais ne suit pas nécessairement les résultats obtenus par les méthodes dites classiques.

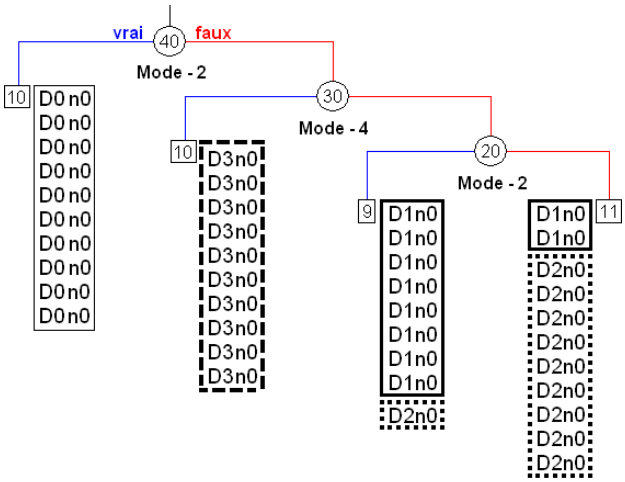


FIGURE 4.15 – Classification des modes propres sans bruit (n0) utilisant les divisions hiérarchiques (histogrammes).

Les simulations concernant le premier niveau de bruit (n1) sont maintenant considérées. De façon similaire à celle qui a été observée à partir de l'utilisation directe des signaux, le nombre de classifications correctes diminue avec la présence de bruit. Dans le cas des histogrammes, une classification relativement adéquate est obtenue : 80% pour D0 et D1, 90% pour D2 et 100% pour D3 (cf. Figure 4.16). Ceci n'est pas observé pour des intervalles (cf. Figure A.12).

Des simulations avec 2% de bruit sont également réalisées. Les Figures 4.17 et A.13 montrent les arbres construits pour chacune des transformations considérées. Pour les deux transformations, une légère diminution des taux d'essais correctement classés est constatée. Il apparaît clairement que l'ajout de bruit cause cette diminution. Il semble également manifeste que les signaux et modes propres sont plus sensibles que les fréquences propres dans ce type d'analyse. Ici, 80% des essais sont correctement classés dans les clusters D0 et D3 alors que 70% ne le sont pour D1 et D2.

Pour le troisième niveau de bruit, seule la classification utilisant des intervalles interquartiles mène à des résultats moins adéquats (cf. Figure A.14). Toutefois, quand la transformation en histogrammes est considérée, l'augmentation du niveau de bruit n'altère pas les taux de classifications obtenus par rapport au niveau précédent : 80% pour les scénarios D0 et D3. La seule

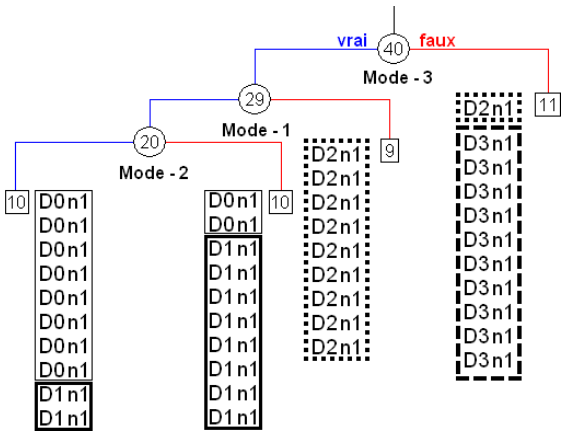


FIGURE 4.16 – Classification des modes propres avec bruit (n1) utilisant les divisions hiérarchiques (histogrammes).

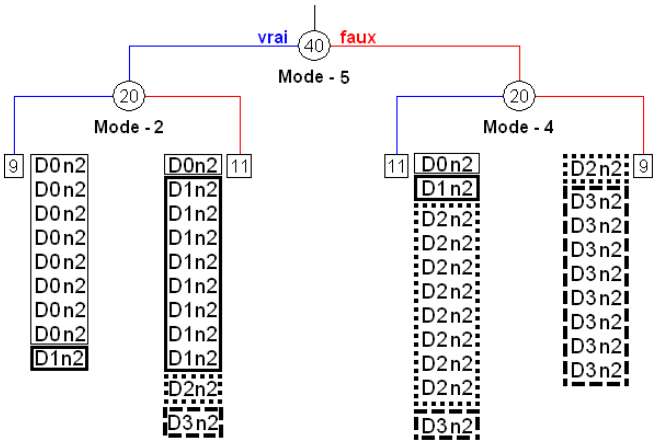


FIGURE 4.17 – Classification des modes propres avec bruit (n2) utilisant les divisions hiérarchiques (histogrammes).

différence est que plus d'essais se mélangent dans le groupe D1, comme il est montré à la Figure 4.18.

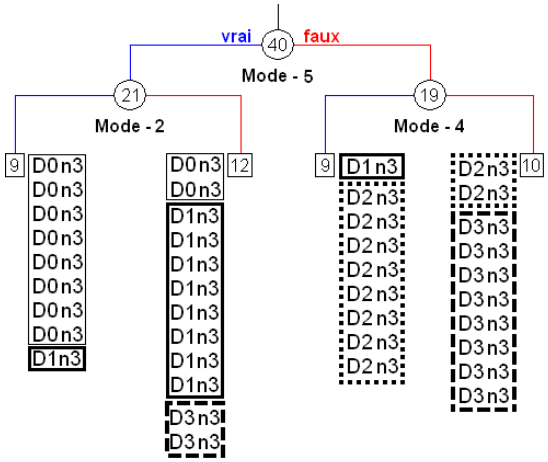


FIGURE 4.18 – Classification des modes propres avec bruit (n3) utilisant les divisions hiérarchiques (histogrammes).

La méthode des agglomérations hiérarchiques est maintenant appliquée aux modes propres identifiés à partir des signaux non bruités. Quand la transformation en histogrammes est considérée, 100% de classification correcte est obtenue pour tous les scénarios (la figure n'est pas présentée pour ne pas alourdir la lecture du mémoire). Les Figures A.16 et 4.19 illustrent les résultats obtenus pour le niveau de bruit n1. Il est constaté que les modes propres sont plus sensibles au bruit que les fréquences propres. Pour des histogrammes, la classification est meilleure, avec 100% de réussite pour les états D1 et D3, et 90% pour les états D0 et D2. Cependant, il existe un essai D0 qui a été classé dans le cluster D3. Ceci peut être encore expliqué par la présence de bruit dans les mesures considérées.

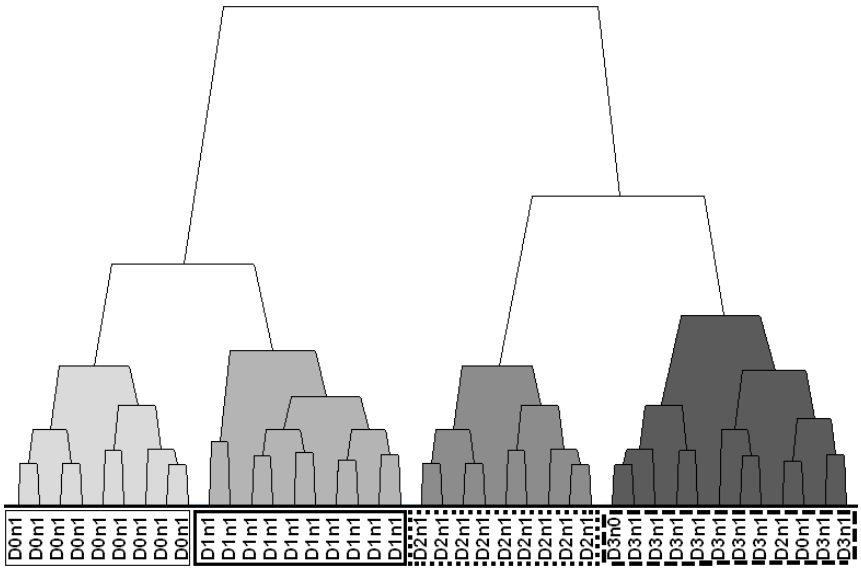


FIGURE 4.19 – Classification des modes propres avec bruit (n1) utilisant les agglomérations hiérarchiques (histogrammes).

Les simulations avec le niveau de bruit n2 sont également considérées. Dans ce cas, plusieurs essais se mélangent, notamment dans les groupes D1 et D2. De façon similaire au cas précédent, les résultats obtenus en utilisant des histogrammes sont meilleurs. Bien que 80% d'essais aient été correctement classés dans tous les groupes, il existe plusieurs essais de type D0 et D2 qui sont classés incorrectement dans le groupe D1 (cf. Figure 4.20). Étant donné qu'ils représentent des conditions structurales « proches », il est compréhensible que la plupart des erreurs de classification se fassent à ce niveau.

Les agglomérations construites pour le dernier niveau de bruit sont présentées aux Figures A.18 et 4.21. Dans le cas des intervalles interquartiles (cf. Figure A.18), plusieurs essais sont mélangés. Cela caractérise l'influence du bruit dans la classification où même que cette méthode n'est pas suffisamment sensible pour discriminer ces états. Toutefois, dans le cas où des histogrammes sont utilisés, cette sensibilité est moins flagrante. En effet, les taux de classification correcte sont relativement plus importants : 80% pour les états D0 et D3, 70% pour D1 et 60% pour D2. Dans ce dernier cas, il est noté que des essais D2 et D3 sont classés dans le groupe D0.

La méthode de clustering basée sur les nuées dynamiques est maintenant appliquée aux modes propres. Le Tableau 4.5 détaille les taux de classification obtenus. Bien que les valeurs ne soient pas aussi élevées comme celles obtenues en utilisant les fréquences propres, des pourcentages adéquats de classification correcte sont obtenus. Par ailleurs, il est noté que la classification utilisant des histogrammes est relativement plus robuste que celles avec des intervalles interquartiles (cf. Tableau 4.5).

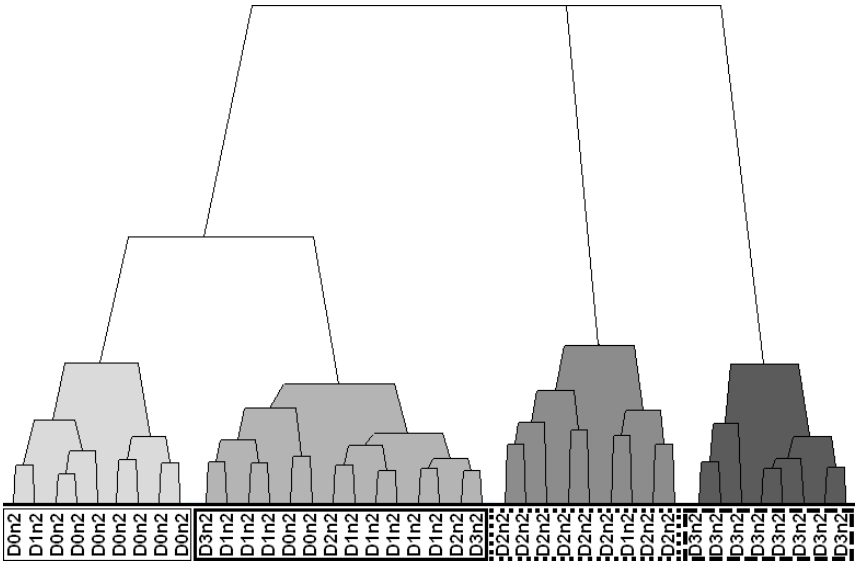


FIGURE 4.20 – Classification des modes propres avec bruit (n2) utilisant les agglomérations hiérarchiques (histogrammes).

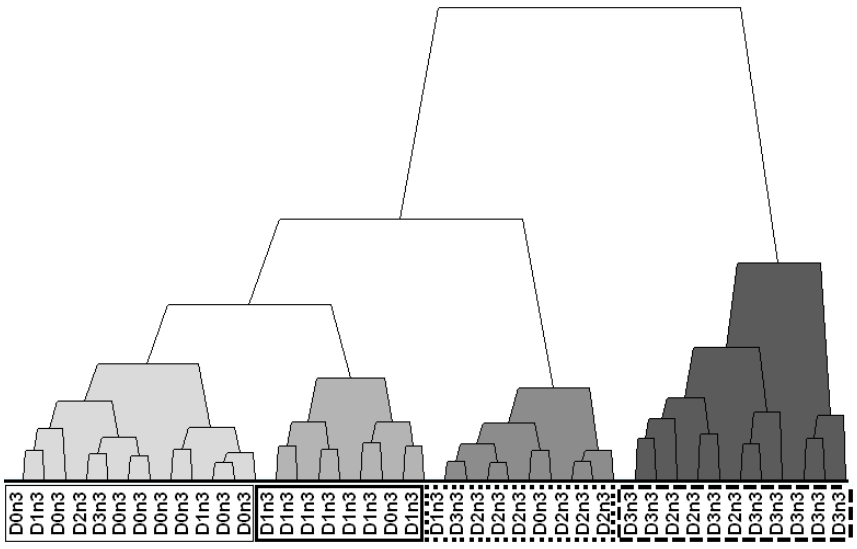


FIGURE 4.21 – Classification des modes propres avec bruit (n3) utilisant les agglomérations hiérarchiques (histogrammes).

	Histogrammes			
	D0	D1	D2	D3
n0	100	100	100	100
n1	90	100	90	90
n2	90	80	90	90
n3	80	70	90	90

TABLEAU 4.5 – Taux de classification correcte obtenus en utilisant les modes propres appliqués aux nuées dynamiques (valeurs en %).

4.3.1.4 Bilan

Afin de synthétiser et mieux comparer les résultats obtenus par les différentes méthodes de partitionnement, les taux de classification correcte sont détaillés dans la suite. La Figure A.19 (cf. Annexe A) regroupe les pourcentages obtenus pour chaque méthode de partitionnement lorsque les signaux sont directement utilisés et convertis soit en intervalles interquartiles ou en histogrammes. Pour chacune de ces figures, les taux de classification sont regroupés par niveau de bruit. De cette façon, chaque barre verticale correspond à un niveau d'endommagement différent. Les Figures A.19(a) et A.19(b) correspondent à la méthode des divisions hiérarchiques appliquées aux signaux transformés en intervalles interquartiles et histogrammes, respectivement. De manière générale, il est noté que la transformation en intervalles interquartiles, pour cette méthode, produit des résultats légèrement meilleurs, compte tenu de tous les niveaux de bruit. Ceci est également observé pour la méthode des agglomérations hiérarchiques (Figures A.19(c) et A.19(d)). Toutefois, des variations importantes sont indiquées, principalement pour le premier niveau de bruit. Cela montre que cette méthode tend à mélanger plus fréquemment les états structuraux, même quand le bruit n'est pas ajouté aux mesures. Finalement, les taux de classification correcte obtenus par la méthode des nuées dynamiques sont montrés aux Figures A.19(e) et A.19(f). Dans cette étude, la méthode des nuées dynamiques produit les meilleurs taux de classification parmi les méthodes considérées. Contrairement à ce qui a été observé pour les deux méthodes précédentes, l'utilisation des signaux transformés en histogrammes mène à une meilleure discrimination des scénarios d'endommagement.

Les résultats obtenus en utilisant les fréquences propres ont montré que leur application est la plus robuste comparée aux signaux et aux modes propres. Pour toutes les méthodes, 100% de classification correcte a été obtenue, à l'exception de celle basée sur les nuées dynamiques appliquée à des intervalles interquartiles.

Les Figures A.20(a) et A.20(b) illustrent les pourcentages obtenus utilisant les divisions hiérarchiques appliquées aux modes propres transformés en intervalles et en histogrammes, respectivement. Contrairement à ce qui a été remarqué dans l'analyse avec les signaux, la représentation en histogrammes produit des résultats légèrement meilleurs, surtout pour les deux derniers niveaux de bruit. Ceci est également observé pour la méthode des agglomérations hiérarchiques. Par ailleurs, des variations relativement élevées sont observées dans le cas de la transformation en intervalles interquartiles (Figures A.20(c) et A.20(d)). Finalement, pour la méthode des nuées dynamiques, des meilleurs taux sont obtenus, notamment dans le cas de la transformation en histogrammes (Figures A.20(e) et A.20(f)).

Pour conclure, la Figure 4.22 synthétise la moyenne globale (tous les niveaux d'endommagement et bruit confondus) des taux de classification pour chaque méthode et chaque type de transformation utilisée. Dans cette analyse, la méthode des nuées dynamiques a montré être la plus performante pour discriminer les scénarios d'endommagement. De plus, son application aux fréquences propres a permis de montrer la robustesse et l'efficacité de l'approche proposée. Cette méthode, appliquée aux modes propres, a aussi produit de résultats adéquats. La méthode des divisions hiérarchiques a également permis d'obtenir des classifications pertinentes. De façon générale, la méthode des agglomérations hiérarchiques est la moins performante. Bien qu'elle ait réussi à avoir des classifications adéquates en utilisant les fréquences propres, les résultats obtenus à partir des signaux et des modes propres sont moins pertinents. En effet, pour ces deux derniers, il a été observé un taux important de classification incorrecte (notamment pour les deux états de dégradation intermédiaires). Les clusters obtenus ont rarement permis de distinguer parfaitement les différents états structuraux.

Il est important de signaler que l'étude de l'ensemble de méthodes est essentielle pour la discrimination adéquate d'un processus d'endommagement. Bien qu'elles ne réussissent pas tou-

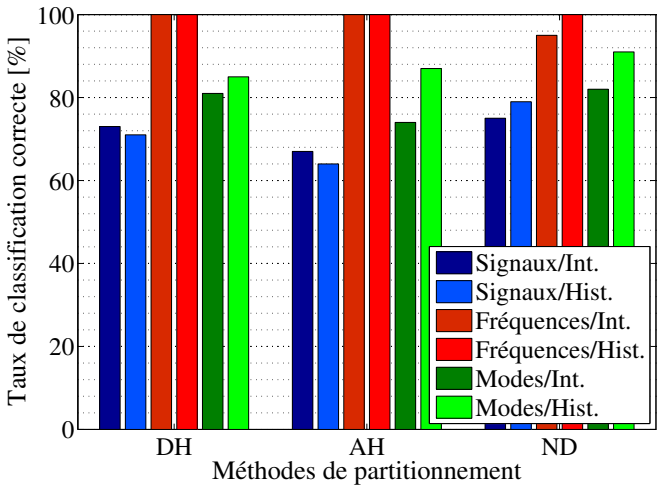


FIGURE 4.22 – Synthèse des résultats obtenus pour les méthodes des divisions hiérarchiques, agglomérations hiérarchiques et nuées dynamiques.

jours à discriminer parfaitement des états structuraux, elles permettent d'observer des relations de proximité (ou de distance) entre les essais. De plus, l'utilisation de différents types de données symboliques peut fournir des informations qui aident à mieux interpréter le cas étudié.

4.3.2 Études complémentaires

Trois analyses complémentaires sont maintenant considérées afin de mieux appréhender les effets du bruit sur les mesures et paramètres modaux. La première analyse consiste à utiliser l'ensemble des essais de type D0 et D3 avec tous les niveaux de bruit confondus. Cela produit donc un ensemble de 80 essais (40 pour D0 et 40 pour D3). Une analyse similaire est ensuite réalisée en ne considérant que les états structuraux D1 et D2. Finalement, l'ensemble de 160 essais est utilisé. Dans ce cas, tous les niveaux d'endommagement et de bruit sont considérés. Le but de ces analyses est de discriminer les différents états structuraux indépendamment du niveau de bruit existant.

Dans un premier temps, les états les plus « différents » D0 et D3 (représentant la structure non dégradée et dégradée, respectivement) sont pris en compte. Le Tableau 4.6 regroupe les taux de classification correcte obtenus par chacune des méthodes de partitionnement. Les résultats confirment la tendance observée dans les simulations précédentes quand un seul niveau de bruit a été considéré séparément. Ici, les divisions hiérarchiques et nuées dynamiques sont les méthodes qui arrivent à mieux discriminer les états D0 et D3. Si les fréquences propres sont considérées, la discrimination est parfaite quelle que soit la méthode. La classification obtenue en utilisant les signaux montre qu'il existe un nombre plus important d'essais mélangés. Cela est principalement dû au fait que la plupart des essais bruités à 5% se mélangent plus fréquemment.

L'analyse suivante consiste à regrouper tous les essais de type D1 et D2. Étant donné que ces états sont plus « proches » (comparés aux états D0 et D3), il est possible que les taux de classification correcte tendent être moins élevés. Les résultats sont regroupés dans le Tableau 4.7. En effet, les résultats obtenus en utilisant les signaux ne sont pas adéquats (la moitié des essais est classée correctement, environ). Dans cette étude, l'utilisation des nuées dynamiques appliquées aux fréquences propres ne permet pas d'obtenir une classification parfaite. Au contraire, les meilleurs taux obtenus sont de l'ordre de 80%. Ceci implique que pour des conditions structurales très proches, il peut ne pas être possible de détecter un endommagement en utilisant les méthodes

Divisions hiérarchiques			
	Signaux	Fréquences propres	Modes propres
D0	65	100	100
D3	70	100	100
Agglomérations hiérarchiques			
	Signaux	Fréquences propres	Modes propres
D0	48	100	78
D3	58	100	80
Nuées dynamiques			
	Signaux	Fréquences propres	Modes propres
D0	70	100	100
D3	65	100	100

TABLEAU 4.6 – Taux de classification correcte obtenus en utilisant les méthodes de partitionnement pour les états D0 et D3 (valeurs en %).

de clustering. En général, la méthode des divisions hiérarchiques arrive à avoir des taux de réussite de l'ordre de 70%, ce qui est plus élevé que les pourcentages obtenus par les agglomérations hiérarchiques (de l'ordre de 60%).

Divisions hiérarchiques			
	Signaux	Fréquences propres	Modes propres
D1	53	75	65
D2	50	78	75
Agglomérations hiérarchiques			
	Signaux	Fréquences propres	Modes propres
D1	43	70	70
D2	48	65	73
Nuées dynamiques			
	Signaux	Fréquences propres	Modes propres
D1	55	80	73
D2	63	80	78

TABLEAU 4.7 – Taux de classification correcte obtenus en utilisant les méthodes de partitionnement pour les états D1 et D2 (valeurs en %).

Finalement, une troisième analyse considérant les 160 essais est réalisée. L'objectif est toujours de séparer les essais en quatre groupes selon le scénario d'endommagement simulé. Le Tableau 4.8 regroupe les pourcentages de classification correcte calculés. Il est intéressant de noter que dans ce cas, les taux obtenus en utilisant des fréquences propres remontent à nouveau. Pour pratiquement tous les scénarios d'endommagement, 100% de classification correcte est obtenue. En effet, la procédure de clustering est différente si deux ou quatre clusters sont considérés. La modification est essentiellement liée à la définition de la variable discriminante (et de la valeur de coupure associée). En ce sens, il est parfaitement possible qu'un partitionnement avec 4 groupes mène à un classement plus satisfaisant qu'un autre utilisant 2 clusters. Bien que la méthode des nuées dynamiques soit toujours la plus performante, les divisions hiérarchiques arrivent aussi à avoir des pourcentages adéquats de classification correcte. Toutefois, aucune de ces méthodes n'arrive à classer plus de 50% des essais correctement quand les signaux sont considérés. En effet, cela montre que la présence de bruit dans les mesures dynamiques peut conduire à des classifications incorrectes.

Divisions hiérarchiques			
	Signaux	Fréquences	Modes
D0	43	100	80
D1	35	98	78
D2	33	100	78
D3	35	100	80
Agglomérations hiérarchiques			
	Signaux	Fréquences propres	Modes propres
D0	38	100	75
D1	40	100	73
D2	35	100	73
D3	38	100	70
Nuées dynamiques			
	Signaux	Fréquences propres	Modes propres
D0	35	100	78
D1	43	100	78
D2	33	100	80
D3	38	100	83

TABLEAU 4.8 – Taux de classification correcte obtenus en utilisant les méthodes de partitionnement pour les états D0, D1, D2, D3 (valeurs en %).

4.3.3 Indices de partitionnement optimal

Afin de compléter l'étude du couplage de l'ADS aux méthodes de clustering, une recherche des indices de partitionnement optimal, présentés au paragraphe 3.1.5, est proposée. L'objectif est de vérifier la capacité de ces indices à indiquer le nombre correct de comportements structuraux simulés en considérant différents niveaux de bruit. Comme expliqué précédemment, les méthodes des divisions hiérarchiques et des nuées dynamiques nécessitent la définition du nombre de clusters existants au préalable, contrairement à la méthode des agglomérations hiérarchiques. Par conséquent, seules ces deux méthodes sont considérées dans les études suivantes.

La démarche utilisée consiste à calculer les indices pour différents nombres de clusters. Ici, le nombre maximal de clusters est défini arbitrairement comme étant égal à 10. De cette façon, il est considéré que le nombre optimal de clusters se trouve dans l'intervalle $[2, 10]$. Les indices CH, C^* et Γ sont évalués pour chaque niveau de bruit, l'objectif étant d'observer comment le bruit affecte la détermination du nombre correct de clusters.

4.3.3.1 Signaux

Le Tableau 4.9 regroupe les valeurs calculées lorsque les signaux sont transformés en histogrammes. À l'exception du niveau de bruit n3, les indices CH et Γ signalent toujours 4 clusters. Comme dans le cas des signaux transformés en intervalles interquartiles (cf. Tableau A.11), l'indice C^* varie beaucoup, ayant des valeurs entre 6 et 9 clusters. Il est rappelé que les 4 clusters correspondent aux scénarios d'endommagements D0, D1, D2 et D3.

Une démarche similaire est menée en appliquant la méthode des nuées dynamiques. Le Tableau 4.10 regroupe les valeurs de chaque indice dans le cas où les signaux sont transformés en histogrammes. Comme remarqué précédemment, les indices CH et Γ indiquent correctement quatre clusters pour tous les niveaux de bruit (à l'exception du niveau n1, pour Γ). Par ailleurs,

	n0			n1			n2			n3		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
10	11,3	0,05	0,75	9,5	0,13	0,58	25,3	0,03	0,47	19,1	0,04	0,64
9	10,5	**0,03**	0,75	10,2	**0,06**	0,61	47,0	**0,01**	0,52	20,4	0,03	0,63
8	10,3	0,06	0,77	2,2	0,13	0,61	44,6	0,02	0,51	14,0	0,03	0,59
7	11,9	0,07	0,74	10,3	0,13	0,56	24,8	0,06	0,54	23,5	0,03	0,60
6	13,5	0,06	0,68	10,0	0,17	0,49	46,0	0,03	0,47	25,2	**0,02**	0,58
5	20,4	0,07	0,86	27,7	0,07	0,76	38,6	0,07	0,41	**50,8**	0,05	**0,89**
4	**25,3**	0,09	**0,88**	**36,3**	0,18	**0,81**	**57,2**	0,05	**0,61**	40,2	0,04	0,72
3	16,6	0,11	0,78	11,8	0,10	0,74	51,9	0,16	0,40	20,0	0,08	0,73
2	15,4	0,18	0,75	18,5	0,03	0,55	10,4	0,21	0,33	26,9	0,08	0,74

TABLEAU 4.9 – Indices basés sur les divisions hiérarchiques appliquées aux signaux transformés en histogrammes.

l'indice Γ a des valeurs très proches pour le niveau de bruit n3. L'indice C^* oscille beaucoup et n'indique jamais le nombre correct de groupes.

	n0			n1			n2			n3		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
10	15,6	3,0	0,78	11,1	5,7	0,69	10,2	7,1	0,80	11,3	4,7	0,74
9	13,6	4,7	0,74	11,5	**4,6**	0,63	9,6	6,3	0,76	15,0	**3,1**	0,81
8	19,7	3,3	0,77	12,8	5,7	0,61	12,2	5,5	0,75	15,0	3,8	0,82
7	17,8	3,9	0,79	9,7	9,9	0,65	12,8	5,6	0,78	14,2	5,2	0,83
6	20,2	**2,6**	0,78	16,2	5,5	**0,73**	14,1	7,3	0,77	16,0	5,8	0,83
5	20,8	8,1	0,74	15,1	10,9	0,55	15,9	**6,4**	0,79	16,9	6,6	0,81
4	**22,9**	5,3	**0,80**	**17,1**	8,3	0,72	**16,6**	6,5	**0,84**	**20,9**	7,1	**0,84**
3	22,6	11,9	0,68	16,5	19,5	0,57	12,3	7,5	0,78	19,2	12,1	0,74
2	15,7	28,8	0,50	13,8	20,4	0,55	10,2	10,6	0,77	18,0	21,9	0,67

TABLEAU 4.10 – Indices basés sur les nuées dynamiques appliquées aux signaux transformés en histogrammes.† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

4.3.3.2 Fréquences propres

L'analyse suivante consiste à évaluer les indices lorsque les fréquences propres sont considérées. Le Tableau 4.11 regroupe les valeurs calculées. Dans ce cas, il est noté un bon accord entre les trois indices et tous indiquent correctement le nombre de clusters, quel que soit le niveau de bruit considéré. En effet, l'utilisation des fréquences propres a toujours mené à des meilleurs résultats pour ce type d'analyse.

La méthode des nuées dynamiques est maintenant considérée. Les résultats obtenus en utilisant la transformation en histogrammes sont présentés dans le Tableau 4.12. Dans ce cas, tous les indices signalent correctement quatre groupes comme partitionnement optimal.

4.3.3.3 Modes propres

L'étape suivante de cette analyse consiste à évaluer les indices lorsque les modes propres sont utilisés. Dans un premier temps, ces indices sont calculés à partir des partitionnements obtenus par la méthode des divisions hiérarchiques. Le Tableau 4.13 regroupe les valeurs calculées. Il est

	n0			n1			n2			n3		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
10	1547	5,7	0,54	1053	3,4	0,51	1201	1,4	0,47	3537	3,9	0,48
9	1750	3,7	0,69	814	2,9	0,70	1169	1,4	0,54	2469	2,6	0,53
8	1703	3,6	0,66	1016	3,3	0,69	1162	1,7	0,60	2692	2,1	0,61
7	1847	3,8	0,78	1183	2,6	0,74	1274	2,1	0,66	2456	1,6	0,68
6	2247	4,8	0,82	1481	1,9	0,77	1473	1,1	0,68	2199	1,4	0,70
5	2443	7,5	0,86	1544	2,8	0,80	1397	1,0	0,73	1349	1,0	0,76
4	**2637**	**1,9**	**0,98**	**1631**	**5,6**	**0,86**	**1739**	**0,9**	**0,89**	**4933**	**0,7**	**0,95**
3	1445	2,4	0,77	714	7,3	0,65	1232	1,5	0,75	4129	1,2	0,82
2	1297	2,5	0,63	668	11,3	0,49	1137	2,9	0,68	3755	1,5	0,76

TABLEAU 4.11 – Indices basés sur les divisions hiérarchiques appliquées aux fréquences transformées en histogrammes.† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

	n0			n1			n2			n3		
	CH	*C*†*	Γ	*CH*	*C*†*	Γ	*CH*	*C*†*	Γ	*CH*	*C*†*	Γ
10	695	4,6	0,93	1089	3,8	0,90	2126	1,1	0,93	3210	1,5	0,88
9	746	4,1	0,96	1234	5,1	0,92	2435	1,7	0,97	3728	1,3	0,93
8	798	6,5	0,97	1372	3,1	0,98	1879	2,6	0,98	3782	1,5	0,96
7	910	6,2	0,96	1244	6,9	0,98	2951	1,2	0,98	4722	0,8	0,96
6	1021	7,4	0,97	1570	5,9	0,98	3657	1,5	0,99	5231	0,9	0,96
5	1508	3,6	0,97	2317	5,3	0,98	4325	0,9	0,99	6581	0,7	0,97
4	**1691**	**3,5**	**0,98**	**2342**	**2,7**	**0,99**	**5465**	**0,8**	**0,99**	**8326**	**0,5**	**0,98**
3	1285	4,5	0,95	2072	6,6	0,97	2754	2,1	0,99	4993	0,6	0,96
2	155	5,8	0,91	159	5,8	0,96	160	1,8	0,98	161	1,0	0,95

TABLEAU 4.12 – Indices basés sur les nuées dynamiques appliquées aux fréquences transformés en histogrammes.† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-3}$).

noté que, contrairement à ce qui a été observé pour les fréquences, l'indice C^* n'indique jamais le nombre correct de groupes. Pour les autres indices, la valeur optimale est toujours égale à quatre. Toutefois, l'indice Γ n'est pas capable d'indiquer correctement le nombre de clusters lorsque le niveau de bruit est élevé (n3).

	n0			n1			n2			n3		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
10	22,8	2,6	0,88	26,8	3,7	0,69	8,8	3,7	0,78	23,9	4,8	0,66
9	24,9	1,3	0,80	27,2	**3,3**	0,71	9,1	3,4	0,66	18,7	4,2	0,64
8	28,6	1,2	0,77	34,3	4,0	0,72	10,9	**3,3**	0,76	14,9	3,8	0,72
7	24,8	0,8	0,88	28,8	6,4	0,72	10,0	4,5	0,77	24,8	3,4	0,75
6	29,8	**0,2**	0,88	41,0	3,6	0,70	12,1	3,9	0,89	15,3	2,6	0,77
5	30,6	0,7	0,90	42,9	5,3	0,68	13,2	5,8	0,83	11,2	**1,2**	**0,79**
4	**33,8**	0,8	**0,91**	**44,0**	4,7	**0,77**	**13,4**	4,9	**0,92**	**34,9**	1,3	0,77
3	33,0	1,4	0,65	35,3	9,0	0,68	12,9	9,3	0,79	33,5	2,6	0,68
2	28,5	2,9	0,54	28,8	10,8	0,67	10,5	14,9	0,71	27,2	3,3	0,55

TABLEAU 4.13 – Indices basés sur les divisions hiérarchiques appliquées aux modes propres transformées en histogrammes.† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Par ailleurs, les indices sont appliqués aux partitionnements obtenus par la méthode des nuées dynamiques. Le Tableau 4.14 contient les valeurs calculées pour les indices en considérant

la transformation des modes propres en histogrammes. Dans ce cas, seuls les indices CH et Γ réussissent à indiquer quatre groupes comme étant le partitionnement optimal pour tous les niveaux de bruit, tandis que l'indice C^* ne le fait que pour le niveau de bruit n2.

	n0			n1			n2			n3		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
10	7,3	0,088	0,39	2,7	0,566	0,31	8,5	0,130	0,21	7,3	0,132	0,51
9	8,6	0,096	0,55	4,6	0,293	0,29	7,1	0,112	0,23	6,8	0,119	0,48
8	8,5	0,118	0,55	4,7	0,238	0,24	5,9	0,095	0,54	8,3	0,145	0,50
7	9,6	0.097	0,57	8,1	0,147	0,44	8,4	0,077	0,50	7,2	0,205	0,50
6	11,7	**0,071**	0,59	8,7	0,116	0,47	6,6	0,075	0,50	9,2	**0,077**	0,52
5	12,1	0,108	0,61	11,2	0,111	0,52	4,8	0,043	0,63	9,6	0,090	0,50
4	**12,3**	0,234	**0,64**	**11,4**	0,049	**0,72**	**12,4**	**0,040**	**0,84**	**11,5**	0,082	**0,64**
3	6,3	0,310	0,47	8,7	**0,040**	0,63	12,3	0,080	0,75	9,8	0,196	0,60
2	6,7	0,479	0,34	7,1	0,067	0,61	10,2	0,092	0,74	7,5	0,243	0,59

TABLEAU 4.14 – Indices basés sur les nuées dynamiques appliquées aux modes propres transformés en histogrammes.

4.3.3.4 Études complémentaires

Dans le paragraphe 4.3.2, des études complémentaires de clustering ont été menées en considérant différents ensembles d'essais. La première simulation tenait compte des essais correspondant aux états D0 et D3 de la structure. L'objectif ici est de vérifier si les indices de partitionnement optimal sont capables d'indiquer correctement le nombre de groupes existants (dans cet exemple, deux) lorsque tous les niveaux de bruit sont mélangés.

Scénarios D0 et D3

Dans cette analyse, le nombre maximal de clusters est défini arbitrairement comme étant égal à 5. Le Tableau 4.15 regroupe les valeurs calculées en utilisant les méthodes des divisions hiérarchiques et nuées dynamiques appliquées aux signaux. Pour la première méthode, les indices CH et C^* indiquent le nombre correct de clusters, tandis que Γ signale quatre groupes. Il est intéressant de remarquer que ce dernier indice présente des valeurs très proches dans toutes les simulations. Cela indique que, dans le cas des signaux transformés en histogrammes, l'indice Γ peut ne pas être le plus adéquat. En utilisant la méthode des nuées dynamiques, seul l'indice CH indique le nombre correct de groupes.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	38,1	0,148	0,66	47,9	**0,037**	0,76
4	43,2	0,13	**0,70**	45,4	0,062	0,78
3	50,7	0,121	0,69	44,8	0,111	**0,84**
2	**55,7**	**0,096**	0,68	**60,2**	0,164	0,82

TABLEAU 4.15 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en histogrammes (scénarios D0 et D3).

Le Tableau 4.16 détaille les indices calculés en utilisant les fréquences propres. Dans ce cas, pour les deux méthodes et les deux types de représentation symbolique, tous les indices indiquent le nombre correct de clusters. Il est intéressant de signaler qu'en utilisant des histogrammes,

l'indice C^* atteint la valeur zéro (le minimum absolu) et l'indice Γ, la valeur 1 (le maximum absolu). Cela confirme le taux de 100% de réussite des classifications obtenues précédemment.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	43,1	1,6	0,74	94,8	5,7	0,7
4	59,4	3,1	0,89	126,9	3,4	0,91
3	143,4	2,4	0,95	172,8	4,5	0,94
2	**223,9**	**0**	**1,00**	**339,8**	**0**	**1,00**

TABLEAU 4.16 – Indices basés sur les méthodes de clustering appliquées aux fréquences transformées en histogrammes (scénarios D0 et D3).† Les valeurs de l'indice C^* sont multipliées par $(\times 10^{-2})$.

Par ailleurs, la même analyse est menée en considérant les modes propres. Les résultats confirment la bonne adéquation des indices CH et Γ, étant donné qu'ils indiquent toujours correctement le nombre de groupes existants. L'indice C^*, toutefois, varie significativement (cf. Tableau 4.17).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	61,8	**4,7**	0,3	166,3	**1,7**	0,24
4	78,8	10,9	0,25	148,1	2,5	0,29
3	112,8	11,8	0,31	155,5	7,2	0,54
2	**187,4**	33,5	**1,00**	**195,9**	4,8	**0,63**

TABLEAU 4.17 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en histogrammes (scénarios D0 et D3).† Les valeurs de l'indice C^* sont multipliées par $(\times 10^{-2})$.

Scénarios D1 et D2

Comme dans l'analyse précédente, le nombre maximal de clusters est défini arbitrairement comme étant égal à 5. La deuxième étude complémentaire a utilisé l'ensemble des essais correspondant aux états structuraux D1 et D2. Il a été remarqué que les taux de classifications correcte ont été moins importants par rapport aux analyses avec les scénarios D0 et D3. Cette observation est également notée dans le calcul des indices de partitionnement optimal. Le Tableau 4.18 regroupe les valeurs calculées pour les deux méthodes de partitionnement en utilisant la transformation en histogrammes. Dans ce cas, il est noté que seuls les indices CH et Γ calculés à partir de la méthode des nuées dynamiques indiquent le nombre correct de clusters. Cela est en accord avec les taux de classification correcte plus importants obtenus par cette méthode par rapport à ceux obtenus par les divisions hiérarchiques.

L'analyse suivante considère l'utilisation des fréquences propres. Dans ce cas, les résultats sont significativement meilleurs que ceux obtenus en utilisant les signaux. En effet, pour toutes les méthodes, les indices CH et Γ signalent le nombre correct de clusters, comme montré dans le Tableau 4.19. Par ailleurs, l'indice C^* oscille entre 2 et 4 clusters, montrant que cet indice est parfois trop sensible aux petites variations structurales.

Par ailleurs, les modes propres sont considérés. Le Tableau 4.20 détaille les valeurs calculées pour chaque indice. Il est noté que l'indice CH n'arrive pas à indiquer correctement le nombre de groupes lorsque la transformation en histogrammes est utilisée. Étant donné que les scénarios

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	82	**4,3**	0,83	85,2	**3,5**	0,73
4	110,4	5,5	0,79	62,5	4,7	0,71
3	**112,3**	8,1	0,91	63	9,2	0,64
2	106,1	12,6	**0,92**	**100,8**	10,9	**0,81**

TABLEAU 4.18 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en histogrammes (scénarios D1 et D2).† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	49,3	8,3	0,45	51,5	43,7	0,38
4	38,1	**5,0**	0,75	52,1	31,4	0,67
3	50,9	15	0,81	63,9	**8,8**	0,95
2	**55,3**	62,4	**0,91**	**71,9**	11,1	**1,00**

TABLEAU 4.19 – Indices basés sur les méthodes de clustering appliquées aux fréquences transformées en histogrammes (scénarios D1 et D2).† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

d'endommagement D1 et D2 correspondent à des états structuraux relativement proches, il se peut que le calcul des indices à partir de cette représentation symbolique ne soit pas suffisamment robuste. Cependant, si des intervalles sont utilisés, cet indice (comme l'indice Γ) signalent 2 clusters comme étant le partitionnement optimal (cf. Tableau A.22).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	28,2	0,065	0,46	23,7	0,114	0,6
4	27,2	0,089	0,44	25,3	**0,078**	0,71
3	**35,3**	**0,061**	0,46	**31,4**	0,194	0,75
2	31,3	0,11	**0,38**	29,3	0,153	**0,78**

TABLEAU 4.20 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en histogrammes (scénarios D1 et D2).

Scénarios D0, D1, D2 et D3

La dernière partie de l'étude complémentaire a considéré l'ensemble complet des essais (160 au total). L'objectif est toujours de séparer les quatre scénarios d'endommagement, en considérant tous les niveaux de bruit confondus. La première analyse porte sur l'utilisation des signaux. Dans cette analyse, le nombre maximal de clusters est défini arbitrairement comme étant égal à 10. Le Tableau 4.21 regroupe les valeurs calculées pour les trois indices à partir des méthodes de partitionnement. Il est noté que les indices CH et Γ signalent quatre clusters lorsque la méthode des nuées dynamiques est utilisée. Pour la méthode des divisions hiérarchiques, toutefois, seul l'indice CH indique le nombre correct. Cela peut être expliqué par le fait que cette méthode a des taux de classification correcte moins importants que ceux obtenus par les nuées dynamiques (cf. Tableau 4.8).

L'analyse est également menée avec les fréquences propres. Étant donné que les classifications obtenues ont été significativement meilleures que celles utilisant les signaux, les indices sont

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
10	58,9	4,1	0,71	62,3	**3,1**	0,8
9	61	3,2	0,76	60,1	3,2	0,8
8	70,2	**2,7**	0,79	65,4	3,4	0,8
7	69,4	3,8	0,72	70,7	4	0,8
6	69,8	5	0,82	60,1	6,3	0,78
5	96,2	5,9	0,82	75,5	5,3	0,8
4	**117,9**	7,1	0,83	**84,7**	7,3	**0,83**
3	86,5	10,6	**0,84**	83,3	11,4	0,73
2	72	16,4	0,81	70,9	15,9	0,71

TABLEAU 4.21 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en histogrammes (scénarios D0, D1, D2 et D3).† Les valeurs de l'indice *C** sont multipliées par ($\times 10^{-2}$).

également plus cohérents. En effet, pour toutes les méthodes et en considérant n'importe quelle transformation symbolique, tous les indices sont capables de confirmer le partitionnement optimal avec quatre clusters (cf. Tableaux 4.22 et A.24). Il est également noté que l'indice Γ varie très peu, mais sa valeur maximale est toujours atteinte pour le nombre correct de groupes.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
10	332	4,8	0,82	413	5,3	0,9
9	318	4,8	0,91	464	4,6	0,93
8	356	4,5	0,95	509	5,9	0,95
7	400	4,6	0,97	614	5,1	0,96
6	474	4,2	0,97	703	4,7	0,98
5	592	2,6	0,98	876	3,4	0,99
4	**764**	**1,5**	**1,00**	**1123**	**0**	**1,00**
3	560	2,1	0,99	841	2,1	0,99
2	44	2,5	0,92	66	3,3	0,95

TABLEAU 4.22 – Indices basés sur les méthodes de clustering appliquées aux fréquences transformées en histogrammes (scénarios D0, D1, D2 et D3).† Les valeurs de l'indice *C** sont multipliées par ($\times 10^{-2}$).

La même démarche est appliquée aux modes propres, comme détaillé dans le Tableau 4.23. De façon générale, les indices *CH* et Γ indiquent dans la plupart des cas quatre clusters pour un partitionnement optimal. Cela n'est pas observé lorsque la méthode des divisions hiérarchiques est appliquée à des intervalles (cf. Tableau A.25) et lorsque les nuées dynamiques sont appliquées aux modes propres transformés en histogrammes. Dans ces cas, l'indice Γ indique cinq clusters pour un partitionnement optimal. L'indice *C** montre, comme dans les cas précédents, une tendance erratique en signalant des valeurs entre 9 et 10 clusters.

4.3.4 Bilan

Dans les paragraphes précédents, trois indices ont été présentés pour déterminer le partitionnement optimal d'un ensemble d'essais. Pour cela, plusieurs simulations ont été réalisées en considérant différents ensembles d'essais bruités en non bruités. Par ailleurs, les indices ont été calculés à partir de l'utilisation des méthodes de clustering - divisions hiérarchiques et nuées

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
10	19,4	2,2	0,47	5,2	3,7	0,37
9	16,2	**1,9**	0,45	9,7	**2,1**	0,38
8	15,9	2,6	0,42	6,4	13,8	0,41
7	19,3	3,3	0,48	6,6	19,7	0,44
6	22,5	3,6	0,54	7,7	19,3	0,43
5	25,1	3,6	0,52	9,3	19,5	**0,46**
4	**27,8**	9,5	**0,58**	**12,1**	20,3	0,44
3	25,3	14,4	0,55	5,1	27,6	0,31
2	18,9	68,2	0,55	5,5	68,2	0,28

TABLEAU 4.23 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en histogrammes (scénarios D0, D1, D2 et D3).

dynamiques - appliquées aux signaux et paramètres modaux. Ces derniers ont été transformés en intervalles interquartiles et histogrammes. Dans un premier temps, de nombreuses simulations ont été réalisées en tenant compte des essais ayant le même niveau de bruit (à partir du niveau n0 jusqu'au niveau n3). L'objectif est d'observer comment les indices changent à mesure que le taux de bruit augmente. De façon générale, il est remarqué que les indices CH et Γ sont plus robustes, notamment si les fréquences propres sont utilisées. Même pour des niveaux de bruit importants, ces indices sont capables d'indiquer correctement le nombre de clusters, ce qui n'est pas le cas de l'indice C^*. En effet, cet indice a varié de manière significative en signalant des valeurs aberrantes telles que 9 ou 10 clusters. Dans un second temps, les niveaux de bruit ont été tous mélangés, l'objectif étant de vérifier la capacité des indices à indiquer le nombre correct de clusters. Il a été observé une cohérence entre les indices lorsque les deux états D0 et D3 sont considérés (en particulier dans le cas où les paramètres modaux sont utilisés). Cependant, lorsque les états D1 et D2 sont considérés, les indices n'indiquent pas les deux groupes comme étant le partitionnement optimal. En effet, la discrimination de ces deux états a montré être plus délicate, étant donné que les niveaux d'endommagements imposés à la structure sont, dans un certain sens, plus similaires. Finalement, lorsque tous les essais sont considérés en même temps, les indices calculés confirment que CH et Γ indiquent correctement le partitionnement optimal.

4.4 ADS couplée aux méthodes de classification supervisée

Dans ce paragraphe, les résultats obtenus à partir du couplage entre l'ADS et les méthodes de classification supervisée sont présentés. Dans un premier temps, ces méthodes sont appliquées aux représentations symboliques des signaux. Dans un deuxième temps, la même procédure est appliquée aux fréquences propres. Finalement, les modes propres sont utilisés pour la discrimination des différents états structuraux.

Paramètres utilisés

Les simulations utilisant les ADB (Arbres de Décision Bayésiens) sont réalisées sous MatLab® à l'aide des routines développées par l'auteur. Pour la méthode basée sur les RN (réseaux de neurones), l'architecture du réseau utilisée est composée d'une couche cachée avec 20 neurones. Par ailleurs, les fonctions d'activation utilisées sont la fonction sigmoïde (cf. Équation (3.41)) dans la couche cachée et la fonction Heaviside (cf. Équation (3.40)) dans la couche de sortie. Le taux d'apprentissage η est fixé à 0,01 et la constante de moment μ est fixé à 0,0001. Ces paramètres ont été choisis après quelques simulations réalisées au préalable et ont montré une bonne adéquation aux problèmes traités (Cury & Crémona 2010*c*,*b*). Les simulations ont été réalisées à l'aide de la

boîte à outils Netlab (Bishop 1995). En ce qui concerne les SVM (Machines à Vecteurs Supports), la fonction noyau RBF (cf. paragraphe 3.2.3.2) est utilisée et le paramètre σ est déterminé de manière itérative lors de la phase de validation. Les simulations ont été réalisées à l'aide de la boîte à outils LibSVM (Chang & Lin 2001).

4.4.1 Analyse de la sensibilité au bruit

Les simulations réalisées dans la suite suivent le même déroulement que dans le cas des méthodes de clustering. L'objectif est d'analyser l'influence du bruit sur les classifications obtenues par les méthodes supervisées. Par ailleurs, étant donné que ces méthodes utilisent des groupes pour l'entraînement, validation et test des données, il est également intéressant d'observer comment les proportions de chacun de ses groupes peut modifier les résultats des classifications. Dans un premier temps, les taux de classification correcte sont calculés par niveau de bruit et en utilisant 30% des essais pour l'entraînement, 10% pour la validation et 60% pour le test. Cela signifie que pour chaque niveau de bruit, les 40 essais (10 essais D0, 10 essais D1, 10 essais D2 et 10 essais D3) sont répartis en 12 essais pour l'entraînement, 4 pour la validation [1] et 24 pour le test. Dans un deuxième temps, 40% d'essais (16) sont utilisés pour l'entraînement, 10% (4) pour la validation et 50% (20) pour le test. Finalement, la moitié des essais (20) font partie du groupe d'entraînement, 10% (4) de validation et 40% (16) de test.

Étant donné que ces distributions peuvent être déterminées de façon aléatoire et afin de mener une étude statistique, 10000 simulations différentes ont été réalisées (i.e., 10000 groupes différents d'entraînement, de validation et de test). Les résultats présentés dans la suite correspondent aux taux de classification maximal, moyen et minimal obtenus parmi les 10000 évalués. Dans le cas des méthodes de clustering, les taux de classification correcte ont été présentés par niveau d'endommagement. Ici, les taux correspondent à la proportion globale de classifications correcte obtenues. En d'autres termes, c'est le ratio entre le nombre total d'essais correctement classés (parmi les états D0, D1, D2 et D3) et le nombre total d'essais existants dans chaque analyse (40).

4.4.1.1 Signaux

Les résultats présentés dans le Tableau 4.24 correspondent aux trois méthodes de classification appliquées aux signaux, transformés en histogrammes. Les premiers résultats montrent que les méthodes des RN et des SVM sont plus robustes à la présence de bruit que les ADB. De façon générale, la méthode des SVM atteint des taux de classification correcte les plus élevés. Il est rappelé cependant que ce sont les taux moyens qui sont comparées. Dans ce cas, c'est la méthode des RN qui présente la meilleure performance (ayant cependant les taux de classification minimal assez faibles par rapport aux autres méthodes). Dans ce type de simulation, il est possible de dire que les RN et les SVM peuvent mieux définir les « seuils » de séparation entre les différents états structuraux.

Les proportions des groupes d'entraînement sont maintenant augmentées (cf. Tableaux 4.25 et 4.26). En général, il est noté que les résultats sont améliorés pour toutes les méthodes, surtout si les taux moyen et minimal sont comparés. En effet, cela montre que ces méthodes arrivent à mieux discriminer les scénarios d'endommagent à mesure que plus d'essais sont utilisés dans l'étape d'apprentissage. Parmi les trois méthodes, les ADB donnent des taux de classification correcte relativement moins importants. Cela peut être expliqué par le fait que les ADB ne sont pas une vraie méthode d'apprentissage. Comme décrit précédemment, les essais du groupe de

1. Pour la méthode des arbres de décision Bayésiens, il n'existe pas de groupes de validation. Dans ce cas, la proportion de ce groupe est ajoutée à celle du groupe d'entraînement.

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	83	79	75	75	100	100	96	92	100	100	100	100
Moyen	61	58	59	44	78	76	74	78	73	71	78	74
Minimal	38	25	33	11	21	25	48	55	17	48	21	38

TABLEAU 4.24 – Taux de classification correcte utilisant les signaux transformés en histogrammes (30% entr., 10% val., 60% test. Valeurs en %).

test sont classés selon des questions logiques. De cette façon, la détermination des seuils pour la séparation des classes peut être moins adéquate par rapport aux méthodes ayant une « vraie » apprentissage. Par ailleurs, il est intéressant de noter que l'augmentation de la proportion d'essais dans le groupe d'entraînement n'a pas montré une amélioration nette pour cette méthode. Si les taux maximaux sont considérés, les RN et SVM ont réussi (parmi les 10000 simulations réalisées) à séparer parfaitement les quatre scénarios d'endommagement. La tendance des pourcentages calculés montre également qu'à mesure que le niveau de bruit augmente, le nombre de classifications correctes diminue.

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	80	80	80	75	100	100	95	95	100	100	100	100
Moyen	66	55	56	48	78	76	78	74	85	81	80	84
Minimal	35	30	35	20	50	50	50	55	63	56	56	63

TABLEAU 4.25 – Taux de classification correcte utilisant les signaux transformés en histogrammes (40% entr., 10% val., 50% test. Valeurs en %).

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	83	83	83	75	100	100	100	100	100	100	100	100
Moyen	68	58	58	50	80	77	73	80	85	83	81	77
Minimal	38	42	33	19	50	44	50	56	66	58	50	58

TABLEAU 4.26 – Taux de classification correcte utilisant les signaux transformés en histogrammes (50% entr., 10% val., 40% test. Valeurs en %).

De plus, il est important de remarquer que l'application de ces méthodes aux données transformées en histogrammes est relativement plus coûteuse (en temps de calcul) que la transformation en intervalles interquartiles. En effet, une donnée symbolique est représentée par deux valeurs si des intervalles sont utilisés (la borne inférieure et supérieure), tandis que vingt valeurs ne sont nécessaires pour la représentation en histogrammes. Dans cette analyse, en utilisant les signaux, il est possible que la détermination des seuils de discrimination soit moins aisée dans certains cas. Les résultats obtenus à partir des intervalles interquartiles sont regroupés dans les Tableaux A.26, A.27 et A.28 (cf. Annexe A). En général, les taux de classification correcte sont légèrement plus faibles par rapport à ceux obtenus par des histogrammes.

4.4.1.2 Fréquences propres

Les Tableaux 4.27, 4.28 et 4.29 regroupent les taux de classification évalués pour les différentes simulations réalisées en utilisant les fréquences propres transformées en histogrammes. Dans ce cas, des résultats pertinents sont obtenus pour toutes les méthodes. Il est noté que même les ADB

arrivent toujours à classer les essais correctement. Bien que l'application des méthodes supervisées aux descriptions en histogrammes soit plus complexe, son utilisation avec les fréquences propres semble être la plus robuste et efficace dans ce type d'analyse. Par ailleurs, les effets de la présence de bruit sont faiblement remarqués dans le cas des ADB et des SVM, étant donné que les pourcentages ne varient pratiquement pas. Cela a été aussi remarqué pour les méthodes de clustering.

	ADB				**RN**				**SVM**			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	100	100	100	100	100	100	100	100	100	100	100	100
Moyen	100	100	100	99	97	97	90	90	96	97	98	92
Minimal	100	100	100	96	58	50	58	58	55	55	55	55

TABLEAU 4.27 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes (30% entr., 10% val., 60% test. Valeurs en %).

	ADB				**RN**				**SVM**			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	100	100	100	100	100	100	100	100	100	100	100	100
Moyen	100	100	100	100	99	95	95	94	99	99	98	86
Minimal	100	100	100	100	65	65	65	50	55	55	55	40

TABLEAU 4.28 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes (40% entr., 10% val., 50% test. Valeurs en %).

	ADB				**RN**				**SVM**			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	100	100	100	100	100	100	100	100	100	100	100	100
Moyen	100	100	100	100	99	99	95	94	99	99	97	95
Minimal	100	100	100	100	88	81	81	58	82	81	82	80

TABLEAU 4.29 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes (50% entr., 10% val., 40% test. Valeurs en %).

4.4.1.3 Modes propres

Les modes propres sont désormais utilisés. Les Tableaux 4.30, 4.31 et 4.32 regroupent les pourcentages évalués en considérant la transformation en histogrammes. Dans ce cas, comme il a été signalé pour les signaux, les taux de classification obtenus sont, en général, moins importants comparés à ceux évalués en utilisant des intervalles (cf. Tableaux A.32, A.33 et A.34). Cependant, il est intéressant de remarquer que les taux maximal et minimal sont relativement plus importants dans ce cas. En effet, il est possible que pour une (ou plus) simulation(s) donnée(s) (parmi les 10000 réalisées) le groupe d'entraînement soit plus « représentatif » en utilisant des histogrammes que des intervalles. Toutefois, en moyenne, cela n'est pas le cas. Par ailleurs, la présence de bruit a aussi affecté les classifications, notamment pour la méthode des ADB. En général, la méthode des RN atteint les pourcentages les plus élevés, en moyenne.

4.4.2 Études complémentaires

De façon similaire à celle présentée pour les méthodes de clustering, trois analyses complémentaires sont maintenant proposées en utilisant différents ensembles d'essais. Dans un premier

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	80	68	58	58	100	96	92	92	92	92	92	92
Moyen	57	33	33	35	73	71	55	57	75	72	67	62
Minimal	33	13	8	8	33	30	21	21	42	33	21	21

TABLEAU 4.30 – Taux de classification correcte utilisant les modes propres transformés en histogrammes (30% entr., 10% val., 60% test. Valeurs en %).

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	75	70	60	60	100	100	95	90	100	95	95	90
Moyen	56	33	35	33	82	77	57	61	64	75	71	59
Minimal	30	10	15	5	40	35	40	25	35	35	35	25

TABLEAU 4.31 – Taux de classification correcte utilisant les modes propres transformés en histogrammes (40% entr., 10% val., 50% test. Valeurs en %).

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	94	91	81	63	100	100	94	94	94	94	94	93
Moyen	57	35	33	33	87	81	58	63	65	75	69	56
Minimal	25	20	10	8	31	44	13	19	19	38	31	13

TABLEAU 4.32 – Taux de classification correcte utilisant les modes propres transformés en histogrammes (50% entr., 10% val., 40% test. Valeurs en %).

temps, les 80 essais correspondant aux états D0 (40 essais) et D3 (40 essais) sont considérés. Cet ensemble d'essais est divisé en groupes d'entraînement, validation et test. L'objectif est aussi de mener une étude statistique en simulant diverses proportions pour chaque groupe. Ainsi, les premières simulations contiennent 30% (24) d'essais pour l'entraînement, 10% (8) pour la validation et 60% (48) pour le test. Des simulations sont également menées en considérant les proportions 40% (32), 10% (8) et 50% (40) ; et finalement, 50% (40), 10% (8) et 40% (32). Ici, au lieu d'avoir quatre classes différentes et des niveaux de bruit séparés comme auparavant, il existe uniquement deux classes, mais tous les niveaux de bruit sont mélangés. L'idée est d'observer si les méthodes sont capables de distinguer ces deux états structuraux pour n'importe quel niveau de bruit. Les paramètres utilisés sont essentiellement les mêmes, à l'exception de la fonction d'activation utilisée dans la couche cachée des RN. Ici, au lieu de la fonction sigmoïde, la fonction tangente hyperbolique est employée (cf. Équation (3.42)). Ce choix est lié aux résultats obtenus à partir des simulations réalisées préalablement.

4.4.2.1 Scénarios D0 et D3

Les premières analyses correspondent à l'application des méthodes de classification aux signaux, transformés en histogrammes (cf. Tableau 4.33). Il est noté que les SVM ont toujours les meilleurs résultats. Par ailleurs, à mesure que le nombre d'essais dans le groupe d'entraînement augmente, les taux de classification améliorent pour toutes les méthodes. Il est également remarqué que, pour cette analyse, les ADB permettent d'obtenir des pourcentages plus importants que ceux des RN. Toutefois, leurs taux de classification correcte sont relativement faibles (de l'ordre de 50%, en moyenne).

L'analyse suivante consiste à appliquer les méthodes supervisées aux fréquences propres. Le

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	71	71	100	73	75	100	75	78	100
Moyen	54	52	76	55	52	78	56	55	80
Minimal	33	28	40	35	30	37	28	31	42

TABLEAU 4.33 – Taux de classification correcte utilisant les signaux transformés en histogrammes pour les scénarios D0 et D3 (valeurs en %).

Tableau 4.34 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en histogrammes est considérée. Comme il a été observé pour les méthodes de clustering, les méthodes de classification supervisée permettent d'obtenir des résultats pertinents. Bien que les ADB n'arrivent pas à classer 100% des essais, les RN et les SVM ne le font pour toutes les simulations. Cela met en évidence encore une fois que les fréquences propres, transformées en données symboliques, peuvent constituer un indicateur efficace pour la détection d'endommagements.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	100	100	100	100	100	100	100	100	100
Moyen	99	100	100	100	100	100	100	100	100
Minimal	98	100	100	100	100	100	100	100	100

TABLEAU 4.34 – Taux de classification correcte utilisant les fréquences transformées en histogrammes pour les scénarios D0 et D3 (valeurs en %).

Par ailleurs, cette démarche est appliquée aux modes propres. Des résultats adéquats sont également obtenus. Le Tableau 4.35 regroupe les taux de classification évalués en utilisant la transformation en histogrammes. Dans ce cas, les ADB ont, en moyenne, des taux moins importants que ceux obtenus par des intervalles (cf. Tableau A.37). Cependant, les meilleurs taux sont observés si des histogrammes sont pris en compte. Les RN et SVM produisent des résultats sensiblement meilleurs, surtout dans le dernier cas (50% E., 10% V., 40% T.) où tous les essais sont classés correctement.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	92	100	100	93	100	100	97	100	100
Moyen	75	96	97	75	96	99	76	97	100
Minimal	52	53	95	50	55	97	50	73	100

TABLEAU 4.35 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les scénarios D0 et D3 (valeurs en %).

4.4.2.2 Scénarios D1 et D2

La deuxième étude complémentaire tient compte de l'ensemble des essais correspondant aux états structuraux D1 et D2. Il a été observé que les taux de classification correcte ont été moins importants par rapport aux analyses avec les scénarios D0 et D3. En effet, cela est aussi remarqué pour les méthodes de classification supervisée. En général, les taux de classification sont de l'ordre de 50%, en moyenne, pour toutes les méthodes, en considérant la transformation des signaux en intervalles (cf. Tableau A.38). Les SVM sont la seule méthode capable d'élever ce taux

à 60% environ, lorsque des histogrammes sont considérés (cf. Tableau 4.36). Bien que ces méthodes aient montré être efficaces pour la détection de modifications structurales, il reste toujours difficile la discrimination des scénarios d'endommagement relativement similaires, notamment si des signaux sont utilisés.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	65	69	72	65	71	76	68	71	72
Moyen	57	50	57	60	54	61	59	55	62
Minimal	48	29	35	53	43	32	50	51	37

TABLEAU 4.36 – Taux de classification correcte utilisant les signaux transformés en histogrammes pour les scénarios D1 et D2 (valeurs en %).

Les fréquences propres sont maintenant considérées. Bien que les taux soient moins importants que ceux obtenus précédemment dans l'étude des états D0 et D3, les pourcentages ici évalués restent pertinents. Il est noté que les SVM arrivent à classer 100% d'essais pour les deux dernières simulations, en considérant des histogrammes (cf. Tableau 4.37). Les RN obtiennent aussi une classification parfaite lors des dernières simulations. Les taux obtenus par des ADB sont également adéquats, en variant entre 80% à 88%.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	88	100	100	92	100	100	94	100	100
Moyen	85	97	99	83	99	100	88	100	100
Minimal	67	95	98	65	97	100	61	100	100

TABLEAU 4.37 – Taux de classification correcte utilisant les fréquences transformées en histogrammes pour les scénarios D1 et D2 (valeurs en %).

La démarche est appliquée aux modes propres. Le Tableau A.40 détaille les pourcentages obtenus en considérant des intervalles. Dans ce cas, les ADB ont des niveaux relativement faibles, contrairement aux RN et aux SVM. De plus, la transformation en histogrammes ne change pas significativement les taux de classification obtenus par les ADB. Cependant, il est noté que cette représentation a amélioré les pourcentages des RN et des SVM (cf. Tableau 4.38).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	65	90	93	65	91	93	69	94	93
Moyen	52	82	88	56	85	86	57	88	88
Minimal	44	51	62	43	52	61	39	51	64

TABLEAU 4.38 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les scénarios D1 et D2 (valeurs en %).

4.4.2.3 Scénarios D0, D1, D2 et D3

La troisième partie de l'étude complémentaire considère l'ensemble total de 160 essais. L'objectif est toujours de séparer les quatre scénarios d'endommagement, en considérant tous les niveaux de bruit ensemble. En cela, les premières simulations contiennent 30% (48) d'essais pour l'entraînement, 10% (16) pour la validation et 60% (96) pour le test. Des simulations sont également menées en considérant les proportions 40% (64), 10% (16) et 50% (80) ; et finalement, 50%

(80), 10% (16) et 40% (64). La première analyse prend en compte les signaux. Il est noté que la transformation en intervalles ne produit pas des bons résultats. En effet, il est montré dans le Tableau A.41 qu'en moyenne, moins de la moitié des essais est classé correctement. Cela est relativement amélioré si la représentation par des histogrammes est considérée (cf. Tableau 4.39). Dans ce cas, seules les ADB et les SVM obtiennent des pourcentages supérieurs à 65%. Les RN atteint des taux très faibles.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	80	60	100	82	65	100	84	65	100
Moyen	75	25	65	76	33	68	76	35	71
Minimal	12	13	25	25	15	20	15	16	22

TABLEAU 4.39 – Taux de classification correcte utilisant les signaux transformés en histogrammes pour tous les scénarios d'endommagement (valeurs en %).

En utilisant les fréquences propres, les résultats sont visiblement meilleurs. Le Tableau 4.40 regroupe les taux calculés en considérant la représentation par des histogrammes. Contrairement à la tendance observée, les classifications sont relativement moins pertinentes dans les cas des RN et des SVM par rapport à celles obtenues en utilisant des intervalles (cf. Tableau A.42). Cependant, les ADB mènent à des pourcentages relativement plus importants dans ce cas.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	98	100	98	98	100	98	98	100	100
Moyen	89	98	87	91	99	87	91	99	88
Minimal	71	88	71	75	85	69	78	94	73

TABLEAU 4.40 – Taux de classification correcte utilisant les fréquences transformées en histogrammes pour tous les scénarios d'endommagement (valeurs en %).

Finalement, les modes propres sont utilisés. Il est noté que les résultats sont meilleurs que ceux obtenus par les signaux, mais les taux sont relativement plus faibles comparés à ceux obtenus en considérant les fréquences propres. À l'exception des SVM, les autres méthodes mènent à des meilleurs taux si des intervalles sont utilisés (cf. Tableau A.43). En utilisant des histogrammes, les pourcentages évalués avec les SVM sont adéquats quel que soit le nombre d'essais utilisés dans le groupe d'entraînement. En moyenne, 97% des essais sont correctement classés dans leurs groupes (cf. Tableaux A.43).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	80	92	100	80	94	100	82	95	100
Moyen	76	78	97	77	81	97	79	83	97
Minimal	35	22	82	47	39	84	49	25	90

TABLEAU 4.41 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour tous les scénarios d'endommagement (valeurs en %).

4.4.3 Bilan

Dans les paragraphes précédents, trois méthodes de classification supervisée ont été utilisées pour la discrimination d'états structuraux. Pour cela, plusieurs simulations ont été réalisées en

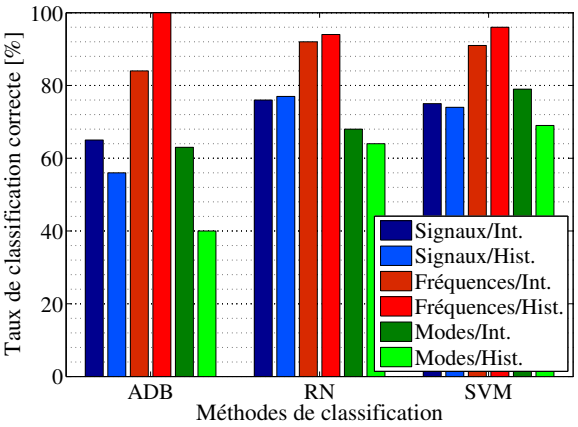
considérant différents scénarios d'endommagent et niveaux de bruit. L'objectif a été de vérifier la robustesse de ces méthodes à mesure que niveau de bruit devient plus important. Afin de synthétiser les résultats obtenus par ces méthodes, les taux moyens de classification correcte sont détaillés dans la suite. La Figure A.21 (cf. Annexe A) regroupe les pourcentages obtenus pour chaque méthode en utilisant les signaux. Pour chacune de ces figures, les taux de classification **moyen** sont regroupés par niveau de bruit. De plus, chaque barre verticale correspond à une configuration différente des essais, étant donné que le nombre d'essais dans les groupes d'entraînement et test ont changé pour chaque simulation. La première paire de graphiques (Figures A.21(a) et A.21(b)) correspond à la méthode des ADB appliquée aux signaux transformés en intervalles interquartiles et histogrammes, respectivement. De manière générale, il est noté que la transformation en intervalles interquartiles, pour cette méthode, produit des résultats légèrement meilleurs, compte tenu de tous les niveaux de bruit. Ceci est également observé pour la méthode des RN (Figures A.21(c) et A.21(d)). Finalement, les taux calculés utilisant la méthode des SVM sont illustrés aux Figures A.21(e) et A.21(f). Dans cette analyse, cette méthode a produit les meilleurs taux de classification correcte parmi les trois présentées. Contrairement à ce qui a été observé pour les ADB et les RN, les SVM appliquées aux signaux transformés en histogrammes ont montré être relativement plus adéquates pour la discrimination des différents niveaux d'endommagements. En général, les taux de classification correcte augmentent avec le nombre d'essais utilisés dans le groupe d'entraînement. Par ailleurs, il est noté que les RN et les SVM sont les méthodes les plus robustes au bruit, étant donné que la variation des taux de classification n'a pas été très importante.

Comme il a été observé avec les méthodes de clustering, les résultats obtenus en utilisant les fréquences propres ont montré que leur application est la plus robuste comparée aux signaux ou aux modes propres (Figure A.22). En effet, les taux de classification correcte les plus importants (de l'ordre de 90%) ont été trouvés par toutes les méthodes, à l'exception de celle basée sur les ADB appliquée à des intervalles interquartiles. Il est noté que dans ce cas, les ADB appliqués à des histogrammes ont permis de classer parfaitement tous les essais pour tous les niveaux de bruit considérés. De façon générale, toutes les méthodes permettent d'atteindre de meilleurs résultats lorsque la transformation en histogrammes est considérée. Les RN et les SVM produisent souvent des classifications similaires.

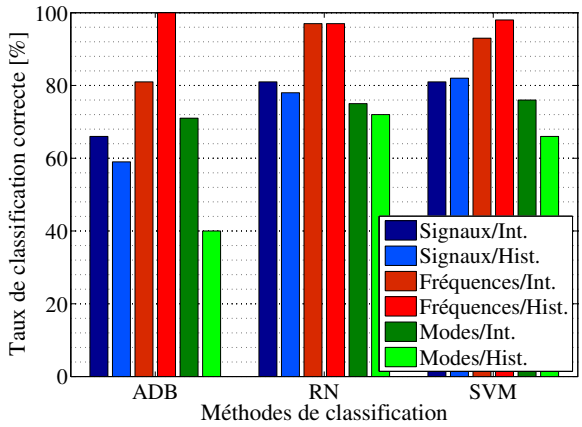
Les Figures A.23(a) et A.23(b) illustrent les pourcentages obtenus en utilisant la méthode des ADB appliquée aux modes propres transformés en intervalles et histogrammes, respectivement. Contrairement à ce qui a été remarqué dans l'analyse avec les fréquences propres, la représentation en intervalles interquartiles produit des résultats sensiblement meilleurs, surtout pour les ADB. Les Figures A.23(c-f) montrent qu'en général, les RN et les SVM sont les méthodes les plus adéquates pour la discrimination des états structuraux. Les taux de classification correcte sont relativement supérieurs à ceux obtenus en utilisant les signaux, mais inférieurs à ceux obtenus par les fréquences propres.

Finalement, la Figure 4.23(a) synthétise la moyenne globale (tous les niveaux d'endommagement et bruit confondus) des taux de classification pour chaque méthode et chaque type de transformation utilisée, en considérant 30% des essais pour l'entraînement et 60% pour le test. Dans cette analyse, lest SVM ont montré être la méthode la plus adéquate pour la discrimination des états structuraux. De plus, l'application de toutes les méthodes aux fréquences propres a montré la robustesse et l'efficacité de l'approche proposée (notamment dans le cas des ADB). Les méthodes de classification appliquées aux modes propres ont aussi mené des résultats relativement meilleurs que ceux obtenus par les signaux. De façon générale, les ADB ont montré être les moins performants. Bien qu'ils aient mené à des bonnes classifications en utilisant les fréquences propres, cela n'a pas été reproduit avec les signaux et les modes propres. Cela pourrait être expliqué par la difficulté de cette méthode à déterminer les seuils de discrimination entre les

scénarios d'endommagement. La Figure 4.23(b) illustre les résultats obtenus par ces méthodes, lorsque le nombre d'essais dans le groupe d'entraînement est plus important. En général, les taux de classification sont plus élevés pour toutes les méthodes, surtout si les signaux sont considérés.



(a) Synthèse des résultats en considérant 30%E., 10%V., 60%T.



(b) Synthèse des résultats en considérant 50%E., 10%V., 40%T.

FIGURE 4.23 – Synthèse des résultats obtenus pour les méthodes des ADB, RN et SVM.

4.5 Conclusions

Dans ce chapitre, une application numérique basée sur un modèle de poutre bi appuyée, discretisée aux éléments finis, a été considérée. Cette application a consisté à simuler quatre scénarios d'endommagement différents, chacun représentant un état structural de la poutre. Par ailleurs, trois niveaux de bruit ont été ajoutés aux mesures dynamiques. L'objectif a été de réaliser une étude de sensibilité sur la capacité des méthodes de classification à détecter des modifications structurales en présence de bruit.

Dans un premier temps, une approche basée sur le couplage de l'ADS avec les méthodes de clustering - divisions hiérarchiques, agglomérations hiérarchiques et nuées dynamiques - a été mise en œuvre. Pour cela, diverses simulations tenant compte de différents sous ensembles d'essais ont été réalisées. Le but initial a été d'observer comment ces méthodes peuvent discriminer les scénarios d'endommagement à mesure que le niveau de bruit augmente. Des analyses ont été menées en utilisant les signaux et paramètres modaux transformés soit en intervalles interquartiles, soit en histogrammes. Les résultats ont montré que, de manière générale, la démarche proposée est capable de détecter des défauts structuraux. De plus, il a été noté que les méthodes des nuées dynamiques et des divisions hiérarchiques ont été plus efficaces pour la discrimination des différents scénarios, notamment si des fréquences propres sont utilisées. Bien que la méthode des agglomérations hiérarchiques n'ait pas permis d'obtenir des résultats entièrement convaincants, elle permet d'observer des relations entre les essais à partir de la construction des graphiques de type pyramidal. Il a été également observé que cette méthode a une tendance à mélanger les essais, surtout pour des niveaux de bruit importants. Par ailleurs, trois études complémentaires ont été menées. Dans la première, seuls les scénarios D0 et D3 (correspondant aux états sain et le plus endommagé de la poutre) sont considérés. Toutes les méthodes sont alors capables de discriminer ces deux états avec des taux de classification correcte adéquats. Lorsque les scénarios D1 et D2 sont considérés, les pourcentages de classification correcte sont significativement plus faibles. Ces états correspondent à des niveaux d'endommagement relativement similaires, la démarche proposée peut ne pas être la plus efficace pour les discriminer. Enfin, lorsque tous

les essais sont utilisés (tous les scénarios d'endommagement et tous les niveaux de bruit pris en même temps), les taux de classification sont relativement meilleurs. En effet, si les fréquences propres sont utilisées, 100% de classification correcte est atteinte.

La deuxième partie de cette étude a consisté à appliquer les indices de partitionnement optimal aux méthodes de clustering. Ces indices sont importants dans les cas où le nombre de groupes de comportement n'est pas connu au préalable. Pour chaque simulation réalisée, les indices ont été calculés. De manière générale, l'utilisation des indices CH et Γ a donné des résultats satisfaisants, en indiquant toujours le nombre correct de clusters, notamment si les fréquences propres sont utilisées. Toutefois, à mesure que le niveau de bruit augmente, des indications incorrectes ont été observées (surtout pour celles où des signaux sont utilisés). L'indice C^* a montré être le plus sensible aux variations structurales, mais aussi aux différents niveaux de bruit. En effet, cet indice a rarement indiqué le nombre correct de groupes, oscillant fortement et présentant des valeurs parfois aberrantes. Ceci montre que, pour l'étude considérée, cet indice n'est pas le plus adéquat pour déterminer le partitionnement optimal de clusters.

Finalement, une démarche basée sur le couplage de l'ADS aux méthodes de classification supervisée - arbres de décision Bayésiens, réseaux de neurones et machines à vecteur support - a été considérée. Pour cela, les mêmes simulations menées en utilisant les méthodes de clustering ont été réalisées. Bien que la méthode des ADB n'ait pas mené à des taux de classification correcte importants, il a été observé que, de façon générale, les RN et les SVM ont produit des résultats cohérents. En effet, en les comparant aux méthodes de clustering, les méthodes de classification supervisée ont atteint des taux de classification correcte plus élevés. Cela est dû, manifestement, au fait que les méthodes de classification supervisées sont basées sur une procédure d'apprentissage. En d'autres termes, avant de classer des essais inconnus, elles sont entraînées avec des essais « étiquetés » de sorte qu'elles apprennent les seuils de séparation entre les différents groupes (ou classes). L'utilisation de ces méthodes a montré être robuste, surtout en considérant les fréquences propres. Dans la pratique, il n'est généralement pas aisé de connaître *a priori* les différents états d'une structure. Toutefois, si cela est possible, l'utilisation de ces méthodes est fortement conseillée pour la détection d'endommagements structuraux.

Chapitre 5

Étude du pont-rail PK 075+317

Après avoir appliqué les méthodes de classification non supervisée et supervisée à un modèle numérique de poutre, une première étude expérimentale est présentée dans ce chapitre. Dans un premier temps, la procédure d'identification modale, appliquée au pont-rail PK 075+317 reliant Paris à Lyon est introduite. Dans un deuxième temps, des analyses classiques sur les paramètres modaux identifiés sont réalisées. Dans un troisième temps, l'ADS est couplée aux méthodes de classification pour la discrimination des états structuraux. Par ailleurs, une analyse des effets de la température sur les paramètres modaux est menée. Pour cela, les techniques de régression linéaire et non linéaire présentées dans le chapitre 3 sont utilisées.

5.1 Présentation du cas d'étude

Dans le cadre du projet Réseau Génie Civil & Urbain « Évaluation Dynamique des Ponts » (Crémona et al. 2003, Crémona 2004), le suivi dynamique d'un pont-rail avait été retenu avec un double objectif : d'une part caractériser et quantifier *in situ* les améliorations apportées par le resserrage de bielles sur le comportement de l'ouvrage, et d'autre part évaluer l'impact des variations environnementales sur les caractéristiques modales à partir de mesures ponctuelles à des périodes différentes de l'année. Le pont-rail PK 075+317 fait partie de la ligne à grande vitesse Paris Sud-Est. Cet ouvrage, franchissant la route départementale 939 entre les communes de Sens et de Soucy dans l'Yonne (Figures 5.1 et 5.2), est apparu un candidat intéressant pour répondre à ces deux objectifs.

FIGURE 5.1 – Localisation du pont-rail PK 075+317.

FIGURE 5.2 – Vue de l'ensemble du pont-rail.

L'ouvrage, d'une portée de 17,5 m, est constitué de deux ponts en poutrelles enrobées. L'instrumentation et les mesures décrites dans cette section concernent le tablier de la voie 1 où les trains circulent dans le sens Paris → Lyon et qui, contrairement au tablier de la voie 2, n'est pas équipé d'un tapis anti-vibratile.

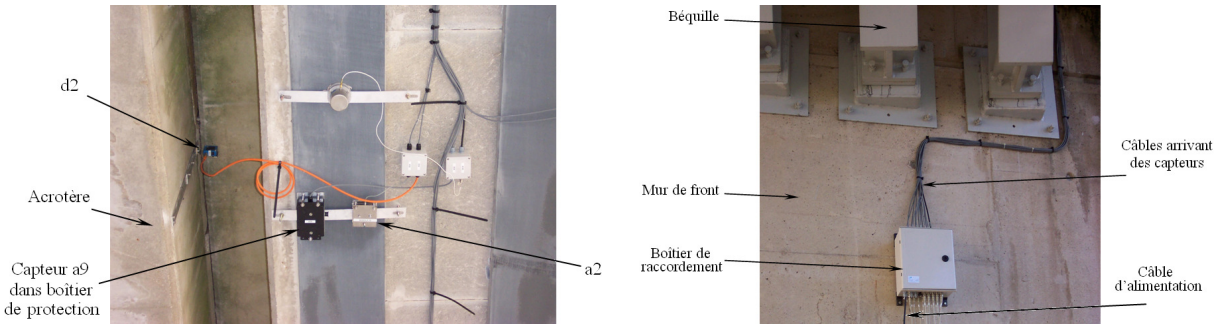
L'instrumentation mise en place et les conditions d'intervention par la société SITES et la SNCF sont détaillées par Crémona (2004). L'instrumentation a été mise en place du 24 au 26 juin 2003 et comporte les capteurs suivants :

- 3 capteurs de déplacement vertical entre l'acrotère et le tablier : chaque capteur de déplacement est monté sur un support fixé sur l'acrotère et mesure le déplacement relatif entre l'acrotère supposé fixe et le tablier.
- 8 accéléromètres verticaux et 2 accéléromètres horizontaux (longitudinal et latéral) sous le tablier : chaque accéléromètre est monté sur un support fixé sous le tablier au niveau des poutrelles.
- 2 sondes de température respectivement en-dessous et au dessus de l'acrotère.
- 2 ponts Q sur la voie : chaque pont Q mesure les charges sur essieux à l'entrée et à la sortie de l'ouvrage.

Les supports sont à la fois collés et maintenus mécaniquement par différents montages qui permettent le réglage de l'orientation de chaque capteur. Les capteurs de déplacement et les accéléromètres sont reliés, par l'intermédiaire de câbles cheminant sous le tablier, à un boîtier de raccordement, fixé à environ 3 m du sol sur un mur de front, depuis lequel est installée la chaîne d'acquisition (Figure 5.3). Les câbles des ponts Q, préalablement installés par la SNCF, arrivent à proximité du boîtier de raccordement mais ne sont pas reliés à ce dernier. L'acquisition de la température est réalisée indépendamment des autres mesures.

La Figure 5.4 présente le schéma de l'implantation des capteurs sous le tablier et la numérotation adoptée pour la détermination des déformées modales expérimentales. La source d'excitation du tablier de la voie 1 est le passage des TGV commerciaux sur cette voie. Les enregistrements se font en continu pendant une durée d'environ 3,5 s.

Les ponts Q ont cependant permis de discriminer la fréquence d'excitation liée au passage du convoi TGV. En effet, le décalage de temps entre les premiers pics de la réponse des ponts Q1 et Q2 permet d'obtenir la vitesse de passage du TGV qui reste à peu près constante (ce qui justifie ainsi la répétabilité des essais). Par exemple, les premiers pics sont décalés en moyenne de 0,223 s, ce qui donne approximativement la vitesse de passage de $17,5/0,223 = 78,47$ m·s^{-1}. Connaissant la distance entre les bogies (18,70 m), la fréquence d'excitation du convoi est évaluée



(a) Vue des capteurs d2, a2 et a9 (b) Vue du boîtier de raccordement et des câblages

FIGURE 5.3 – Vue des capteurs positionnés sur l'ouvrage.

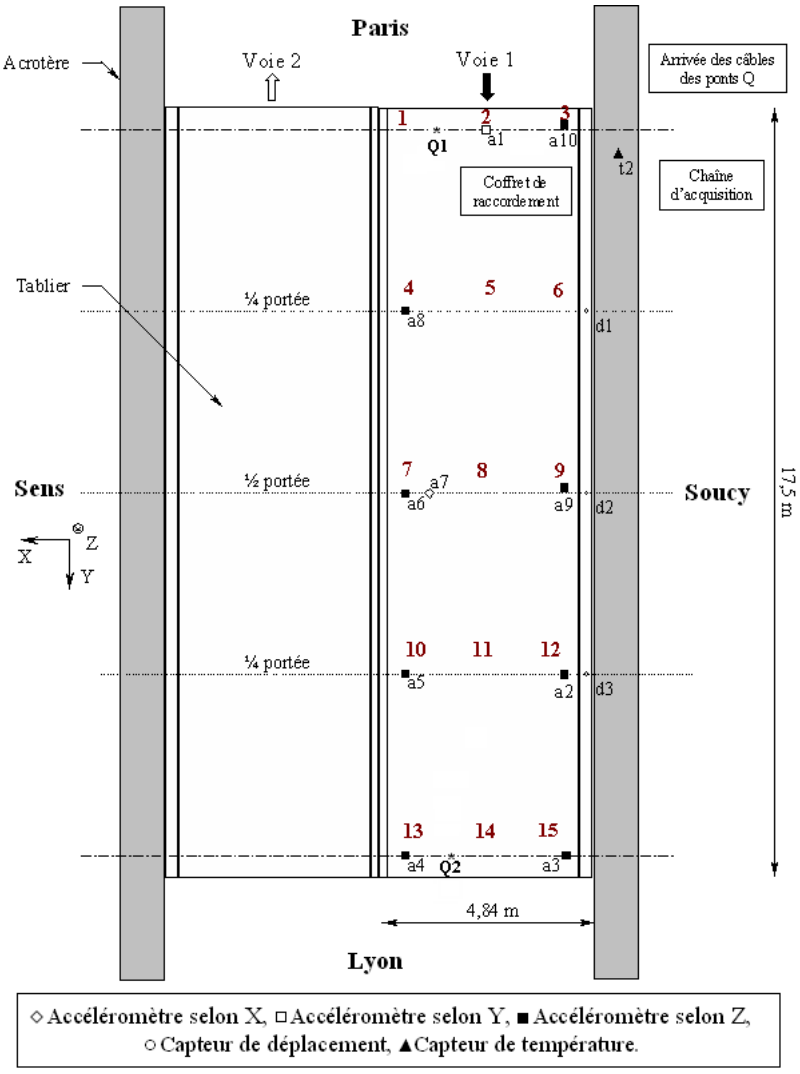


FIGURE 5.4 – Implantation des capteurs sous le tablier.

à 4,20 Hz.

Les précédentes mesures réalisées par la SNCF sur cet ouvrage ont mis en évidence des vibrations et des déformations excessives du tablier lors du passage des TGV. Ces constatations ont conduit le Département des Ouvrages d'Art de la SNCF à modifier les conditions d'appui

du tablier, précédemment en appui simple, par la réalisation d'un encastrement mécanique des abouts du tablier sur les culées. En effet, une procédure de rigidification de l'ouvrage a été mise en œuvre afin d'éloigner la fréquence d'excitation liée au passage du convoi TGV de la première fréquence propre du pont. Le risque de résonance était d'autant plus fort que le rechargement en ballast apporte une masse additionnelle qui tend à faire diminuer les fréquences propres du pont. La SCNF a donc mis en place un système de bielle aux extrémités de l'ouvrage (Figure 5.5) qui sont resserrées par clé dynamométrique. Ce resserrage apporte une rigidification qui a pour but d'augmenter les fréquences propres et d'éviter la mise en résonance (Crémona 2004).

(a) Vue inférieure des bielles

(b) Détaille des bielles

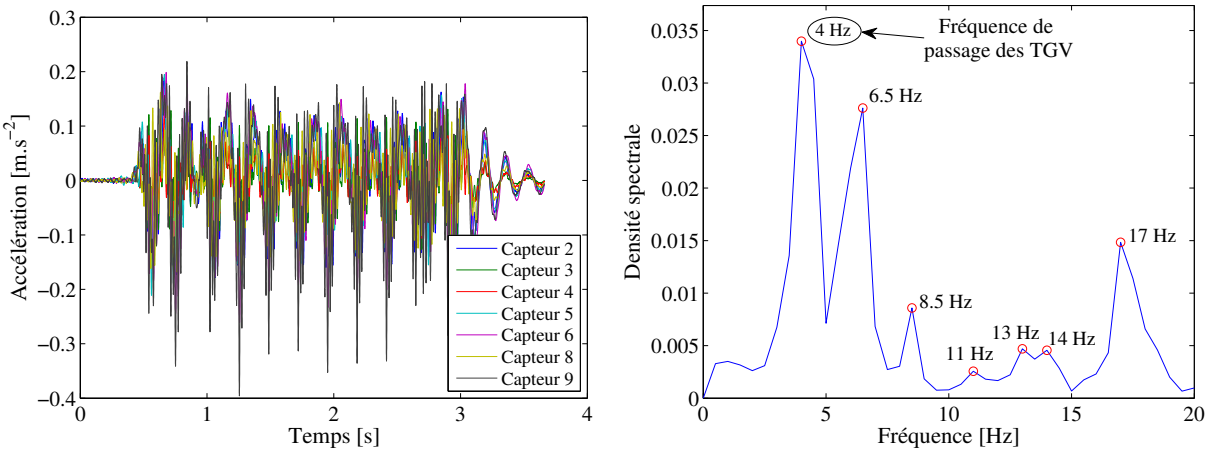
(c) Procédure de resserrage

FIGURE 5.5 – Vue des tabliers et des bielles du pont-rail.

5.2 Identification modale

Préalablement à l'analyse temporelle des signaux, une brève étude des densités spectrales de puissance a été réalisée afin d'estimer grossièrement les fréquences excitées pendant les essais. La Figure 5.6 présente les allures temporelles des signaux et la densité spectrale pour le capteur 5 (voir Figure refplansens).

Les mesures accélérométriques verticales permettent l'identification des principales caractéristiques modales de la structure. En observant la densité spectrale évaluée, deux pics très proches sont mis en évidence. Le premier pic, à environ 4 Hz, correspond à la fréquence de passage du TGV. Les autres pics correspondent aux quatre premières fréquences propres identifiées lors du passage du TGV. De manière générale, pour l'ensemble d'essais, les fréquences se trouvent



(a) Réponse temporelle des capteurs verticaux

(b) Densité spectrale du capteur 5

FIGURE 5.6 – Fréquences propres identifiées à partir de la densité spectrale du capteur 5.

comprises entre les bandes fréquentielles suivantes (représentées par chaque colonne)

$$\left[\begin{array}{c|c|c|c} 5 & 8 & 12 & 16 \\ 6 & 10 & 14 & 20 \end{array}\right] \text{Hz}$$

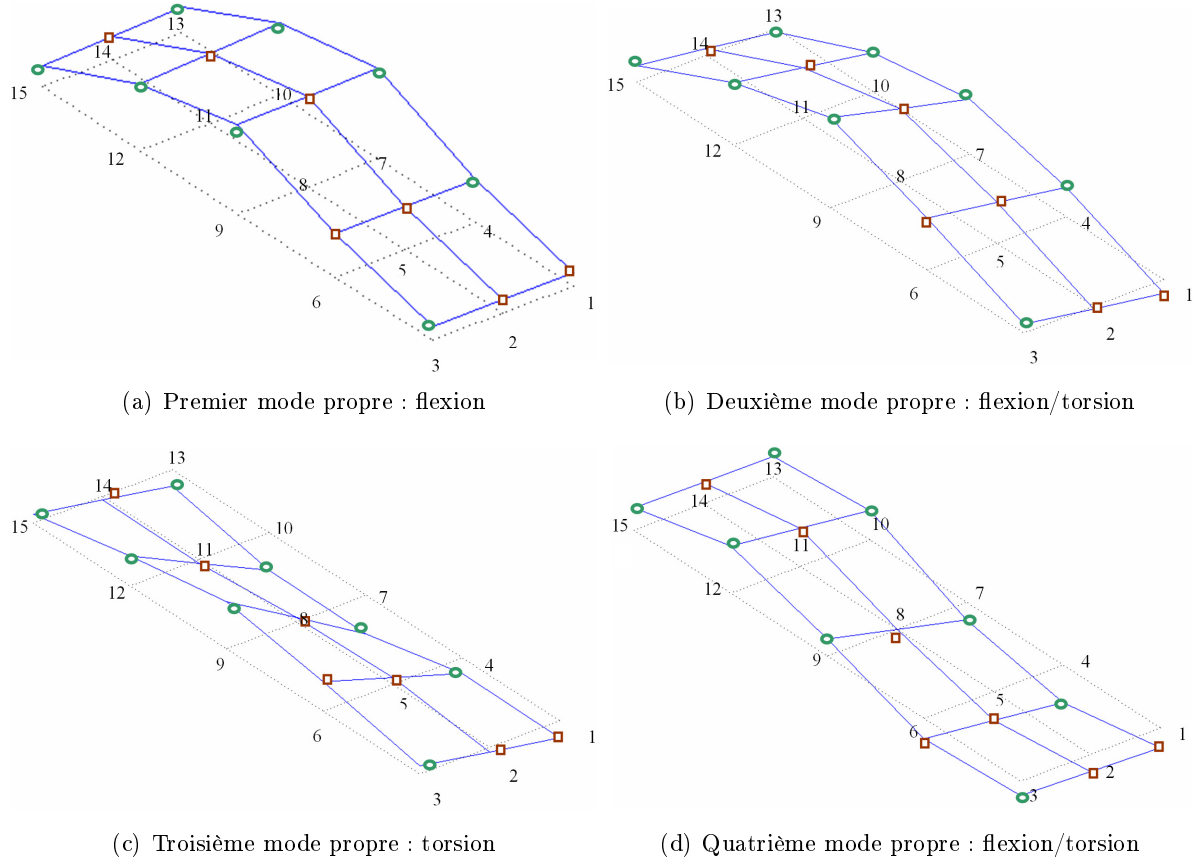
La boîte à outils Sys-Ident (SysIdent-LCPC 2007) a été utilisée pour l'identification modale du pont. Après le filtrage autour des plages de fréquences identifiées avec les densités spectrales, la méthode du décrément aléatoire est appliquée. La fréquence d'échantillonnage des signaux est de 1/4096 s. Les points de référence pour le calcul des décréments et la normalisation des modes propres sont respectivement le point 9 (voie de mesure 2) pour les 3 premiers modes et le point 4 (voie de mesure 8) pour le 4$^{\text{ème}}$ mode. Ces points sont choisis car situés au niveau des ventres de vibration pour les modes concernés. Afin d'avoir une bonne représentation des paramètres modaux identifiés, 10 réalisations sont considérées pour chaque essai. Dans cette analyse, chaque pas du décrément est considéré comme une réalisation (cf. Annexe D).

Les quatre premiers modes propres identifiés sont présentés à la Figure 5.7 (Crémona et al. 2004). Les points non mesurés (représentés par les carrés) du maillage ont été interpolés linéairement.

Comme mentionné précédemment, l'un des intérêts des campagnes de mesures accélérométriques sur ce pont, est de caractériser et quantifier les améliorations apportées par le resserrage de bielles sur le comportement dynamique de l'ouvrage. Pour cela, trois campagnes de mesures ont été réalisées. La première, réalisée le 24 juin 2003 a consisté à évaluer les caractéristiques modales du pont avant les travaux de resserrage. Pour cela, 15 essais dynamiques (appelés TGV1A à TGV15A) ont été enregistrés. Le Tableau B.1 (cf. Annexe B) regroupe les heures d'enregistrement des essais, les valeurs moyennes (calculées à partir des 10 réalisations) et les écarts-types respectifs des fréquences propres identifiées pour chaque essai.

Par ailleurs, le Tableau 5.1 synthétise les valeurs moyennes et les écarts-types globaux des fréquences propres et des taux d'amortissements évalués.

La deuxième campagne de mesures a été réalisée le 25 juin, le jour concernant la mise en œuvre des travaux. La procédure de resserrage a été effectuée en 4 phases pour le demi-pont instrumenté. Pendant la première phase, quatre premières bielles ont été resserrées. Puis, trois autres bielles ont été resserrées pendant la deuxième phase. Ces deux phases ont considéré les bielles situées côté Lyon. Pendant les troisième et quatrième phases, quatre, puis trois bielles



(a) Premier mode propre : flexion

(b) Deuxième mode propre : flexion/torsion

(c) Troisième mode propre : torsion

(d) Quatrième mode propre : flexion/torsion

FIGURE 5.7 – Allure des quatre premiers modes propres identifiés. (Crémona et al. 2004).

Mode	$\overline{f}$ [Hz]	σ_f [Hz]	$\overline{\xi}$ (%)	σ_ξ (%)
1	5,84	0,042	2,62	0,107
2	8,74	0,148	1,60	0,250
3	13,09	0,150	1,15	0,120
4	16,95	0,130	1,08	0,087

TABLEAU 5.1 – Valeurs moyennes et écarts-types des fréquences propres et amortissements (avant resserrage) (Crémona 2004).

ont été resserrées côté Paris, respectivement. Au total, 13 essais dynamiques ont été enregistrés (notés TGV1R à TGV13R). Le Tableau B.2 (cf. Annexe B) regroupe les valeurs moyennes et les écarts-types respectifs calculés pour chacun de ces essais. De plus, les valeurs moyennes et les écarts-types globaux des fréquences propres et coefficients d'amortissements sont détaillés dans les Tableaux 5.2 et B.3 pour chaque phase (cf. Annexe B), respectivement. Dans ces tableaux, la dernière ligne correspond aux valeurs moyennes par rapport aux quatre phases du resserrage. Il est noté que les fréquences propres et les taux d'amortissement augmentent significativement, ce qui en accord avec les travaux réalisés.

Finalement, la troisième campagne a eu lieu le 26 juin 2003. L'objectif a été de vérifier les effets du resserrage des bielles juste après la fin des travaux. Le Tableau B.4 (cf. Annexe B) regroupe les horaires, les valeurs moyennes des fréquences propres identifiées et les écarts-types pour les 13 essais enregistrés. Le Tableau 5.3 synthétise les valeurs moyennes et les écarts-types globaux des fréquences propres et des coefficients d'amortissements évalués.

Le Tableau 5.4 résume les fréquences propres identifiées pour chaque campagne de mesures.

Phase	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4	
	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
1	5,91	0,17	8,85	0,85	13,01	0,44	17,87	0,63
2	6,28	0,14	9,01	1,06	13,12	0,63	17,25	0,40
3	6,32	0,20	8,83	0,93	13,07	1,18	17,88	0,41
4	6,42	0,22	8,61	0,95	13,15	0,76	18,06	0,95
Moyenne	**6,25**	0,18	**8,81**	0,95	**13,08**	0,75	**17,79**	0,63

TABLEAU 5.2 – Valeurs moyennes et écarts-types des fréquences propres (pendant resserrage) (Crémona 2004). Valeurs en Hz.

Mode	$\overline{f}$ [Hz]	σ_f [Hz]	$\overline{\xi}$ (%)	σ_ξ (%)
1	6,47	0,16	4,15	2,62
2	8,87	0,42	2,98	1,62
3	13,15	0,30	2,35	2,08
4	17,94	0,66	2,19	5,24

TABLEAU 5.3 – Valeurs moyennes et écarts-types des fréquences propres et amortissements (après resserrage) (Crémona 2004).

Il est noté une augmentation des valeurs moyennes des quatre fréquences.

Mode	Avant resserrage	Pendant resserrage	Après resserrage
	$\overline{f}$ [Hz]	$\overline{f}$ [Hz]	$\overline{f}$ [Hz]
1	5,84	6,25	6,47
2	8,74	8,81	8,87
3	13,09	13,08	13,15
4	16,95	17,79	17,94

TABLEAU 5.4 – Comparatif des quatre premières fréquences propres identifiées dans chaque campagne de mesures.

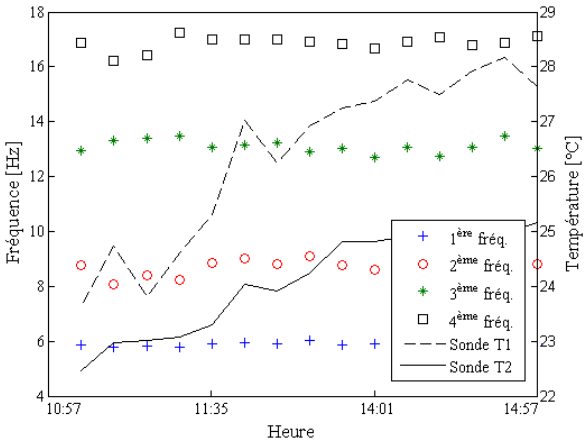
L'un des objectifs majeurs de ce chapitre est d'utiliser l'approche basée sur le couplage de l'ADS aux méthodes de classification non supervisée et supervisée, afin de discriminer les trois états structuraux du pont-rail PK 075+317 (i.e., avant, pendant et après le resserrage des bielles). Toutefois, avant de mettre en œuvre cette démarche, une étude des effets thermiques sur les paramètres modaux est menée.

5.3 Étude des effets thermiques sur les paramètres modaux

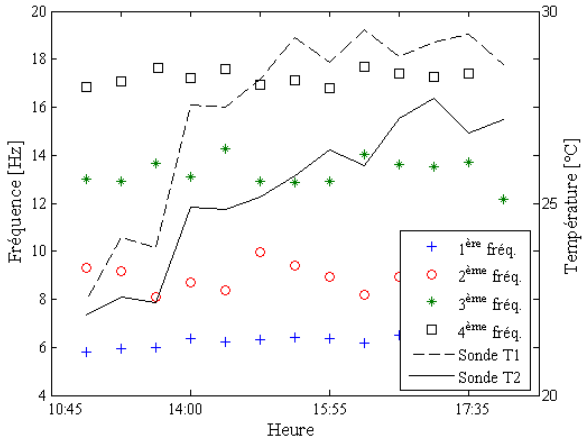
Ce paragraphe propose une étude des effets de la variation de la température sur les paramètres modaux. La démarche adoptée prend en compte les techniques de régression linéaire et non linéaire présentées dans le paragraphe 3.4. L'idée est de créer des lois de correction permettant de corréler les températures mesurées aux fréquences propres et modes propres identifiés. Une fois les lois déterminées, il est possible d'évaluer les nouvelles valeurs pour ces paramètres, maintenant considérés comme « compensés » ou, d'une certaine façon, « corrigés » des effets thermiques.

La Figure 5.8 illustre la variation des quatre premières fréquences propres et des températures mesurées par les des deux sondes T1 et T2 au cours des enregistrements des trois campagnes de 2003. Pour rappel, la sonde T1 a été placée au-dessus de l'acrotère et la sonde T2, au-dessous.

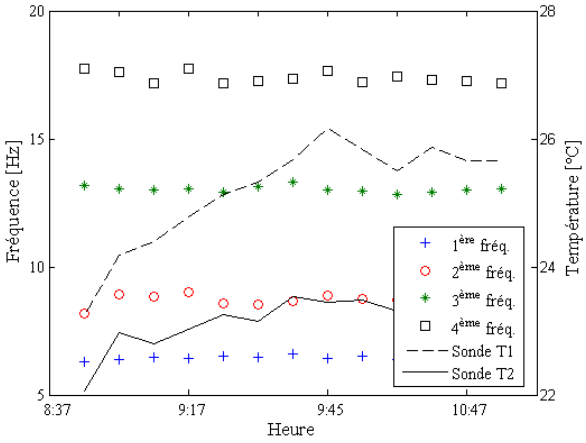
Il est noté une variation plus importante pour la première que pour la la seconde. Il est observé que les fréquences varient relativement peu au cours des enregistrements. Par ailleurs, la plage de température pendant les trois jours des campagnes de mesures est restée pratiquement la même, en oscillant entre 22 °C et 30 °C.



(a) Campagne avant resserrage



(b) Campagne pendant resserrage



(c) Campagne après resserrage

FIGURE 5.8 – Allure de la variation des quatre premières fréquences propres et des sondes de températures durant les campagnes d'acquisition.

Comme il a été décrit dans le paragraphe 3.4, avant d'appliquer les techniques de régression, les valeurs des fréquences propres et des composantes modales identifiées sont normalisées par rapport à une valeur de référence. Dans cette étude, la température 23 °C mesurée par la sonde T2 est utilisée comme température de référence. Ainsi, les fréquences (et composantes modales) sont normalisées par rapport à la valeur identifiée à cette température. Il est important de remarquer que les deux sondes de températures sont prises en compte pour la détermination des lois de correction. Dans les analyses suivantes, les ratios des paramètres modaux identifiés lors de la campagne avant resserrage des bielles sont utilisés pour la construction des lois de correction.

Le Tableau 5.5 regroupe les vecteurs des poids évalués pour chaque ratio des fréquences propres en utilisant des filtres linéaires. Dans ce tableau, β est le biais du modèle, w_{T1} et w_{T2} sont les poids qui multiplient les enregistrements de température des sondes T1 et T2, respectivement.

Les lois de correction basées sur la régression non linéaire sont déterminées en utilisant des RN (réseaux de neurones). L'architecture utilisée est composée de 10 neurones dans la couche cachée

	Fréquence 1	Fréquence 2	Fréquence 3	Fréquence 4
β	0.9838	0.8898	1.2493	0.9378
w_{T1}	0.0088	0.0319	0.0064	0.0031
w_{T2}	-0.0084	-0.0286	-0.0182	0.0003

TABLEAU 5.5 – Vecteurs des poids évalués pour chaque loi de correction thermique (régression linéaire).

et les fonctions d'activation utilisées sont la tangente hyperbolique (dans la couche cachée) et la sigmoïde (dans la couche de sortie). Les données d'entrée sont les enregistrements de température des sondes T1 et T2 et les ratios des paramètres modaux, les données de sortie. L'entraînement du réseau est réalisé en utilisant 10 essais de l'état « avant ». La validation utilise les 5 autres essais (également de l'état « avant »). La démarche consiste alors à réaliser 1000 tirages aléatoires (pour différents groupes d'entraînement et de validation). Le modèle choisi est celui dont l'erreur commise lors de phase de validation est la plus faible. Étant donné que le nombre de poids du réseau calibré est conséquent (20 dans la couche cachée + 10 dans la couche de sorties + 11 biais), les valeurs ne sont pas détaillés.

Une fois les modèles de régression construits, les fréquences propres (et composantes modales) compensées peuvent être évaluées comme décrit dans le paragraphe 3.4. La Figure 5.9 donne un aperçu de la variation des quatre premières fréquences propres compensées et non compensées des effets thermiques pour chaque état structural. Dans cette étude, il est observé que les deux filtres ne modifient pas significativement la première fréquence propre. Toutefois, dans le cas des troisième et quatrième fréquences propres, les différences sont plus marquées.

L'objectif majeur de l'étude courante est de comparer les résultats obtenus par les méthodes de clustering et de classification supervisée appliquées aux paramètres modaux corrigés, avec ceux évalués en considérant la prise en compte des effets thermiques. Une analyse classique est toutefois proposée au préalable. Cette analyse s'appuie sur des outils statistiques et sur quelques techniques citées dans le Chapitre 1 pour discriminer les différents comportements structuraux.

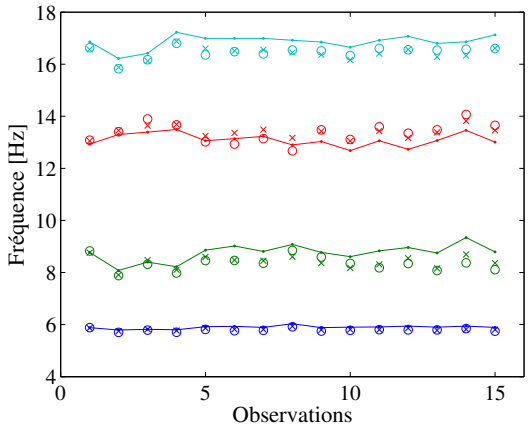
5.4 Analyse Classique

5.4.1 Sans prise en compte des effets thermiques

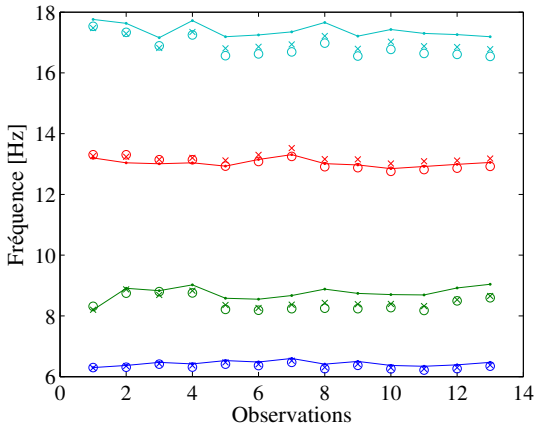
Le but de cette analyse est de montrer que l'utilisation d'outils et de méthodes dits classiques ne permet généralement pas de conclure, avec un niveau de confiance satisfaisant, que les trois états structuraux - avant, pendant et après resserrage - du pont PK 075+317 sont associés à un processus de modification structurale.

La Figure 5.10 illustre les histogrammes pour les quatre premières fréquences propres identifiées dans les états avant et après resserrage. Dans cette figure, les traits verticaux pointillés représentent les intervalles de confiance à 95% pour les valeurs moyennes de fréquences calculées. Bien que la différence entre les deux états structuraux soit évidente pour la première fréquence (les traits verticaux ne se chevauchent pas), ceci n'est pas le cas pour les autres. En analysant uniquement ces histogrammes, il n'est donc pas possible d'affirmer, avec 95% de confiance, que les états structuraux sont différents.

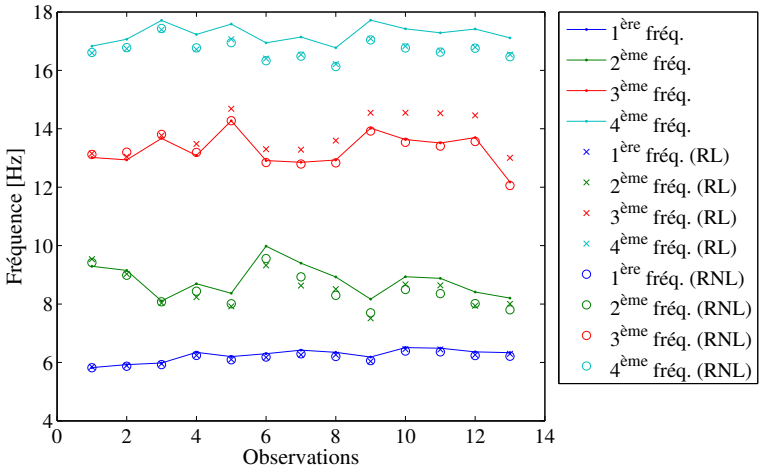
Par ailleurs, des tests d'hypothèse sont menés en utilisant les fréquences propres identifiées avant et après resserrage. De façon similaire à celle présentée dans le chapitre précédent, l'hypothèse nulle du test est que les fréquences sont issues d'une même distribution (ce qui implique à la non détection des effets des travaux réalisés) et l'hypothèse alternative, le contraire. Le niveau



(a) Campagne avant resserrage



(b) Campagne pendant resserrage



(c) Campagne après resserrage

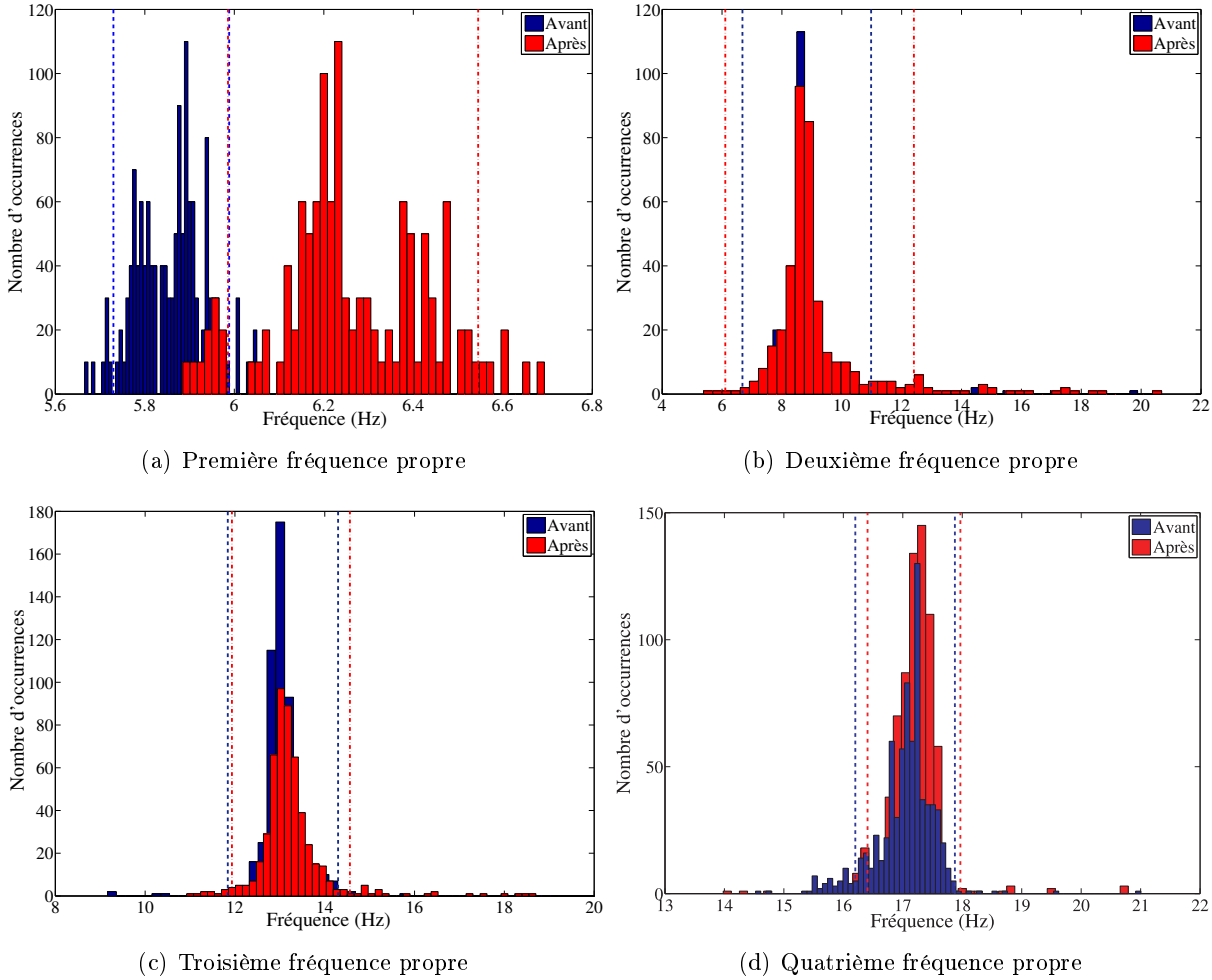
FIGURE 5.9 – Allure de la variation des quatre premières fréquences propres sans et avec correction des effets thermiques par la régression linéaire (RL) et non linéaire (RNL).

d'acceptation des tests est toujours fixé à 5%. Il est rappelé que l'hypothèse nulle est acceptée dans le cas où p^* est supérieur au niveau d'acceptation. Dans le cas contraire, l'hypothèse nulle est rejetée. La valeur de p^* est calculée en utilisant l'Équation (4.2). Le Tableau 5.6 regroupe les résultats obtenus pour chaque test réalisé.

$1^{ère}$ fréq.		$2^{ème}$ fréq.		$3^{ème}$ fréq.		$4^{ème}$ fréq.	
H	p^* (%)	H	p^* (%)	H	p^* (%)	H	p^* (%)
1	0	0	8,8	1	3,5	1	2,6

TABLEAU 5.6 – Résultats des tests d'hypothèse de Kolmogorov-Smirnov pour les quatre premières fréquences propres (états avant et après resserrage des bielles).

Les résultats obtenus montrent que seules les distributions de la deuxième fréquence propre sont considérées comme similaires par le test d'hypothèse. Il est remarqué que pour la première fréquence, la valeur p^* est égal à zéro, ce qui ne suscite aucun soupçon sur la non acceptation de l'hypothèse nulle. Par ailleurs, il est noté que pour les troisième et quatrième fréquences propres,



(a) Première fréquence propre

(b) Deuxième fréquence propre

(c) Troisième fréquence propre

(d) Quatrième fréquence propre

FIGURE 5.10 – Comparatif des histogrammes pour les quatre premières fréquences propres avant et après resserrage.

l'hypothèse nulle est également rejetée. En effet, les valeurs p^* sont inférieures à 5%, dans ces deux derniers cas.

Une analyse comparative tenant compte des déformées modales est également proposée. Dans cette analyse, les coefficients MAC sont calculés pour chacune des lignes de capteurs (voir Figure 5.4) afin d'évaluer la corrélation entre les modes propres. La Figure 5.11 illustre les quatre premières déformées modales et les coefficients MAC respectifs. Il est noté que la variation des modes est relativement faible, à l'exception des modes 2 et 3 qui semblent être plus sensibles aux modifications structurales apportées par les travaux de resserrage. Cela est directement reflété sur le calcul des coefficients MAC. En effet, pour les premier et quatrième modes propres, les coefficients varient entre 86% et 100%, ce qui dénote, pour cet indicateur, une absence de sensibilité des modes propres par rapport aux effets du resserrage. Toutefois, les modes 2 et 3 ont des variations légèrement plus importantes, ce qui implique des coefficients MAC plus faibles (entre 50 et 70%).

Les résultats obtenus par l'analyse classique sont, d'une certaine façon, antagoniques. En analysant les histogrammes et les tests d'hypothèse, il a été montré que les distributions des première, troisième et quatrième fréquences propres sont différentes. Toutefois, en observant les coefficients MAC, les déformées modales associées sont celles les moins sensibles aux travaux menés. D'ailleurs, le même paradoxe est observé pour la deuxième fréquence propre, pour laquelle

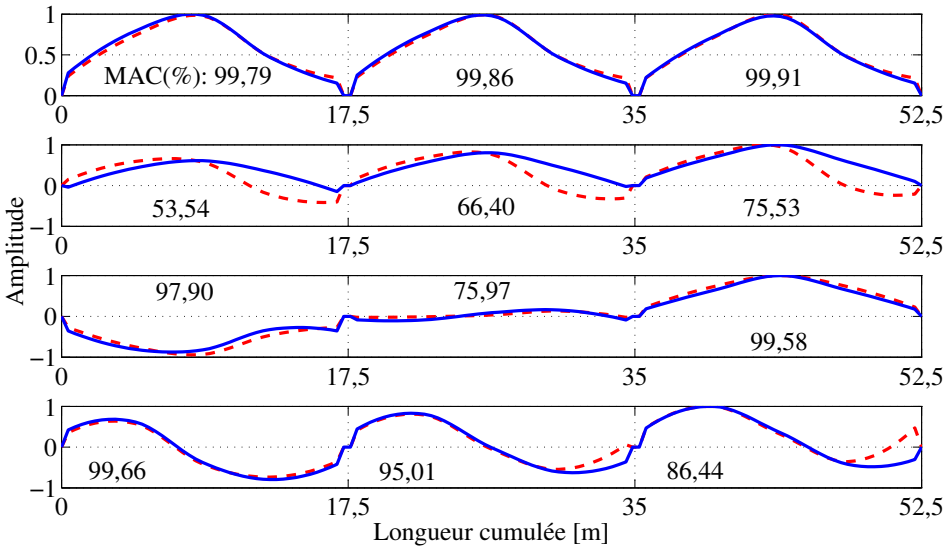


FIGURE 5.11 – Comparatif des quatre premiers modes propres avant et après resserrage et calcul de MAC.

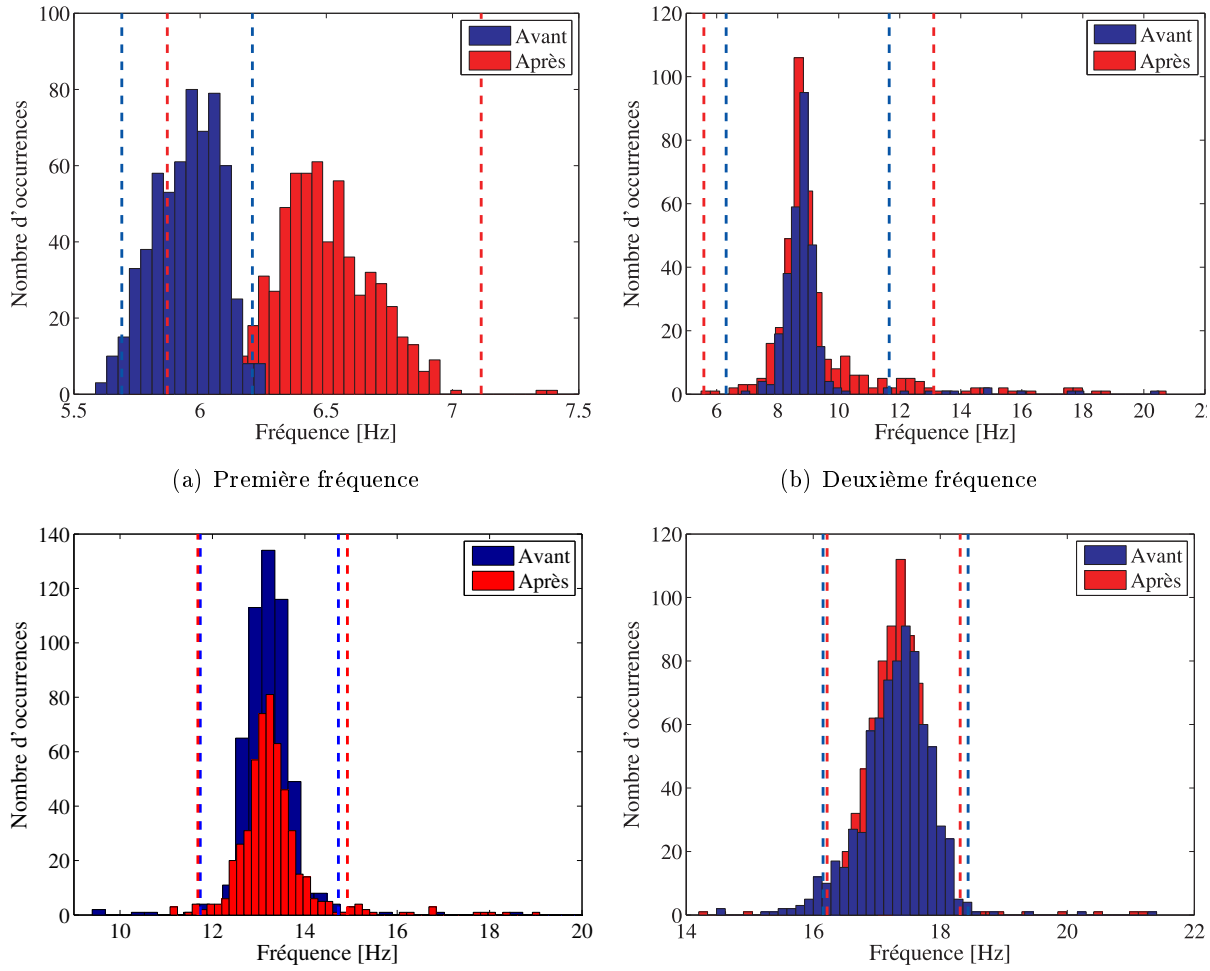
le test d'hypothèse a accepté l'hypothèse nulle et les coefficients MAC des modes propres ont été plus faibles.

Ainsi, comme il a été montré dans le chapitre précédent, à partir de l'étude de sensibilité d'une poutre numérique, les analyses classiques ne sont généralement pas suffisamment sensibles pour discriminer des modifications structurales. Par ailleurs, lorsque les paramètres modaux sont analysés directement, il est difficile de conclure si les différences entre leurs valeurs sont liées aux effets des travaux réalisés ou à d'autres phénomènes. Il est donc justifié d'appliquer des analyses complémentaires afin de mieux discriminer différents états structuraux.

5.4.2 Avec prise en compte des effets thermiques

Préalablement à l'étude symbolique des paramètres modaux corrigés, les analyses classiques exposées précédemment sont reconsidérées. La Figure 5.12 montre les histogrammes des quatre premières fréquences propres corrigées par le filtre linéaire, en comparant les états avant et après resserrage des bielles. Il est intéressant de noter que les observations faites précédemment en analysant l'évolution des moyennes des fréquences propres (cf. Figure 5.9) ne sont plus valables. En effet, seuls les intervalles de confiance calculés pour la première fréquence propre ne se chevauchent pas. Toutefois, pour les autres fréquences il n'est pas possible de conclure, avec 95% de confiance, que les distributions sont différentes. Une analyse similaire est menée en considérant la correction par des filtres non linéaires (cf. Figure B.1). Il est noté qu'il y a peu de différence entre les deux types de correction dans l'exemple considéré. En effet, pour des échantillons de données relativement petits, il est parfaitement envisageable que les lois construites par des modèles linéaires et non linéaires soient similaires.

Les tests d'hypothèse sont également reconsidérés. Les échantillons des fréquences propres identifiées avant et après resserrage des bielles sont soumis au test. De façon similaire à celle présentée précédemment, le niveau d'acceptation des tests est fixé à 5%. Pour rappel, si la valeur p^* (cf. Équation (4.2)) du test est supérieure au niveau d'acceptation, il est considéré que les échantillons sont issus de distributions différentes. Si p^* est inférieur à 5%, les échantillons peuvent être considérés issus d'une même distribution. Le Tableau 5.7 regroupe les résultats obtenus en



(a) Première fréquence

(b) Deuxième fréquence

(c) Troisième fréquence

(d) Quatrième fréquence

FIGURE 5.12 – Comparatif des histogrammes pour les quatre premières fréquences propres des états avant et après resserrage (régression linéaire).

considérant les quatre fréquences propres corrigées par le filtre linéaire.

$1^{\text{ère}}$ fréq.		$2^{\text{ème}}$ fréq.		$3^{\text{ème}}$ fréq.		$4^{\text{ème}}$ fréq.	
H	p^* (%)	H	p^* (%)	H	p^* (%)	H	p^* (%)
1	0	1	0,73	1	2,45	0	38,9

TABLEAU 5.7 – Résultats des tests d'hypothèse de Kolmogorov-Smirnov pour les quatre premières fréquences propres (états avant et après ; régression linéaire).

Contrairement à ce qui a été observé dans le Tableau 5.6 (sans correction thermique), les résultats obtenus montrent que seules les distributions de la quatrième fréquence propre sont considérées comme différentes par le test d'hypothèse. Il est remarqué que pour la première fréquence, la valeur p^* est toujours égale à zéro, ce qui ne suscite aucun soupçon sur la non acceptation de l'hypothèse nulle. Par ailleurs, il est noté que pour les deuxième et troisième fréquences propres, le test a échoué également. Seules les distributions de la quatrième fréquence propre sont considérées similaires. Dans ce dernier cas, la valeur p^* est significativement supérieure à 5% (le niveau de signification du test). Mis à part les valeurs p^* calculées, des résultats identiques sont obtenus en considérant le filtre non linéaire (cf. Tableau B.5). Toutefois, il est

noté que la valeur p^* pour la quatrième fréquence a diminué. Cela montre que le filtre non linéaire a pu, dans un certain sens, « éloigner » les distributions des fréquences propres.

Une analyse comparative tenant compte des déformées modales compensées est également proposée. Dans cette analyse, les coefficients MAC sont calculés pour chacune des lignes de capteurs (voir Figure 5.4) afin d'évaluer la corrélation entre les modes propres. La Figure 5.13 illustre les quatre premières déformées modales corrigées par le filtre linéaire et les MAC respectifs. Il est noté que la variation des modes reste relativement faible pour les modes 1 et 4. Inversement, les modes 2 et 3 sont plus sensibles aux modifications structurales, même après la correction thermique. En général, le calcul des coefficients MAC mène à des valeurs légèrement plus faibles pour tous les modes. Cela indique que la correction thermique permet d'écarter les composantes modales des modes propres avant et après resserrage. La même démarche est appliquée en utilisant la régression non linéaire. Dans ce cas, les coefficients MAC sont encore plus faibles que ceux évalués précédemment (cf. Figure B.2).

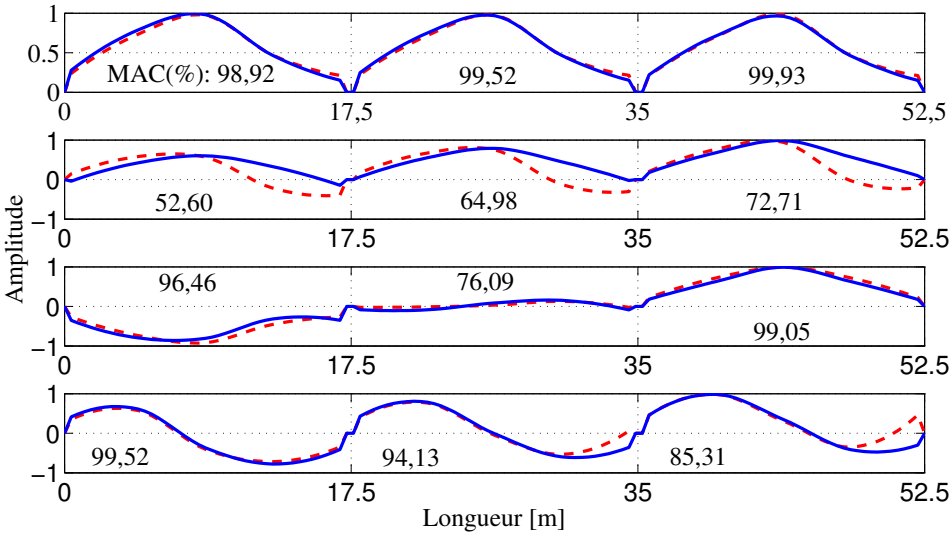


FIGURE 5.13 – Comparatif des quatre premiers modes propres avant et après resserrage et calcul de MAC (régression linéaire).

Mis à part les tests d'hypothèse qui permettent une meilleure discrimination des comportements structuraux, les analyses classiques ne donnent généralement pas une conclusion claire sur les effets du renforcement mené, même après correction thermique. Les méthodes de classification couplées à l'ADS sont maintenant appliquées dans la suite aux paramètres modaux corrigés. L'objectif est d'apprécier d'une part la pertinence des techniques de régression proposées et d'autre part, l'impact sur les classifications obtenues.

5.5 ADS couplée aux méthodes de clustering

À l'instar des applications numériques présentées au Chapitre 4, le couplage de l'ADS avec les méthodes de clustering est maintenant appliqué aux essais réalisés sur le pont-rail PK 075+317. Dans les trois campagnes de mesures, un total de 41 essais a été enregistré : 15 essais pour l'état avant resserrage, 13 essais pour l'état de l'ouvrage pendant les travaux et 13 essais après la réalisation du renforcement. L'objectif est d'utiliser les techniques de classification des données (cf. Chapitre 3) pour séparer l'ensemble des 41 essais dynamiques en trois groupes distincts. Dans un premier temps, ces méthodes sont appliquées aux représentations symboliques des signaux. Dans

un deuxième temps, la même procédure est appliquée aux fréquences propres. Finalement, les modes propres sont utilisés pour la discrimination des états structuraux. Toutes les simulations de clustering décrites dans les paragraphes suivants considèrent les quatre premières fréquences propres et leurs déformées modales associées. Dans ce chapitre, deux représentations symboliques sont utilisées. La première, basée sur des intervalles interquartiles et la deuxième, sur des histogrammes à 20 catégories. Dans les analyses suivantes, seuls les résultats obtenus à partir de la transformation en histogrammes sont donnés. Dans le cas des intervalles, les résultats sont montrés dans l'Annexe B de ce mémoire.

5.5.1 Signaux

Les premiers résultats correspondent à l'application de la méthode des divisions hiérarchiques aux signaux. Il est manifeste que le partitionnement idéal est celui capable de discriminer les 3 groupes d'essais de manière homogène (i.e., chaque groupe ne contient que des essais correspondant à un seul état structural). La Figure 5.14 illustre l'arbre construit pour les signaux transformés en histogrammes. Les taux de classification sont légèrement plus faibles que ceux obtenus par des intervalles (cf. Figure B.3) : 73% (11/15) pour l'état avant, 62% (8/13) pour l'état pendant et 85% (11/13) pour l'état après. Dans ce cas, il est noté que les capteurs 10 et 3 (situés proche de l'appui) ont été choisis comme les variables discriminantes pour la première et seconde divisions, respectivement. Par ailleurs, un essai « avant » est classé dans le cluster « après », ce qui montre que l'utilisation des histogrammes dans cette analyse peut ne pas être la plus adéquate. Il est observé que les essais de l'état pendant tendent à se mélanger relativement plus que les essais des autres états. Par exemple, le premier cluster contient pratiquement tous les essais « avant », mais il existe aussi deux essais (TGV1R et TGV2R) qui, théoriquement, ne devraient pas y être. Toutefois, étant donné que le resserrage des bielles a été une effectué de manière progressive, il est possible de dire que les réponses dynamiques des premiers essais « pendant » les travaux sont plus proches de celles enregistrées « avant » resserrage.

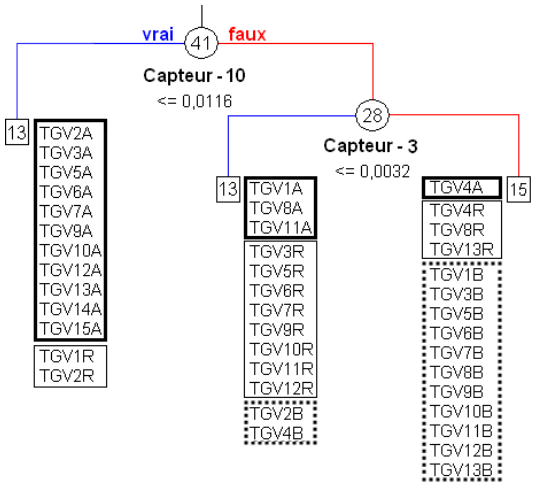


FIGURE 5.14 – Classification des signaux en utilisant les divisions hiérarchiques (histogrammes).

La méthode basée sur les agglomérations hiérarchiques est maintenant utilisée. Les trois clusters obtenus sont identifiés à la Figure 5.15. En général, un nombre plus important d'essais a été classé incorrectement. En effet, les essais sont beaucoup plus mélangés par rapport aux partitions obtenues par la méthode des divisions hiérarchiques. Bien que la plupart des essais « avant » soient classés dans un cluster séparé, il y en d'autres qui se mélangent dans le cluster de l'état « après ». Le même résultat est observé pour les essais « pendant ». Comme il a été

noté dans le chapitre de l'application numérique, le critère de distance utilisé dans cette méthode peut ne pas être suffisamment sensible pour discriminer les états structuraux. Dans le cluster « avant », 67% (10/15) des essais sont correctement classés. Dans les clusters « pendant » et « après », les taux sont de 62% (8/13) et 54% (7/13), respectivement. Il est toutefois noté qu'il existe un changement graduel entre les états structuraux pour les deux hiérarchies obtenues (cf. Figure B.4, dans le cas des intervalles). En d'autres termes, le groupe d'essais « avant » est contigu au cluster « pendant », qui est voisin du groupe « après ». Cela montre qu'en général, la méthode met en évidence la notion de dissimilarité (distance) entre les états structuraux.

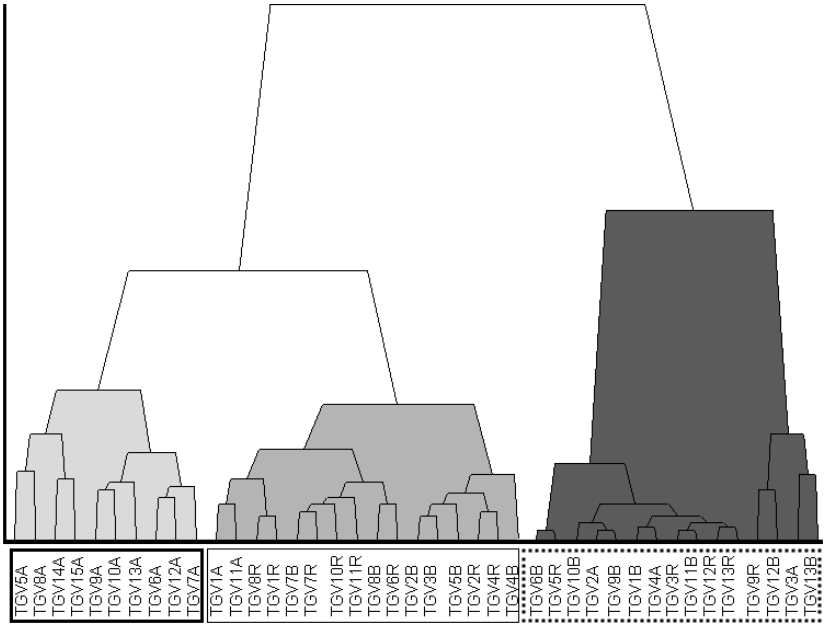


FIGURE 5.15 – Classification des signaux en utilisant les agglomérations hiérarchiques (histogrammes).

Par ailleurs, la méthode de clustering, basée sur les nuées dynamiques, est appliquée aux signaux. En général, cette méthode permet d'obtenir des meilleurs résultats que ceux obtenus par les divisions hiérarchiques et les agglomérations hiérarchiques. Le taux de classification correcte est de 80% (12/15) pour le cluster « avant », 85% (11/13) pour l'état « pendant » et 85% (11/13) pour l'état « après ». Dans le cas de l'application considérée, il est possible de dire que cette méthode est la plus adaptée. Il est également noté que les taux de classification obtenus utilisant des histogrammes sont relativement plus importants que ceux évalués en utilisant des intervalles (cf. Annexe B).

5.5.2 Fréquences propres

5.5.2.1 Sans prise en compte des effets thermiques

Les méthodes de partitionnement sont maintenant appliquées aux quatre premières fréquences propres transformées en données symboliques. Si la transformation en histogrammes est considérée (Figure 5.16), la classification pour le cluster « avant » est meilleure que celle obtenue en utilisant la transformation en intervalles (cf. Figure B.5) : 93% (14/15). Toutefois, deux essais du groupe « après » (ici, 85%) se retrouvent dans le cluster « pendant » (85%). La variable la plus discriminante est la fréquence 2 pour des deux divisions. Bien que la différence entre la première fréquence des états « avant » et « après » soit plus nette (cf. Figure 5.10), il est intéressant

d'observer qu'elle n'est jamais choisie pour la séparation des groupes. En effet, la détermination des clusters est réalisée à partir de la minimisation des intra-inerties, ce qui ne suit pas nécessairement le schéma d'une discrimination (ou observation) classique.

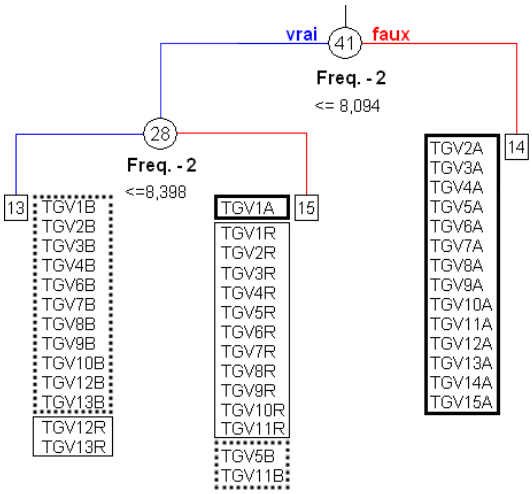


FIGURE 5.16 – Classification des fréquences propres en utilisant les divisions hiérarchiques (histogrammes).

La méthode des agglomérations hiérarchiques est maintenant considérée. La Figure 5.17 illustre les partitionnements obtenus lorsque des histogrammes sont utilisés. Dans ce cas, la disposition des clusters (avant → pendant → après) suit correctement la tendance des états structuraux existants. Les taux de classification correcte sont en général moins importants que ceux obtenus par des intervalles (cf. Figure B.6), étant donné que 87% (13/15) des essais sont bien classés dans le cluster « avant », 54% (7/13) le sont pour l'état « pendant » et 62% (8/13) pour l'état « après ».

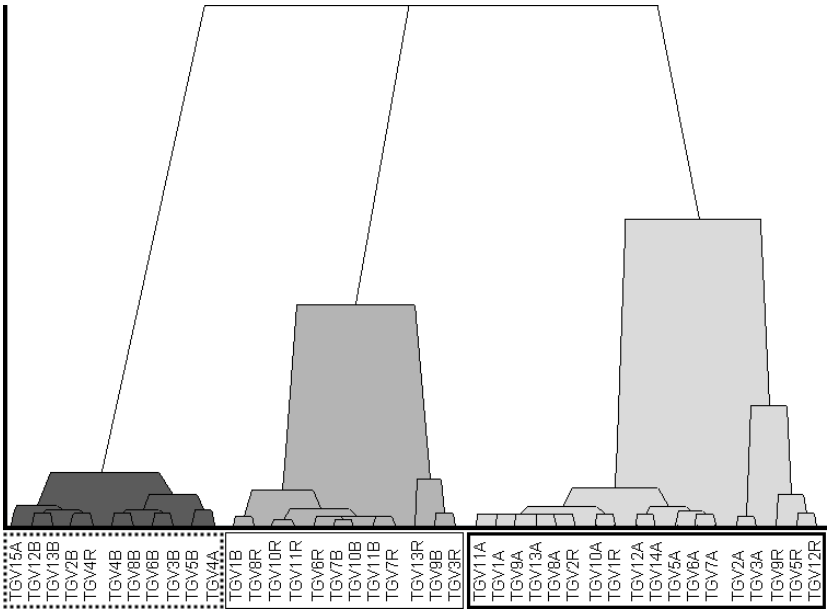


FIGURE 5.17 – Classification des fréquences propres en utilisant les agglomérations hiérarchiques (histogrammes).

Les résultats obtenus en utilisant les nuées dynamiques montrent que cette méthode semble

être plus robuste et adéquate pour la détection de modifications structurales. Dans le cas de la transformation en histogrammes, le taux de classification correcte pour l'état « avant » est de 87%, 80% pour l'état « pendant » et 85% pour l'état « après ». Ces résultats sont pratiquement du même ordre que ceux obtenus par des intervalles interquartiles. Toutefois, il est rappelé que les résultats obtenus par les deux types de transformation en données symboliques doivent être analysés de manière complémentaire. Étant donné que chaque application présente des caractéristiques différentes, il est parfaitement envisageable qu'une représentation soit plus adéquate que l'autre. Cette observation est également valable pour les méthodes de clustering utilisées.

5.5.2.2 Avec prise en compte des effets thermiques

Dans un premier temps, les fréquences corrigées par le filtre linéaire transformées en histogrammes sont utilisées pour appliquer la méthode des divisions hiérarchiques. La Figure 5.18 illustre l'arbre de décision construit et les clusters obtenus pour chacun des états. Il est noté que le partitionnement obtenu ne présente pas de différences par rapport à celui évalué sans la correction thermique (cf. Figure 5.16). En effet, la classification pour le cluster « avant » est la même : 87% (13/15). Toutefois, les taux pour les autres clusters sont moins importants : 77% (10/13) pour l'état « pendant » et 85% (11/13) pour l'état « après ».

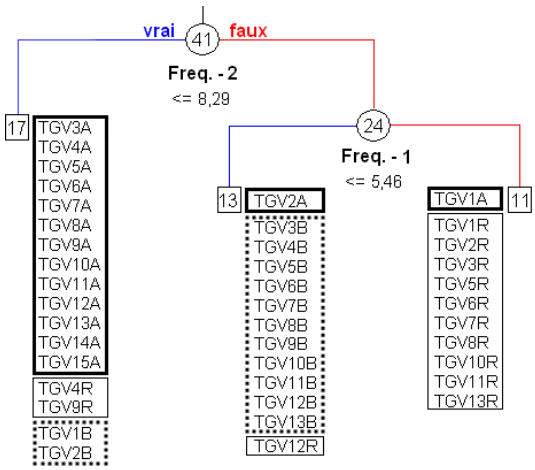


FIGURE 5.18 – Classification des fréquences propres en utilisant les divisions hiérarchiques (histogrammes ; régression linéaire).

La méthode des agglomérations hiérarchiques est également utilisée. Il est noté que l'utilisation du filtre n'altère pas significativement les taux de classification correcte obtenus : en utilisant des histogrammes, les taux sont de 87%, 54% et 85% pour ces trois états, respectivement (cf. Figure 5.19). Il existe quelques essais qui sont regroupés différemment pour chaque type de transformation symbolique, mais il est généralement observé une quantité élevée d'essais mélangés.

La méthode des nuées dynamiques est aussi considérée. Les résultats obtenus confirment les remarques précédentes. Le taux de classification correcte est de 87% (13/15) pour le cluster « avant », 77% (10/13) pour l'état « pendant » et 85% (11/13) pour l'état « après ».

Les simulations suivantes considèrent des filtres non linéaires (basés sur des réseaux de neurones) pour la correction des fréquences propres. L'objectif de ces simulations est non seulement de corriger les effets liés à la variation de températures, mais aussi de comparer les résultats avec ceux obtenus en utilisant le filtre linéaire.

En menant une procédure similaire à celle déjà décrite, les divisions hiérarchiques sont ap-

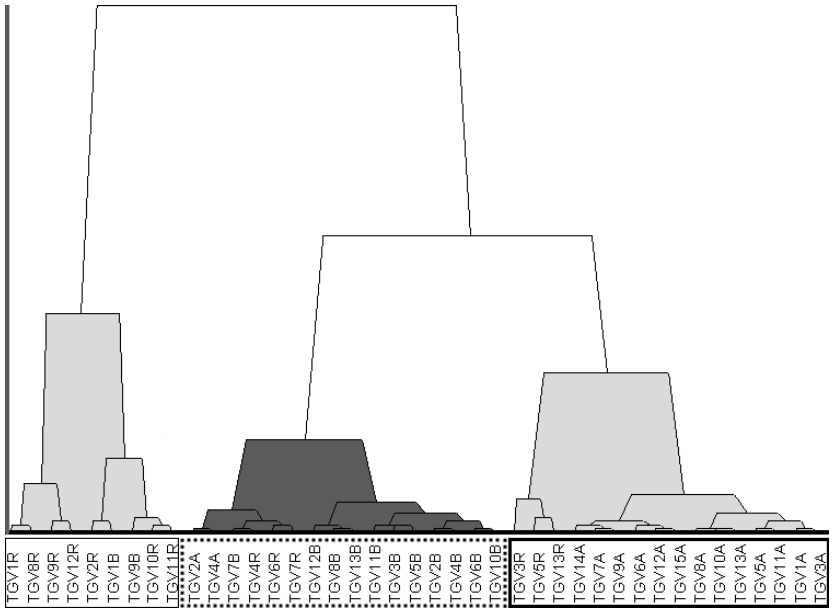


FIGURE 5.19 – Méthode des agglomérations hiérarchiques appliquée aux fréquences propres corrigées (histogrammes ; régression linéaire).

pliquées aux fréquences propres corrigées par le filtre non linéaire. L'arbre de décision obtenu en utilisant des histogrammes est montré à la Figure 5.20. Les résultats sont identiques à ceux obtenus en utilisant le filtre linéaire (cf. Figure 5.18). Cela permet d'indiquer qu'une correction des effets thermiques sur les mesures peut ne pas être nécessaire dans ce cas particulier.

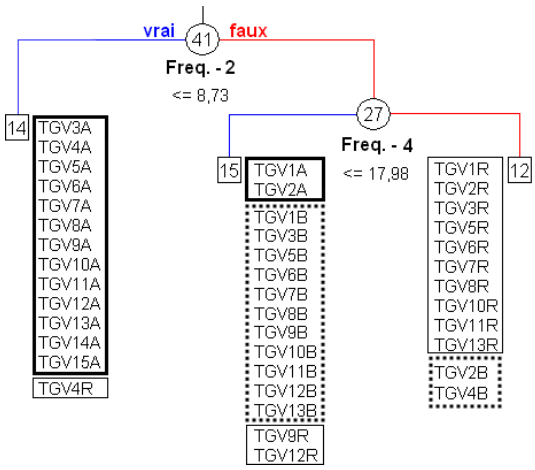


FIGURE 5.20 – Classification des fréquences propres en utilisant les divisions hiérarchiques (histogrammes ; régression non linéaire).

En considérant les agglomérations hiérarchiques, les résultats sont moins pertinents en utilisant le filtre non linéaire, comme il est montré à la Figure 5.21. Dans ce cas, le meilleur taux de classification correcte obtenu est de 70% pour l'état « après ». Pour l'état « avant », le taux est de 73% et 54% pour l'état « pendant ».

En utilisant la méthode des nuées dynamiques, les taux de classification sont relativement plus élevés que ceux obtenus par le filtre linéaire. Toutefois, les résultats ne sont pas meilleurs que ceux obtenus avec une analyse réalisée sans la correction thermique. Dans le cas de la transformation en histogrammes, le taux de classification correcte est de 80% (12/15) pour le cluster « avant »,

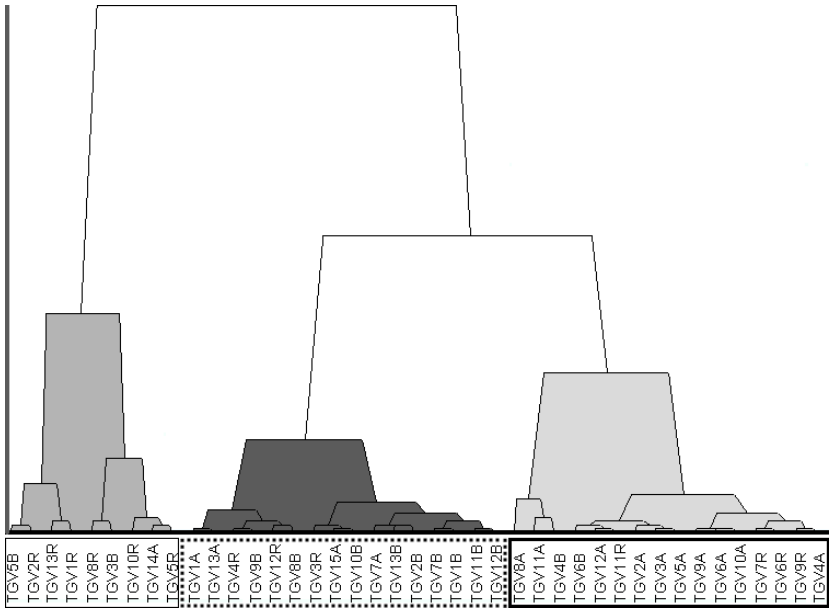


FIGURE 5.21 – Méthode des agglomérations hiérarchiques appliquée aux fréquences propres corrigées (histogrammes ; régression non linéaire).

77% (10/13) pour l'état « pendant » et 77% (10/13) pour l'état « après ».

5.5.3 Modes propres

5.5.3.1 Sans prise en compte des effets thermiques

Les quatre premiers modes propres sont transformés en données symboliques et utilisés pour appliquer la méthode des divisions hiérarchiques. Le partitionnement obtenu en utilisant la transformation en histogrammes est montré à la Figure 5.22. Dans ce cas, il existe un essai « après » que se retrouve dans le groupe « avant ». Bien que le taux de classification correcte pour ce dernier soit élevé (93%, soit 14 essais sur 15), il est plus faible pour les clusters « pendant » et « après » avec 77% de classification correcte dans les deux cas, soit 10 essais sur 13. En général, les résultats ne sont pas aussi pertinents que ceux obtenus en utilisant les fréquences propres.

La méthode des agglomérations hiérarchiques est maintenant appliquée en considérant les modes propres. La Figure 5.23 illustre la hiérarchie obtenue lorsque la représentation par des histogrammes est considérée. Il est noté que des résultats légèrement meilleurs sont obtenus. Pour le cluster « avant », 62% (8/15) d'essais sont correctement classés, 54% (7/13) le sont pour l'état « pendant » et 69% (9/13) pour l'état « après ». Ce dernier est le seul cluster homogène obtenu. De plus, il est noté que dans ces deux analyses (cf. Figure B.12, pour des intervalles), les dispositions des clusters sont inversées (avant → après → pendant). Comme il a été mentionné dans le cas des fréquences propres, cette méthode peut ne pas être la plus pertinente pour une étude de discrimination d'états structuraux.

La méthode des nuées dynamiques est maintenant envisagée. Le taux de classification correcte est de 80% (12/15) pour le cluster « avant », 77% (10/13) pour l'état « pendant » et 77% (10/13) pour l'état « après ». Ces résultats sont significativement meilleurs que ceux obtenus par la représentation en intervalles interquartiles (cf. Annexe B).

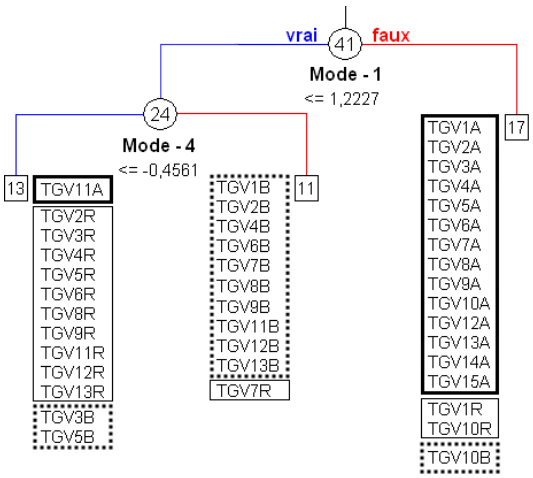


FIGURE 5.22 – Classification des modes propres en utilisant les divisions hiérarchiques (histogrammes).

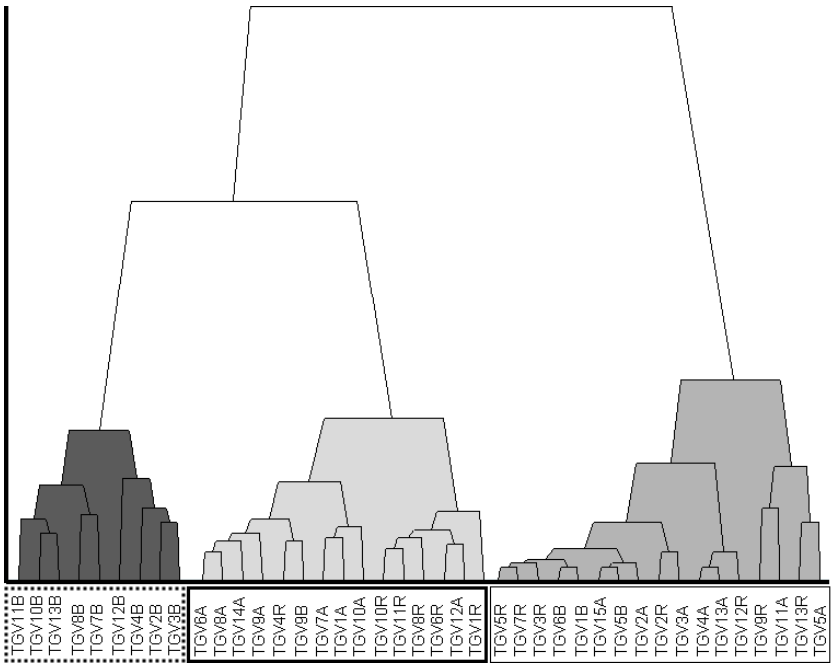


FIGURE 5.23 – Classification des modes propres en utilisant les agglomérations hiérarchiques (histogrammes).

5.5.3.2 Avec prise en compte des effets thermiques

Les modes propres sont maintenant considérés en effectuant la correction de chacune des composantes modales. Le partitionnement obtenu en utilisant la transformation en histogrammes est montrée à la Figure 5.24. Bien que l'arbre de décision soit relativement différent de celui construit sans prise en compte des effets thermiques (cf. Figure 5.22), les résultats sont relativement similaires : 73%, soit 11 essais sur 15, pour l'état « avant », 70% pour l'état « pendant » et 62% pour « après ».

La méthode des agglomérations hiérarchiques est maintenant considérée. La Figure 5.25 illustre la hiérarchie obtenue si la représentation par des histogrammes est considérée. Il est noté que les résultats sont pratiquement les mêmes que ceux obtenus par des intervalles inter-

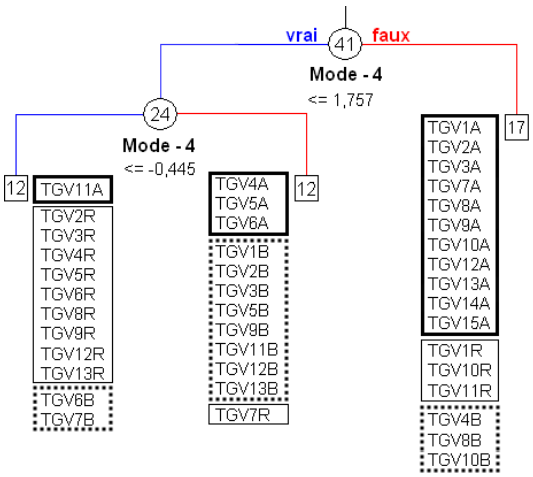


FIGURE 5.24 – Classification des modes propres en utilisant les divisions hiérarchiques (histogrammes ; régression linéaire).

quartiles (cf. Figure B.14) . Pour le cluster « avant », 53% (8/15) d'essais sont correctement classés, 62% (8/13) pour l'état « pendant » et 46% (6/13) pour l'état « après ».

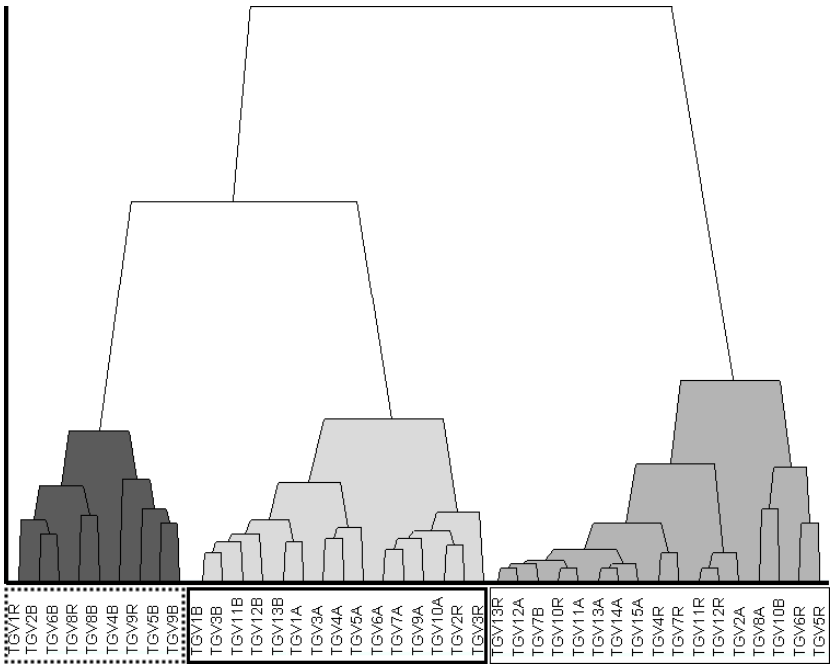


FIGURE 5.25 – Méthode des agglomérations hiérarchiques appliquée aux modes propres corrigées (histogrammes ; régression linéaire).

Enfin, la méthode des nuées dynamiques est utilisée. Bien que les résultats utilisant des intervalles interquartiles ne soient pas très pertinents (cf. Annexe B), une amélioration relative est observée lorsque la représentation par des histogrammes est utilisée. Dans ce cas, le taux de classification correcte est égal à 73% (11/15) pour l'état « avant », 77% (10/13) pour l'état « pendant » et 77% (10/13) pour l'état « après ». En général, les taux de classification ici évalués dans cette étude ne présentent aucune évolution par rapport à ceux obtenus sans l'approche de correction thermique.

La technique de régression non linéaire est maintenant appliquée aux modes propres corrigés.

La Figure 5.26 illustre l'arbre de décision construit en considérant des histogrammes. Dans ce cas, il est noté une sensible amélioration (par rapport au filtre linéaire) dans la classification des essais. Dans ce cas, 100% de essais sont classés correctement dans le cluster « avant », 77% pour l'état « pendant » et 70% pour l'état « après ». Toutefois, ces résultats ne permettent pas de conclure si le filtre de correction est cohérent, étant donné que les taux de classification ne sont pas largement différents de ceux déjà obtenus (sans correction thermique).

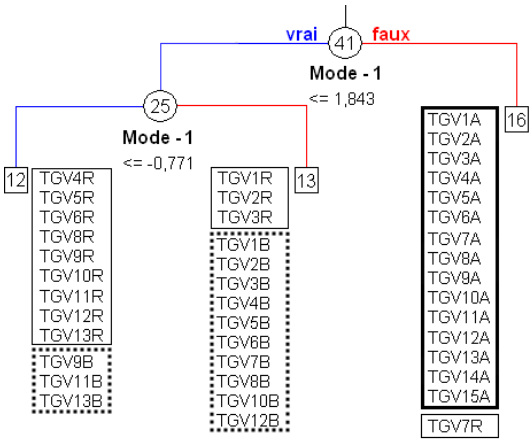


FIGURE 5.26 – Classification des modes propres en utilisant les divisions hiérarchiques (histogrammes ; régression non linéaire).

En utilisant la méthode des agglomérations hiérarchiques, les taux de classification correcte sont significativement plus faibles. En considérant des histogrammes, les résultats sont en général similaires à ceux obtenus par des intervalles (cf. Figure B.16) : 67% pour l'état « avant », 70% pour l'état « pendant » et 46% pour l'état « après » (Figure 5.27). De manière générale, les agglomérations obtenues ne changent significativement pas quel que soit le type de filtre utilisé. Bien que les clusters pour chaque analyse soient différents, cette méthode a déjà montré être la moins adéquate pour la discrimination d'états structuraux.

Les résultats obtenus par la méthode des nuées dynamiques sont légèrement meilleurs que ceux obtenus par le filtre linéaire. Les taux de classification correcte sont de 80% (12/15) pour l'état « avant », 77% (10/13) pour l'état « pendant » et 70% (9/13) pour l'état « après ». Toutefois, la prise en compte de la correction thermique ne permet pas d'améliorer significativement la classification des essais.

Les résultats mettent en évidence le fait que l'utilisation des filtres thermiques peut, au lieu d'améliorer les taux de classification, les dégrader dans certains cas. Deux hypothèses sont formulés pour expliquer ce phénomène : i) les variations thermiques peuvent ne pas jouer un rôle conséquent sur les paramètres modaux (les campagnes des mesures étant réalisées sur une même plage de températures, il n'est pas forcément nécessaire d'utiliser les procédures de correction thermique) ; ii) les erreurs introduites par les filtres thermiques causent un effet « parasite » sur les valeurs originales des paramètres modaux, en dégradant les taux de classification obtenus.

5.5.4 Étude complémentaire

Une brève étude complémentaire est proposée ci-dessous. Les simulations de partitionnement réalisées dans les paragraphes précédents ont considéré les essais correspondant aux trois états structuraux du pont-rail. Étant donné que l'état « pendant resserrage » représente une condition structurale relativement floue, il se peut que cet état ne corresponde pas lui-même à une condition unique de l'ouvrage.

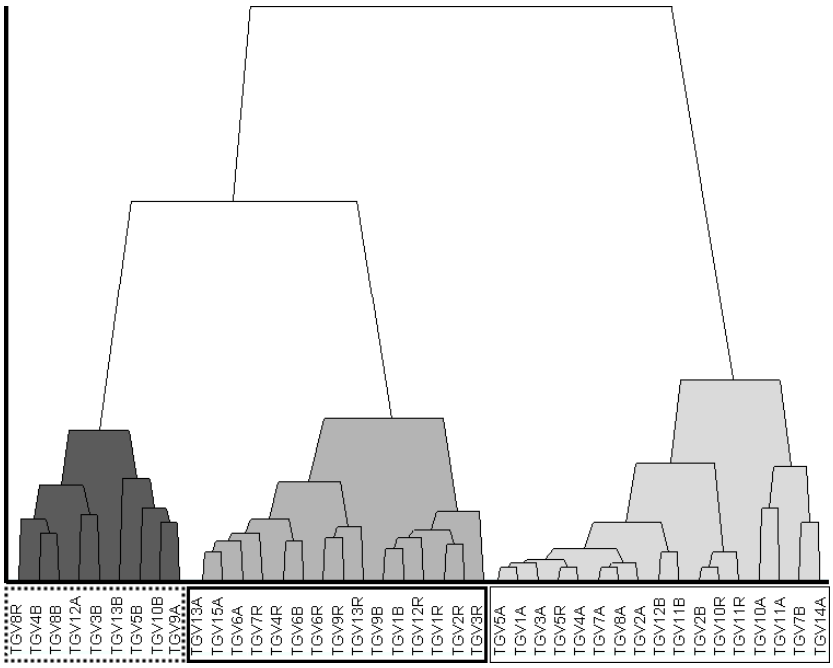


FIGURE 5.27 – Méthode des agglomérations hiérarchiques appliquée aux modes propres corrigées (histogrammes ; régression non linéaire).

Ainsi, il est également intéressant d'analyser la performance des méthodes de clustering maintenant appliquées aux 28 essais correspondant aux états « avant » et « après » les travaux. Bien que l'analyse classique présentée auparavant n'ait pas pu distinguer ces états nettement, il est espéré que ces méthodes puissent le faire de manière adéquate. Il est important de remarquer que les simulations de clustering considérant seulement ces deux états structuraux mènent à de nouveaux calculs de distances entre essais et, par conséquent, à de nouvelles inerties intra et inter clusters.

5.5.4.1 Signaux

La méthode des divisions hiérarchiques est appliquée dans ce paragraphe aux signaux transformés en histogrammes. Les Figures B.17 et 5.28 présentent les arbres de décision obtenus. Il est noté que dans les deux cas, les taux de classification sont significativement meilleurs. Si des histogrammes sont considérés, tous les essais « après » sont regroupés ensemble et 11 essais sur 15 (73%) « avant » sont classés correctement. Le capteur 2 (situé à 1/4 de portée) est choisi comme variable discriminante.

La méthode des agglomérations hiérarchiques est maintenant utilisée. Il est noté que le type de transformation choisi (intervalles ou histogrammes) n'a pas d'impact sur les taux de classification correcte : 60% (9/15) pour l'état « avant » et 54% (7/13) pour l'état « après » (Figure B.18 et 5.29). En effet, les hiérarchies sont très proches. Il existe quelques essais qui sont regroupés différemment pour chaque type de transformation symbolique, mais en général, il est observé une quantité élevée d'essais mélangés.

L'application de la méthode des nuées dynamiques mène à de meilleurs résultats. En effet, quelle que soit la représentation symbolique, tous les essais « avant » sont classés correctement. Pour l'état « après » resserrage, le taux de classification correcte est 62% (8/13).

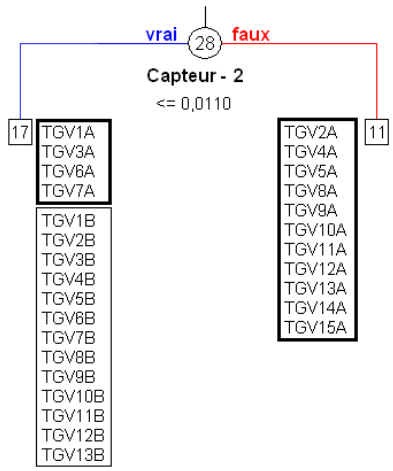


FIGURE 5.28 – Classification des signaux en utilisant les divisions hiérarchiques (histogrammes).

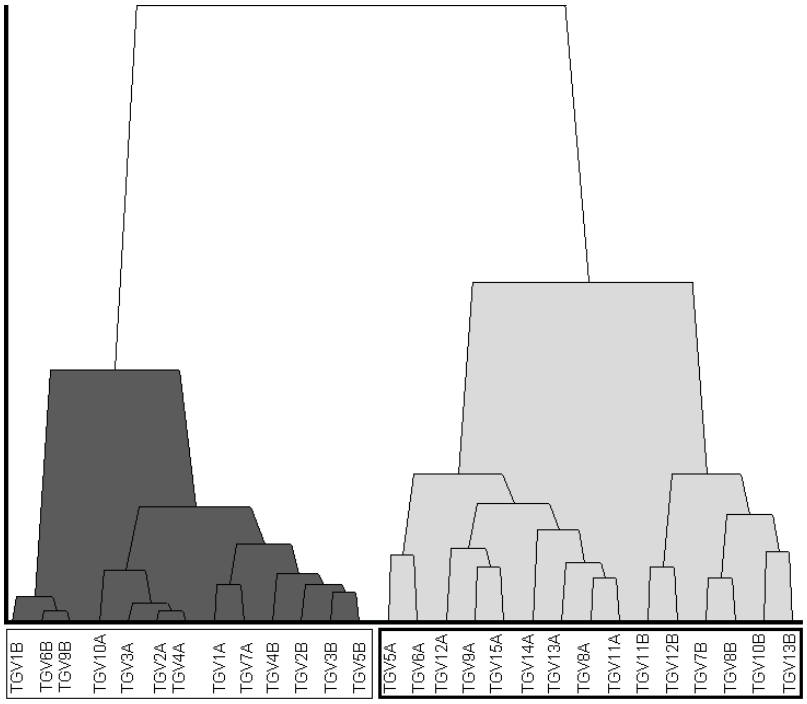


FIGURE 5.29 – Classification des signaux en utilisant les agglomérations hiérarchiques (histogrammes).

5.5.4.2 Fréquences propres

Dans ce paragraphe, les divisions hiérarchiques sont appliquées aux quatre premières fréquences propres transformées en histogrammes. Comme il a déjà été observé précédemment, les résultats obtenus sont meilleurs que ceux utilisant directement les signaux. Si la transformation en histogrammes est prise en compte, le partitionnement est optimal (100% pour les deux clusters) (cf. Figure 5.30). Il est à noter que la deuxième fréquence a été choisie comme la variable la plus discriminante. Ce résultat est intéressant car cette fréquence n'est pas identifiée comme la plus sensible dans les analyses classiques (à l'exception des tests d'hypothèse). Toutefois, il faut prendre en compte que la détermination des variables les plus discriminantes est strictement basée sur la minimisation d'inerties intra-clusters. La variable la plus discriminante n'est donc

pas forcément la même dans une analyse classique et dans une analyse en données symboliques.

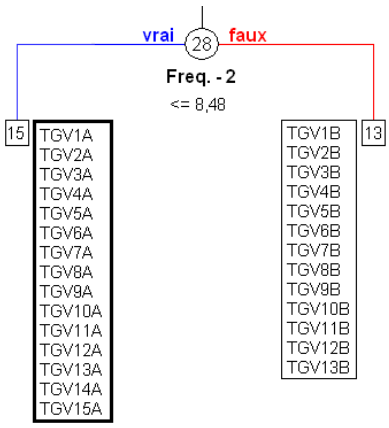


FIGURE 5.30 – Classification des fréquences propres en utilisant les divisions hiérarchiques (histogrammes).

Les résultats obtenus par la méthode des agglomérations hiérarchiques ne sont pas aussi pertinents. Bien que la plupart des essais « avant » soient classés dans un seul cluster, quelques essais « après » y sont mélangés. La Figure 5.31 illustre le partitionnement obtenu en utilisant des histogrammes. Dans ce cas, les taux de classification correcte sont de 73% (11/15) pour l'état « avant » et 92% (12/13) pour l'état « après ». Cela montre que cette méthode n'est généralement pas capable de bien discriminer ces états structuraux, même si des fréquences propres sont utilisées.

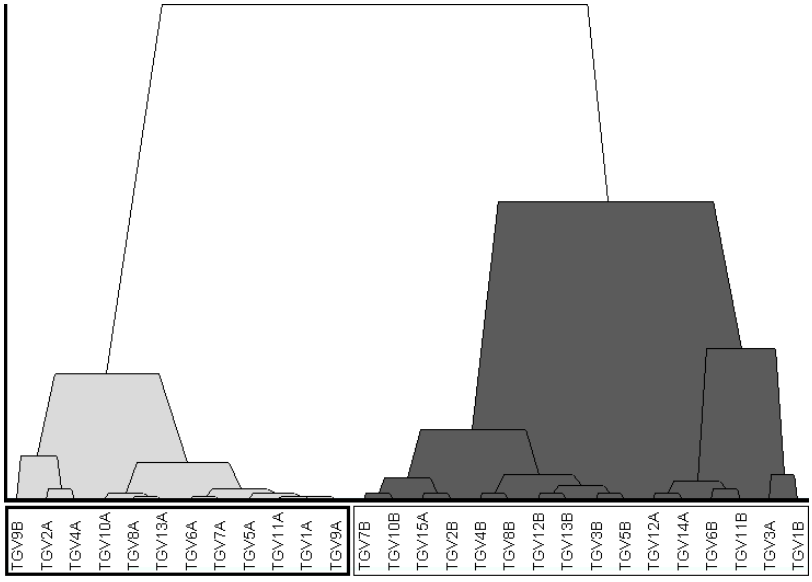


FIGURE 5.31 – Classification des fréquences propres en utilisant les agglomérations hiérarchiques (histogrammes).

En utilisant les nuées dynamiques, les résultats sont meilleurs que ceux obtenus par les agglomérations hiérarchiques. Le taux de classification correcte est de 100% pour le cluster « avant » et 92% (12/13) pour l'état « après ».

5.5.4.3 Modes propres

Les méthodes de clustering sont maintenant appliquées aux quatre premiers modes propres. En général, les taux de classification sont significativement meilleurs que ceux obtenus en utilisant les trois états structuraux en même temps. Cela montre que les calculs des distances entre les essais, des inerties des clusters et surtout des valeurs de coupure, influent directement sur les partitionnements obtenus. Les arbres de décision construits utilisant des intervalles et des histogrammes sont montrés dans les Figures B.21 et 5.32. Dans le cas des histogrammes, 100% des essais « avant » sont classés correctement et 85% (11/13) le sont pour les essais « après ».

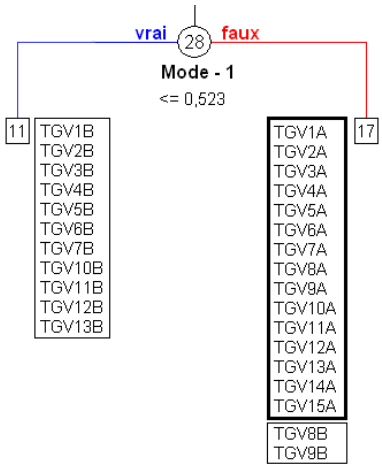


FIGURE 5.32 – Classification des modes propres en utilisant les divisions hiérarchiques (histogrammes).

En utilisant les agglomérations hiérarchiques, des taux de classification correcte plus faibles sont obtenus. Bien que les classifications soient meilleures que celles obtenues avec des signaux, un nombre significatif d'essais est systématiquement mélangé. La Figure 5.33 illustre la hiérarchie obtenue en utilisant des histogrammes. Les résultats sont relativement meilleurs que ceux obtenus en utilisant des intervalles (cf. Figure B.22) : 87% (13/15) pour les essais « avant » et 77% (10/13) pour les essais « après ».

Par ailleurs, la méthode des nuées dynamiques est considérée. Comme il a déjà été remarqué, cette méthode permet d'obtenir de meilleurs résultats que ceux obtenus par les divisions hiérarchiques et les agglomérations hiérarchiques. Dans cette analyse, 100% (15/15) des essais sont classés correctement dans le cluster « avant » et 85% (11/13) le sont dans le cluster « après ».

Tenant compte des simulations réalisées, il est possible de dire que les méthodes des divisions hiérarchiques et des nuées dynamiques sont les méthodes les plus adéquates pour la discrimination d'états structuraux. Même qu'elles soient capables de distinguer de manière satisfaisante les trois états originaux, il est intéressant d'observer que de meilleurs résultats sont obtenus en n'utilisant que les états « extrêmes » (avant et après le resserrage des bielles du pont).

5.5.4.4 Application aux paramètres modaux corrigés

Les analyses suivantes considèrent les méthodes de clustering appliquées aux paramètres modaux corrigés à partir des filtres thermiques proposés. Le but est de comparer les résultats obtenus avec ceux évalués sans prise en compte des variations de températures. Dans la suite, les résultats sont synthétisés sous la forme des tableaux.

Les premières simulations portent sur la méthode des divisions hiérarchiques appliquées aux

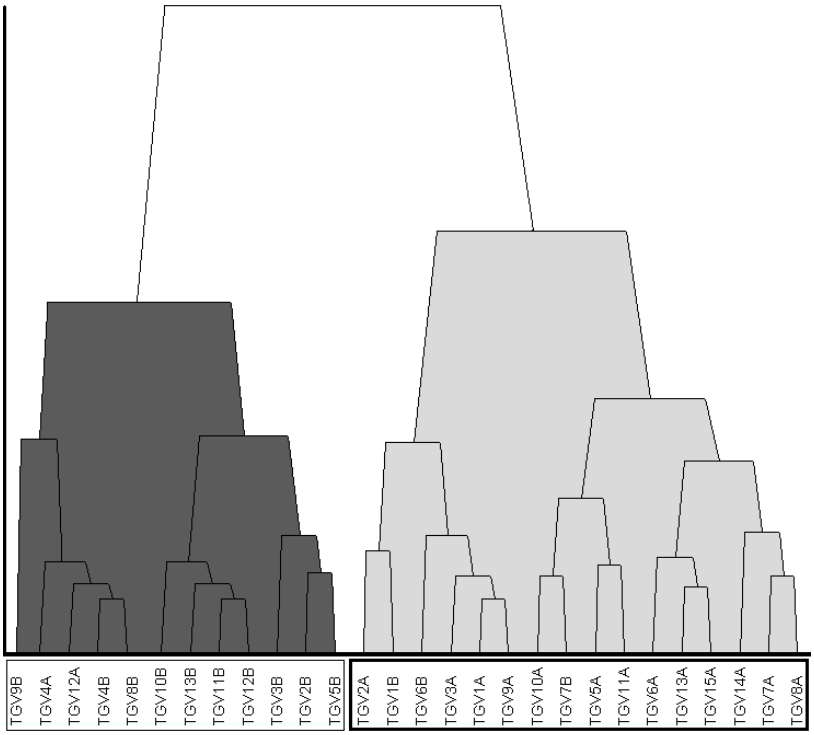


FIGURE 5.33 – Classification des modes propres en utilisant les agglomérations hiérarchiques (histogrammes).

paramètres modaux corrigés par le filtre linéaire (cf. Tableaux B.6 et 5.8). En général, le taux de classification correcte obtenu est pratiquement du même ordre que ceux évalués précédemment. Par ailleurs, il est noté une légère diminution de ces taux lorsque la transformation en histogrammes est considérée, que ce soit pour les fréquences propres ou les modes propres.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	93 (14/15)	93 (14/15)	100 (15/15)	100 (15/15)
Après	92 (12/13)	92 (12/13)	100 (15/13)	85 (13/13)

TABLEAU 5.8 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (régression linéaire ; valeurs en %).

Le Tableau 5.9 regroupe les classifications obtenues en utilisant la méthode des agglomérations hiérarchiques. Bien que les pourcentages soient significativement plus faibles que ceux de la méthode précédente, il est noté une sensible amélioration pour les fréquences propres. Toutefois, ceci n'est pas observé dans le cas des déformées modales, où les taux sont légèrement plus faibles. Ces résultats ambigus laissent penser que l'utilisation des filtres thermiques peut ne pas être pertinente dans cette étude.

Lorsque la méthode des nuées dynamiques est considérée, les résultats sont plus pertinents que ceux obtenus avec la méthode des divisions hiérarchiques. Cependant, les taux de classification correcte sont relativement plus faibles que ceux obtenus sans utiliser des filtres. Le Tableau 5.10 donne les pourcentages évalués en considérant les fréquences et les modes propres. Il est noté que la représentation par des histogrammes semble être plus adéquate par rapport à celle utilisant des intervalles (cf. Tableau 5.10).

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	80 (12/15)	80 (12/15)	73 (11/15)	87 (12/15)
Après	77 (10/13)	77 (10/13)	92 (12/13)	77 (10/13)

TABLEAU 5.9 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (régression linéaire ; valeurs en %).

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	93 (14/15)	87 (13/15)	100 (15/15)	100 (15/15)
Après	77 (10/13)	92 (12/13)	92 (12/13)	85 (11/13)

TABLEAU 5.10 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (régression linéaire ; valeurs en %).

Une deuxième technique, basée sur la régression non linéaire à partir des réseaux de neurones, est proposée ci-après. La méthode des divisions hiérarchiques est appliquée aux fréquences et modes propres. Le Tableau 5.11 regroupe les taux de classification obtenus en utilisant des histogrammes. Dans les deux cas, les résultats ne changent pas significativement. En considérant les paramètres modaux transformés en intervalles (cf. Tableau B.9), les résultats sont légèrement meilleurs. Toutefois, ceci n'est pas observé si la représentation par des histogrammes est considérée. Il est donc difficile de conclure si le filtre non linéaire est plus pertinent que celui linéaire.

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	100 (15/15)	93 (14/15)	100 (15/15)	100 (15/15)
Après	85 (11/13)	100 (13/13)	100 (15/13)	85 (13/13)

TABLEAU 5.11 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (régression non linéaire ; valeurs en %).

En considérant la méthode des agglomérations hiérarchiques, il est noté une légère amélioration des taux de classification correcte par rapport au filtre linéaire (cf. Tableau 5.12). Toutefois, les résultats sont pratiquement du même ordre que ceux évalués sans la prise en compte des effets thermiques.

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	73 (11/15)	80 (12/15)	73 (11/15)	87 (12/15)
Après	85 (10/13)	85 (11/13)	92 (12/13)	77 (10/13)

TABLEAU 5.12 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (régression non linéaire ; valeurs en %).

Enfin, la méthode des nuées dynamiques est appliquée aux paramètres modaux corrigés par le filtre non linéaire. Le Tableau 5.13 regroupe les pourcentages de classification correcte obtenus.

Il est noté que les résultats sont moins pertinents que ceux obtenus par le filtre linéaire (ou encore sans aucun filtre) dans le cas où les fréquences propres sont utilisées. Cependant, les taux sont relativement plus importants pour les déformées modales, notamment les taux du cluster « après ».

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	87 (13/15)	80 (12/15)	100 (15/15)	100 (15/15)
Après	77 (10/13)	100 (13/13)	92 (12/13)	85 (11/13)

TABLEAU 5.13 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (régression non linéaire ; valeurs en %).

5.5.5 Indices de partitionnement optimal

Comme il a été présenté dans le Chapitre 4 à l'aide des simulations numériques, les indices de partitionnement optimal introduits dans le paragraphe 3.1.5 peuvent être employés dans une démarche de discrimination des différents états structuraux. L'objectif est ici de vérifier l'aptitude de ces indices à indiquer le nombre d'états structuraux existants dans le cadre de l'application expérimentale sur le pont-rail PK 075+317. Comme expliqué au Chapitre 3, les indices de partitionnement optimal sont évalués uniquement dans les cas des méthodes des divisions hiérarchiques et des nuées dynamiques. En effet, ces deux techniques nécessitent la définition du nombre de clusters existants au préalable, ce qui n'est pas le cas des agglomérations hiérarchiques.

La démarche proposée est similaire à celle déjà présentée dans le Chapitre 4 et consiste à calculer les indices pour différents nombres de clusters. Le nombre maximal de clusters est fixé à 5 dans la suite. Le but est d'évaluer les indices tenant compte des partitionnements obtenus à partir de l'ensemble des 41 essais. En d'autres termes, l'idée est d'observer si les indices indiquent l'existence de trois comportements structuraux : avant, pendant et après resserrage.

5.5.5.1 Signaux

Les indices sont d'abord calculés en utilisant les deux méthodes de partitionnement appliquées aux signaux. L'analyse du Tableau 5.14 permet de vérifier que l'indice CH indique de manière consistante le nombre de clusters attendu (i.e., trois clusters dans l'exemple étudié). L'indice Γ signale toutefois 2 ou 4 clusters. L'indice C^* n'indique jamais le nombre d'états structuraux théoriquement existants. Il est intéressant d'observer que cet indice signale systématiquement un nombre de clusters toujours plus élevé que dans la réalité. Une sensibilité forte à des petites modifications pourrait justifier l'indication d'autres groupes de comportement. Les résultats obtenus par des intervalles sont relativement plus adéquats (cf. Tableau B.12).

5.5.5.2 Fréquences propres

L'analyse suivante consiste à évaluer les indices lorsque les fréquences propres sont considérées. Le Tableau 5.15 regroupe les valeurs calculées. Dans ce cas, il est noté qu'un bon accord est trouvé entre les indices CH et Γ, pour les deux méthodes de clustering. Si la transformation en histogrammes est considérée, l'indice C^* est différent de CH et Γ qui sont eux identiques. Il est noté que la variation de C^* est très faible lorsque la méthode des nuées dynamiques est utilisée. Il est donc impossible de conclure quand au partitionnement optimal dans ce dernier

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	23,7	**0,004**	0,95	17,9	0,135	0,66
4	25,9	0,006	**0,98**	22,7	**0,112**	0,65
3	**28,8**	0,016	0,96	**37,5**	0,12	0,69
2	26,6	0,031	0,97	27,9	0,14	**0,73**

TABLEAU 5.14 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en histogrammes (états « avant », « pendant » et « après »).

cas. Les trois indices indiquent le nombre théorique d'états structuraux lorsque des intervalles sont considérés, (cf. Tableau B.13).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	25,9	**0,025**	0,67	33,3	0,037	0,7
4	30,5	0,062	0,62	37,1	**0,035**	0,83
3	**37,1**	0,09	**0,84**	**43,2**	0,036	**0,90**
2	34,4	0,135	0,7	28,6	0,038	0,85

TABLEAU 5.15 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en histogrammes (états « avant », « pendant » et « après »).

Le Tableau 5.16 regroupe les indices calculés à partir des partitionnements obtenus en utilisant les fréquences propres corrigées par le filtre linéaire. Il est noté que les indices CH et Γ concordent dans la plupart des cas en indiquant trois clusters comme le partitionnement optimal (cette même valeur a été obtenue sans prise en compte des variations thermiques). L'indice C^* indique toujours un nombre de clusters plus élevé que ceux obtenus avec les autres indices.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	18,1	**0,222**	0,88	34,5	**0,029**	0,8
4	18,5	0,292	0,88	29,7	0,036	0,88
3	**22,2**	0,259	**0,89**	**39,7**	0,037	**0,88**
2	21,8	0,258	0,86	36,1	0,027	0,8

TABLEAU 5.16 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformées en histogrammes (régression linéaire).

La même procédure est considérée lorsque le filtre non linéaire est utilisé. En considérant les fréquences propres, il est noté que la détermination du nombre attendu de clusters est plus erratique que celle observée précédemment (cf. Tableau 5.17). Un partitionnement moins satisfaisant est donc obtenu en utilisant les fréquences propres corrigées. Dans ce cas, l'indice CH permet d'indiquer correctement le nombre de clusters dans la plupart de cas, suivi de l'indice Γ.

5.5.5.3 Modes propres

Par ailleurs, ces indices sont calculés pour chaque partitionnement obtenu en utilisant les modes propres. Dans ce cas, il est noté que les indices indiquent rarement le partitionnement optimal ayant trois clusters. Pour la méthode des nuées dynamiques, les indices CH et Γ indiquent toujours trois clusters. Pour les divisions hiérarchiques, toutefois, seul l'indice CH est associé à trois groupes (cf. Tableau5.18).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ
5	13,61	4,4	0,85	21,08	1,8	0,88
4	16,93	3,9	**0,88**	27,14	1,4	0,88
3	**17,35**	**3,5**	0,81	**32,44**	1,7	**0,92**
2	11,65	6,7	0,73	27,62	**1,1**	0,89

TABLEAU 5.17 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformées en histogrammes (régression non linéaire). † Les valeurs de l'indice C^* sont multipliées par $(\times 10^{-2})$.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	*CH*	C^*	Γ
5	14,3	0,123	0,53	26	0,053	0,73
4	34	**0,048**	0,63	28,3	**0,042**	0,84
3	**39,0**	0,157	0,74	**37,9**	0,072	**0,87**
2	23,3	0,461	**0,75**	32,2	0,103	0,77

TABLEAU 5.18 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformées en histogrammes (états « avant », « pendant » et « après »).

Si les modes propres sont corrigés par le filtre linéaire, il est noté que les indices oscillent significativement lorsque la méthode des divisions hiérarchiques est considérée. Dans l'exemple considéré, seul l'indice Γ signale le nombre attendu de groupes. Dans le cas des nuées dynamiques, toutefois, tous les indices indiquent trois clusters comme partitionnement optimal (cf. Tableau 5.19).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	*CH*	C^*	Γ
5	**29,9**	0,012	0,6	42,4	0,009	0,69
4	24,1	**0,011**	0,59	36,5	0,008	0,7
3	21,9	0,016	**0,62**	**43,3**	**0,005**	**0,72**
2	21,2	0,016	0,62	36	0,009	0,69

TABLEAU 5.19 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigés » transformés en histogrammes (régression linéaire).

De plus, les indices sont calculés en considérant les composantes modales corrigées par le filtre non linéaire. Les valeurs obtenues avec la méthode de nuées dynamiques sont généralement plus pertinentes que celles obtenus avec la méthode des divisions hiérarchiques. En effet, en considérant cette dernière méthode, seul l'indice Γ signale trois clusters comme le partitionnement optimal. Dans le cas des nuées dynamiques, les indices CH et Γ signalent toujours trois clusters (cf. Tableau 5.20).

5.5.5.4 Étude complémentaire

Dans les paragraphes précédents, une étude complémentaire a été menée en ne considérant que les états « avant » et « après » resserrage. Il a été observé que les clusters obtenus étaient plus homogènes et, par conséquent, les taux de classification correcte étaient plus importants. Les indices de partitionnement optimal sont calculés à partir des signaux et des paramètres modaux. L'objectif de ce paragraphe est de vérifier si les indices de partitionnement sont capables

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	23,8	0,128	0,73	17,3	0,086	0,87
4	**26,5**	**0,078**	0,74	15,1	**0,057**	0,84
3	23,5	0,134	**0,77**	**18,2**	0,073	**0,88**
2	25,5	0,092	0,74	17,1	0,101	0,88

TABLEAU 5.20 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigés » transformés en histogrammes (régression non linéaire).

d'indiquer correctement le nombre de groupes existants (à savoir deux groupes dans ce dernier cas).

Signaux

La première simulation considère les méthodes de clustering appliquées aux signaux. Les valeurs calculées pour chaque indice sont présentées dans le Tableau 5.21. L'indice CH indique deux groupes pour toutes les méthodes, ce qui est en accord avec le cas considéré. L'indice Γ ne signale toutefois pas deux clusters comme étant le partitionnement optimal lorsque la méthode des nuées dynamiques est considérée. Enfin, l'indice C^*, donne des résultats différents de ceux obtenus en considérant CH et Γ.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	18,4	**0,065**	0,55	13,3	0,2	0,67
4	24,8	0,096	0,67	14,5	0,14	0,53
3	25,1	0,23	0,72	15,9	**0,07**	**0,71**
2	**27,6**	0,294	**0,75**	**17,0**	0,1	0,68

TABLEAU 5.21 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en histogrammes (états « avant » et « après »).

Fréquences propres

Lorsque les fréquences propres sont considérées (cf. Tableau 5.22), il est montré que tous les indices indiquent deux clusters si la méthode des nuées dynamiques est utilisée. Pour les divisions hiérarchiques, seul l'indice C^* indique un nombre plus important de groupes (quatre clusters).

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	17,1	0,024	0,7	37,1	0,194	0,6
4	24,4	**0,012**	0,6	36,8	0,176	0,68
3	23,1	0,018	0,68	30,9	0,029	0,88
2	**26,8**	0,014	**0,83**	**37,5**	**0,021**	**0,93**

TABLEAU 5.22 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en histogrammes (états « avant », et « après »).

La procédure de correction des fréquences propres par filtre linéaire est maintenant considéré. Le Tableau 5.23 regroupe les valeurs calculées à partir des partitionnements obtenus en utilisant les fréquences propres. Il est noté que les indices CH et Γ indiquent toujours deux groupes comme étant le partitionnement optimal. Toutefois, l'indice C^* oscille entre 3 et 5 clusters. Ceci est également observé lorsque les modes propres sont utilisés (cf. Tableau 5.26). Par ailleurs, l'indice Γ signale trois clusters lorsque la méthode des nuées dynamiques est considérée.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	C^*	Γ	CH	C^*	Γ
5	25,5	**0,029**	0,78	21,92	0,079	0,73
4	27,1	0,148	0,88	23,09	0,044	0,76
3	34	0,281	0,93	25,38	**0,041**	0,8
2	**34,1**	0,394	**0,96**	**26,85**	0,061	**0,82**

TABLEAU 5.23 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformés en histogrammes (états « avant » et « après » ; régression linéaire).

Dans un deuxième temps, le filtre non linéaire est considéré. Les résultats sont pratiquement identiques. Il est noté que l'indice Γ (évalué pour le partitionnement obtenu par les nuées dynamiques) n'indique pas le nombre attendue de clusters (cf. Tableau 5.24). En considérant la méthode des divisions hiérarchiques, seuls les indices CH et Γ concordent en indiquant deux groupes.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	C^*	Γ	CH	C^*	Γ
5	25,5	0,031	0,57	37,2	**0,140**	0,55
4	26,4	**0,029**	0,55	41,7	0,145	0,56
3	32,4	0,05	0,64	49,6	0,163	**0,65**
2	**38,3**	0,077	**0,69**	**55,0**	0,175	0,6

TABLEAU 5.24 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformés en histogrammes (états « avant » et « après » ; régression non linéaire).

Modes propres

Finalement, lorsque les modes propres sont considérés (cf. Tableau 5.25), les valeurs obtenues pour la méthode de nuées dynamiques sont généralement plus adéquates que celles obtenues en utilisant la méthode des divisions hiérarchiques. Dans ce dernier cas, seul l'indice CH signale deux clusters comme partitionnement étant le partitionnement optimal. Pour les nuées dynamiques, tous les indices indiquent correctement le nombre optimal de clusters.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	C^*	Γ	CH	C^*	Γ
5	15,6	**0,056**	0,65	23,8	0,123	0,43
4	15,1	0,062	0,7	22,1	0,18	0,52
3	16,8	0,082	**0,78**	26,3	0,091	0,52
2	**19,7**	0,068	0,73	**30,7**	**0,082**	**0,61**

TABLEAU 5.25 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en histogrammes (états « avant », et « après »).

Comme dans le cas des fréquences propres, le filtre linéaire est considéré pour la correction des composantes modales. Il est noté que les indices CH et Γ indiquent toujours deux groupes comme étant le nombre optimal, dans le cas des divisions hiérarchiques. Toutefois, l'indice Γ signale trois clusters lorsque la méthode des nuées dynamiques est considérée. L'indice C^* oscille entre 3 et 4 clusters (cf. Tableau 5.26).

Le Tableau 5.27 regroupe les valeurs calculées en considérant le filtre non linéaire. Dans ce cas, il est noté un nombre important d'indications incorrectes. En considérant les divisions hiérarchiques, par exemple, l'indice CH n'indique jamais deux clusters comme étant le partitionnement

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	32,3	0,075	0,58	17,4	0,023	0,74
4	36,5	0,05	0,61	19,7	**0,023**	0,72
3	41,7	**0,047**	0,64	21,4	0,029	**0,76**
2	**45,9**	0,063	**0,68**	**22,9**	0,038	0,68

TABLEAU 5.26 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigés » transformés en histogrammes (états « avant » et « après » ; régression linéaire).

optimal. Par ailleurs, pour la méthode des nuées dynamiques, l'indice Γ signale trois clusters.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	44	0,069	0,66	30,59	0,057	0,77
4	46,4	**0,067**	0,65	33,78	**0,045**	0,78
3	**62,4**	0,1	0,74	38,57	0,053	**0,84**
2	61,5	0,131	**0,82**	**41,91**	0,069	0,84

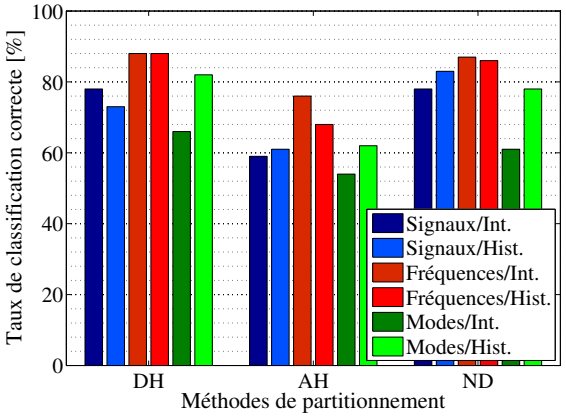
TABLEAU 5.27 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigés » transformés en histogrammes (états « avant » et « après » ; régression non linéaire).

5.5.6 Bilan

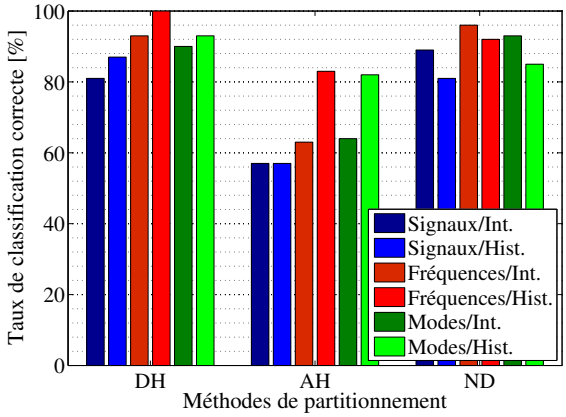
Compte tenu des résultats obtenus, il est possible de dire que l'approche basée sur le couplage de l'ADS aux méthodes de clustering (notamment lorsque les fréquences propres sont considérées), permet une meilleure discrimination des différents états structuraux. En effet, les méthodes des divisions hiérarchiques et des nuées dynamiques ont permis d'obtenir non seulement les meilleurs taux de classification correcte, mais aussi des clusters plus homogènes. La méthode des agglomérations hiérarchiques a cependant conduit à des résultats moins satisfaisants (taux de classification correcte de l'ordre de 60%), ayant des groupes généralement très hétérogènes. Cela montre une certaine inaptitude de cette méthode à bien discriminer différents états structuraux. Il est noté que pour les différentes méthodes considérées, la transformation en histogrammes mène généralement à de meilleurs résultats, surtout si les signaux et les modes propres sont considérés. Il faut néanmoins rester prudent et signaler que les taux de classification correcte ne dépassent généralement pas 90%. Ceci montre que les résultats doivent être analysés avec précaution et que des méthodes plus robustes (telles que des méthodes de classification supervisée) peuvent être considérées.

La Figure 5.34(a) montre une synthèse de la moyenne globale (pour tous les états structuraux) des taux de classification correcte pour chaque méthode et chaque type de transformation utilisée, en considérant l'ensemble d'essais des états avant, pendant et après resserrage des bielles. De plus, les résultats obtenus à partir de l'étude complémentaire, où seulement les essais « avant » et « après » ont été considérés, sont illustrés à la Figure 5.34(b). Dans cette analyse, la méthode des divisions hiérarchiques a montré être la plus pertinente pour la discrimination des états structuraux. Son application à des fréquences propres a par ailleurs permis de montrer l'efficacité de l'approche proposée. Cette méthode appliquée aux signaux produit également de bons résultats. La méthode des nuées dynamiques permet également d'obtenir des bonnes classifications, surtout pour l'application aux modes propres. De façon générale, la méthode des agglomérations hiérarchiques a montré être la moins adaptée, même lorsque les fréquences propres sont utilisées. Il est également remarqué que la prise en compte des essais correspondant aux états « avant »

et « après » mène à de meilleurs résultats, notamment dans le cas où les divisions hiérarchiques sont appliquées aux fréquences propres transformées en histogrammes. Dans ce cas particulier, tous les essais sont classés correctement.



(a) Taux de classification pour les états « avant », « pendant » et « après »



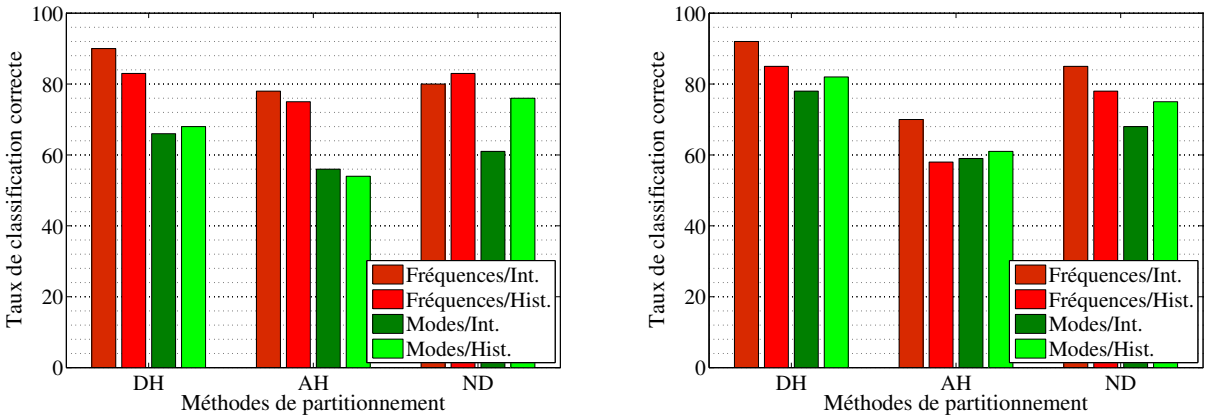
(b) Taux de classification pour les états « avant » et « après »

FIGURE 5.34 – Synthèse des résultats obtenus en utilisant les méthodes de clustering.

L'application des indices de partitionnement optimal aux méthodes de clustering a été également étudiée. Ces indices permettent, dans un certain sens, de valider l'étude de discrimination d'états structuraux. Le but a été d'utiliser les indices appliqués à l'ensemble des 41 essais afin de vérifier le partitionnement en trois clusters distincts. Il a été observé que deux de ces indices - CH et Γ - ont invariablement indiqué le nombre attendu, contrairement à l'indice C^*, qui a souvent oscillé entre le nombre attendu et un nombre plus important de clusters. De plus, l'utilisation des fréquences propres a encore une fois menée aux résultats les plus satisfaisants, en signalant le partitionnement théorique comme optimal dans la plupart des cas. Finalement, les indices ont été calculés pour les simulations de partitionnement en n'utilisant que les états extrêmes du pont-rail : « avant » et « après » le renforcement. Les résultats obtenus ont montré une meilleure adéquation des indices qui ont rarement indiqué un nombre incorrect de groupes. En général, l'indice CH est celui qui a été le plus souvent en accord avec les analyses menées, suivi de l'indice Γ. L'indice C^* s'est toutefois montré encore très sensible dans ce dernier cas, en indiquant plusieurs clusters, ce qui peut être du à la présence d'effets extérieurs.

Par ailleurs, deux techniques de régression ont été proposées afin de corriger les effets de la variation de température sur les paramètres modaux identifiés. Pour cela, deux études ont été menées. Dans la première, l'ensemble de 41 essais des états avant, pendant et après resserrage des bielles ont été considérés. Dans la deuxième étude, seuls les essais enregistrés durant les phases avant et après travaux ont été pris en compte. L'application de lois de correction linéaire et non linéaire n'a pas montré être pertinente pour les cas considérés. En effet, les taux de classification correcte n'ont pratiquement pas changé après la correction des paramètres modaux. Certes, les classifications ont été améliorées dans certains cas. Le nombre d'essais classés incorrectement a cependant été plus important dans d'autres cas. Deux explications sont envisageables. D'une part, il a été observé que la variation des températures lors des trois campagnes d'essais n'était pas très importante. En effet, les séries de mesures ont été réalisées dans des journées successives et surtout dans des plages horaires similaires. Il se peut donc qu'il n'existe aucun effet thermique à corriger. D'autre part, en utilisant les techniques de régression linéaire et non linéaire, les erreurs inhérentes à ces méthodes ont peut être causé des perturbations sur les paramètres modaux et, par conséquent, sur le processus de partitionnement.

Les Figures 5.35(a) et 5.35(b) synthétisent les résultats obtenus pour les deux types de filtres thermiques utilisés, en considérant l'ensemble de 41 essais. Il est noté que les taux de classification correcte sont relativement proches pour les deux approches. Ces taux sont également proches de ceux obtenus sans prise en compte des effets de la variation de températures (cf. Figure 5.34(a)).

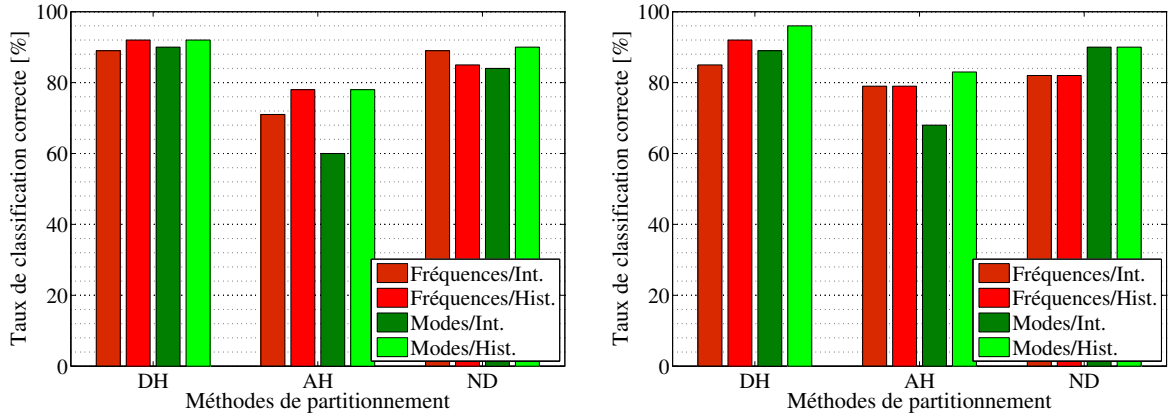


(a) Taux de classification obtenus en utilisant le filtre linéaire

(b) Taux de classification obtenus en utilisant le filtre non linéaire

FIGURE 5.35 – Synthèse des résultats obtenus en utilisant les méthodes de partitionnement de données appliquées aux paramètres modaux « corrigés » (états avant, pendant et après).

Par ailleurs, les Figures 5.36(a) et 5.36(b) synthétisent les pourcentages évalués dans les simulations réalisées en tenant compte des états avant et après resserrage, seulement. De façon similaire à l'analyse précédente, les résultats sont particulièrement ambigus, étant donné qu'ils sont améliorés dans certains cas et dégradés dans d'autres cas.



(a) Taux de classification obtenus en utilisant le filtre linéaire

(b) Taux de classification obtenus en utilisant le filtre non linéaire

FIGURE 5.36 – Synthèse des résultats obtenus en utilisant les méthodes de partitionnement de données appliquées aux paramètres modaux « corrigés » (états avant et après).

5.6 ADS couplée aux méthodes de classification supervisée

Dans ce paragraphe, les résultats obtenus à partir du couplage entre l'ADS et les méthodes de classification supervisée, appliquées aux signaux et paramètres modaux, sont présentés. Les simulations réalisées dans ce paragraphe utilisent la démarche présentée au Chapitre 4 à l'aide des applications numériques, en créant des groupes d'entraînement, de validation et de test à partir de l'ensemble des essais. Afin de vérifier l'adéquation et la robustesse de l'approche proposée, trois études différentes sont réalisées. Dans un premier temps, le groupe d'entraînement comporte 28 essais (15 essais correspondant à l'état « avant » et 13 essais à l'état « après »). Dans ce cas, l'état « avant » représente la classe 1 et l'état « après », la classe 2. Le groupe de test comporte les 13 essais d'état « pendant » les travaux de renforcement. Le but de cette simulation est de classer ces différents essais dans l'un des groupes définis précédemment. Étant donné que le resserrage des bielles a été une procédure progressive, il est intéressant d'observer comment ces méthodes classent les essais « pendant » dans les classes « avant » et « après ». Dans un deuxième temps, une analyse statistique est conduite, en utilisant l'ensemble des 41 essais et en simulant différents groupes d'entraînement, de validation et de test. Dans ce cas, trois classes sont créées, chacune représentant un état structural. Pour cela, 10000 simulations sont réalisées en considérant trois configurations :

- 30% (12) d'essais pour l'entraînement, 10% (4) pour la validation et 60% (25) pour le test,
- 40% (16) d'essais pour l'entraînement, 10% (4) pour la validation et 50% (21) pour le test,
- 50% (21) d'essais pour l'entraînement, 10% (4) pour la validation et 40% (16) pour le test.

Finalement, en n'utilisant que les 28 essais des états « avant » et « après », une nouvelle analyse statistique est réalisée. Les proportions pour chaque groupe sont alors les suivantes :

- 30% (8) d'essais pour l'entraînement, 10% (3) pour la validation et 60% (17) pour le test,
- 40% (11) d'essais pour l'entraînement, 10% (3) pour la validation et 50% (14) pour le test,
- 50% (14) d'essais pour l'entraînement, 10% (3) pour la validation et 40% (11) pour le test.

Paramètres utilisés

Les simulations utilisant les ADB (Arbres de Décision Bayésiens) sont réalisées sous MatLab® à l'aide de routines développées par l'auteur. Pour la méthode basée sur les RN (réseaux de neurones), l'architecture du réseau utilisé est composée d'une couche cachée avec 10 neurones (dans la première étude) et 20 neurones (dans les deuxième et troisième études). Par ailleurs, les fonctions d'activation utilisées sont la fonction tangente hyperbolique (cf. Équation (3.42)) dans la couche cachée et la fonction Heaviside (cf. Équation (3.40)) dans la couche de sortie. Le taux d'apprentissage η est fixé à 0,05 et la constante de moment μ est fixée à 0,001. Ces paramètres ont été choisis après quelques simulations réalisées au préalable et ont montré une bonne adéquation aux problèmes traités (Cury & Crémona 2010*c*,*b*). Les simulations ont été réalisées à l'aide de la boîte à outils Netlab (Bishop 1995). En ce qui concerne les SVM (Machines à Vecteurs Supports), la fonction noyau RBF (cf. paragraphe 3.2.3.2) est utilisée et le paramètre σ est déterminé de manière itérative lors de la phase de validation. Les simulations ont été réalisées à l'aide de la boîte à outils LibSVM (Chang & Lin 2001).

5.6.1 Première étude

Le Tableau 5.28 rappelle l'ensemble des essais dynamiques enregistrés lors de campagnes de mesures sur le pont-rail PK 075+317 et détaille les 4 phases des travaux de resserrage réalisés.

Bien que l'état « pendant » de l'ouvrage soit considéré comme une condition structurale différente et, d'une certaine manière, indépendante des états « avant » et « après », il est intéressant d'appliquer les méthodes de classification supervisée à ces essais et d'observer comment

État	Phase	Essais
Avant	-	TGV1A, TGV2A, ..., TGV15A
Pendant	1	TGV1R, TGV2R, TGV3R
Pendant	2	TGV4R, TGV5R, TGV6R
Pendant	3	TGV7R, TGV8R, TGV9R
Pendant	4	TGV10R, TGV11R, TGV12R, TGV13R
Après	-	TGV1B, TGV2B, ..., TGV13B

TABLEAU 5.28 – Description des essais réalisés sur le pont-rail PK 075+317.

ils se « rapprochent » de l'un de ces états. Par ailleurs, l'hypothèse suivante est considérée dans la suite : étant donné que le resserrage des bielles ait été une procédure continue, il peut être considéré que le comportement de l'ouvrage pendant les phases 1 et 2 des travaux est plus « similaire » à son comportement avant les travaux. De même, les phases 3 et 4 peuvent être considérées comme plus « similaire » à l'état « après ». En ce sens, si les essais TGV1R,..., TGV6R sont placés dans la classe 1, la classification est considérée correcte. La même approche est considérée pour la classification des essais TGV7R,..., TGV13R dans la classe 2.

Le Tableau 5.29 regroupe les résultats obtenus par la méthode des ADB appliquée aux signaux et paramètres modaux, transformés en histogrammes. Dans ce cas, et en considérant les critères précédents, les taux de classification pour toutes les analyses sont égaux à 70% (9/13). Il est intéressant de remarquer que l'essai TGV6R est systématiquement classé incorrectement. Ceci peut être expliqué par le fait que cet essai se trouve dans la frontière de transition entre les phases 2 et 3, ce qui justifierait son caractère ambigu.

Essai	Classement espéré	Classement obtenu		
		Signaux	Fréquences	Modes
TGV1R	1	1	1	1
TGV2R	1	1	1	1
TGV3R	1	1	2	2
TGV4R	1	1	1	2
TGV5R	1	1	2	1
TGV6R	1	2	2	2
TGV7R	2	1	2	2
TGV8R	2	1	2	2
TGV9R	2	1	1	1
TGV10R	2	2	2	2
TGV11R	2	2	2	2
TGV12R	2	2	2	2
TGV13R	2	2	2	2

TABLEAU 5.29 – Classification d'essais obtenue par les ADB (histogrammes).

En considérant les procédures de correction thermique, il est noté des modifications sur les résultats. Dans le cas des fréquences propres, par exemple, l'essai TGV5R est maintenant classé au groupe 1, différemment des résultats obtenus sans prise en compte des effets thermiques. De façon similaire, l'essais TGV9R qui était incorrectement classé dans le groupe 1, avant les corrections thermiques, est maintenant affecté à la classe 2. Toutefois, l'essai TGV6R est toujours classé au groupe 2 (cf. Tableau 5.30).

Les classifications obtenues en utilisant les RN sont détaillées dans le Tableau 5.31. En consi-

Essai	Classement espéré	Classement obtenu			
		Filtre linéaire		Filtre non linéaire	
		Fréquences	Modes	Fréquences	Modes
TGV1R	1	1	1	1	1
TGV2R	1	1	1	1	1
TGV3R	1	1	2	1	1
TGV4R	1	1	1	1	2
TGV5R	1	2	2	1	2
TGV6R	1	2	2	2	2
TGV7R	2	2	2	2	1
TGV8R	2	2	2	2	2
TGV9R	2	2	2	2	2
TGV10R	2	2	2	2	2
TGV11R	2	2	2	2	2
TGV12R	2	2	2	2	2
TGV13R	2	2	2	2	2

TABLEAU 5.30 – Classification d'essais obtenue par les ADB (correction thermique ; histogrammes).

dérant les signaux, les taux de classification correcte sont plus faibles que ceux obtenus par les ADB : 62% (8/13). Les taux sont significativement meilleurs en considérant les paramètres modaux. Dans ce cas, 12 essais sur 13 (92%) sont correctement classés, en considérant les fréquences propres. En utilisant les modes propres, le taux est de 85%, soit 11 essais sur 13. La remarque vis-à-vis de l'essai TGV6R reste valable en utilisant cette méthode.

Essai	Classement espéré	Classement obtenu		
		Signaux	Fréquences	Modes
TGV1R	1	1	1	1
TGV2R	1	1	1	1
TGV3R	1	2	1	1
TGV4R	1	2	1	1
TGV5R	1	2	1	1
TGV6R	1	2	2	2
TGV7R	2	2	2	1
TGV8R	2	2	2	2
TGV9R	2	1	2	2
TGV10R	2	2	2	2
TGV11R	2	2	2	2
TGV12R	2	2	2	2
TGV13R	2	2	2	2

TABLEAU 5.31 – Classification d'essais obtenue par les RN (histogrammes).

Des résultats similaires sont obtenus en considérant les procédures de correction thermique. En réalité, rien ne change dans le cas des fréquences propres. Toutefois, si des modes propres sont analysés, il est noté que l'utilisation des filtres à dégradé les classifications. Les essais TGV4R et TGV10R qui était classés correctement sans prise en compte des effets thermiques, ne le sont plus dans les analyses courantes. Cela laisse penser que l'utilisation des filtres thermiques peut ne pas être pertinente dans cette étude (cf. Tableau 5.32).

Essai	Classement espéré	Classement obtenu			
		Filtre linéaire		Filtre non linéaire	
		Fréquences	Modes	Fréquences	Modes
TGV1R	1	1	1	1	1
TGV2R	1	1	1	1	1
TGV3R	1	1	1	1	1
TGV4R	1	1	2	1	1
TGV5R	1	1	1	1	1
TGV6R	1	2	2	2	2
TGV7R	2	2	2	2	2
TGV8R	2	2	2	2	2
TGV9R	2	2	2	2	2
TGV10R	2	2	1	2	1
TGV11R	2	2	2	2	2
TGV12R	2	2	2	2	2
TGV13R	2	2	2	2	2

TABLEAU 5.32 – Classification d'essais obtenue par les RN (correction thermique ; histogrammes).

Finalement, le Tableau 5.33 regroupe les résultats obtenus par les SVM. La classification obtenue en utilisant les signaux transformés en histogrammes est la meilleure parmi les trois méthodes : 85% (11/13). Les taux de classification en considérant les paramètres modaux ont diminué par rapport à ceux obtenus en utilisant les RN. Pour les fréquences propres, 10 essais sur 13 (77%) sont correctement classés. Le taux obtenus pour les modes propres est égal à 70% (9/13).

Essai	Classement espéré	Classement obtenu		
		Signaux	Fréquences	Modes
TGV1R	1	1	1	1
TGV2R	1	1	1	1
TGV3R	1	1	1	1
TGV4R	1	1	1	1
TGV5R	1	1	1	2
TGV6R	1	2	2	2
TGV7R	2	2	1	1
TGV8R	2	2	1	1
TGV9R	2	1	2	2
TGV10R	2	2	2	2
TGV11R	2	2	2	2
TGV12R	2	2	2	2
TGV13R	2	2	2	2

TABLEAU 5.33 – Classification d'essais obtenue par les SVM (histogrammes).

L'utilisation des filtres thermiques couplés à la méthode des SVM améliore légèrement les taux de classification correcte. Il est à noté que le filtre linéaire semble être plus adéquat dans cette analyse. Dans le cas des fréquences propres, par exemple, seuls les TGV6R et TGV7R sont classés incorrectement. D'ailleurs, le même est observé pour les modes propres. L'utilisation d'un filtre non linéaire mène à des résultats relativement moins pertinents, notamment dans le cas des déformées modales (cf. Tableau 5.34).

Essai	Classement espéré	Classement obtenu			
		Filtre linéaire		Filtre non linéaire	
		Fréquences	Modes	Fréquences	Modes
TGV1R	1	1	1	1	1
TGV2R	1	1	1	1	1
TGV3R	1	1	1	1	1
TGV4R	1	1	1	1	1
TGV5R	1	1	1	1	1
TGV6R	1	2	2	2	2
TGV7R	2	1	1	2	1
TGV8R	2	2	2	2	2
TGV9R	2	2	2	2	2
TGV10R	2	2	2	1	1
TGV11R	2	2	2	2	2
TGV12R	2	2	2	2	2
TGV13R	2	2	2	2	2

TABLEAU 5.34 – Classification d'essais obtenue par les SVM (correction thermique ; histogrammes).

5.6.2 Deuxième étude

La deuxième étude a pour objectif de vérifier la robustesse et efficacité des méthodes de classification supervisée pour la discrimination des trois états structuraux. Pour cela, une étude statistique est menée en faisant varier la taille des groupes d'entraînement et de test. Au total, 10000 simulations sont réalisées pour chaque configuration et les meilleurs taux de classification correcte sont évalués ainsi que les taux moyens et minimaux (cf. paragraphe 5.5).

5.6.2.1 Signaux

Les premières simulations correspondent à l'application des méthodes de classification aux signaux. Dans le premier cas, où seulement 30% des essais sont utilisés pour l'entraînement, les trois méthodes donnent des résultats similaires. En effet, seuls les RN et SVM réussissent à classer parfaitement tous les essais. Par ailleurs, il apparaît clairement que les taux de classification sont plus importants pour toutes les méthodes à mesure que le nombre d'essais dans le groupe d'entraînement augmente. Il est également remarqué dans cette analyse que les RN permettent d'obtenir des pourcentages moyens plus importants que ceux obtenus avec les autres méthodes. Le Tableau 5.35 regroupe les taux de classification obtenus en utilisant la transformation en histogrammes. Mis à part le cas des RN, les résultats sont relativement moins satisfaisants que ceux obtenus avec les autres méthodes. Bien que les SVM permettent de classer 100% des essais correctement dans une (ou plus) des simulations réalisées, le taux moyen est faible pour la première configuration. Les ADB mènent à des taux de classification correcte relativement faibles pour tous les configurations (de l'ordre de 65%, en moyenne).

5.6.2.2 Fréquences propres

En appliquant les méthodes de classification supervisée aux fréquences propres, des résultats plus pertinents sont obtenus (comme dans le cas des méthodes de clustering). Les pourcentages de classification correcte obtenus lorsque la transformation en histogrammes sont montrés dans le Tableau5.36, respectivement. Bien que les ADB n'arrivent pas à classer 100% des essais, les

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	90	100	100	94	100	100	94	100	100
Moyen	64	83	67	65	89	80	64	93	85
Minimal	21	17	24	30	26	25	38	40	43

TABLEAU 5.35 – Taux de classification correcte utilisant les signaux transformés en histogrammes pour les états « avant », « pendant » et « après » (valeurs en %).

RN et les SVM le font pour toutes les simulations en considérant des intervalles interquartiles (cf. TableauB.33). Cela met en évidence, encore une fois, que les fréquences propres, transformées en données symboliques, peuvent constituer un indicateur efficace pour la détection d'endommagements. Il est intéressant de noter que les performances des RN et SVM sont « complémentaires » par rapport au type de transformation symbolique utilisée. En effet, les RN sont les plus performants lorsque des histogrammes sont utilisés. Au contraire, ce sont les SVM les plus performant lorsque des intervalles sont considérés. Cette observation renforce l'importance de toujours analyser l'ensemble des résultats et de tenir compte des spécificités de chaque méthode et du type de représentation des données utilisée.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	80	100	88	83	100	95	89	100	97
Moyen	57	90	53	66	93	56	75	95	54
Minimal	12	20	16	25	33	15	31	50	17

TABLEAU 5.36 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour les états « avant », « pendant » et « après » (valeurs en %).

Les simulations suivantes correspondent à l'application des méthodes de classification aux fréquences propres corrigées par le filtre linéaire. Le Tableau 5.37 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en histogrammes est considérée. Bien que les ADB n'arrivent pas à classer 100% des essais, les RN et les SVM le font pour quelques simulations. En général, les méthodes de classification supervisée permettent d'atteindre des résultats aussi satisfaisants que ceux obtenus sans le filtre thermique (cf. Tableau 5.36). Ceci renvoie à la remarque déjà mentionnée de la faible variation des températures entre les campagnes d'essais. Ainsi, la correction n'améliore pas les résultats des classifications.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	77	92	86	77	96	89	83	100	100
Moyen	60	87	61	60	86	59	63	90	57
Minimal	25	22	9	28	30	12	31	25	23

TABLEAU 5.37 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour les états « avant », « pendant » et « après » (régression linéaire ; valeurs en %).

L'utilisation du filtre non linéaire est maintenant considérée. Les résultats obtenus en considérant la transformation des fréquences propres en histogrammes est détaillée dans le Tableau 5.38. Il est noté que les classifications obtenues sont relativement similaires à celles trouvées par des intervalles (cf. Tableau B.35), avec une légère amélioration pour les SVM. Dans le sens inverse, il est possible que les corrections réalisées sur les fréquences aient produit une meilleure dis-

crimination des classes pour cette dernière méthode. En d'autres termes, la détermination des hyperplans de séparation a été, parfois, plus aisée en considérant cette analyse.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	75	92	89	77	92	93	78	91	91
Moyen	59	88	61	62	90	66	66	93	66
Minimal	20	25	19	25	22	20	29	44	16

TABLEAU 5.38 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour les états « avant », « pendant » et « après » (régression non linéaire ; valeurs en %).

5.6.2.3 Modes propres

La démarche proposée est maintenant appliquée aux modes propres. Les résultats montrent de faibles taux de classification correcte pour toutes les méthodes. Cela confirme la faible sensibilité des modes propres aux travaux de resserrage des bielles réalisés. Étant donné que la variation des modes propres est relativement faible d'un état à un autre, il est possible que l'approche proposée ne soit pas capable de discriminer les modifications structurales. Dans le cas de la transformation des modes propres en histogrammes (cf. Tableau 5.39), les ADB ont, en moyenne, des taux moins importants que ceux obtenus dans le cas des intervalles (cf. Tableau B.36). Le même résultat est observé pour les RN et SVM. Enfin, même si des valeurs plus importantes pour les taux maximaux sont atteints, ces méthodes sont, en général, moins performantes en utilisant des histogrammes.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	77	85	90	81	81	85	80	85	90
Moyen	33	38	34	31	40	50	35	41	55
Minimal	15	18	23	22	20	25	15	23	30

TABLEAU 5.39 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les états « avant », « pendant » et « après » (valeurs en %).

L'utilisation des modes propres corrigés par le filtre linéaire ne permet pas d'obtenir des taux de classification correcte aussi bons que ceux trouvés avec les fréquences propres. De plus, quel que soit le type de transformation en données symboliques utilisé, les résultats obtenus par les ADB et les RN évoluent par rapport à ceux obtenus sans prise en compte des effets thermiques (cf. Tableau5.39). Par ailleurs, les taux de classification correcte obtenus par les SVM sont relativement plus faibles (cf. Tableau 5.40).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	69	78	88	73	73	86	76	77	88
Moyen	35	39	40	35	38	40	37	43	45
Minimal	26	4	11	22	19	22	18	25	33

TABLEAU 5.40 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les états « avant », « pendant » et « après » (régression linéaire ; valeurs en %).

La considération des déformées modales corrigées par le filtre non linéaire mène à des résultats antagoniques. Bien que l'utilisation des intervalles interquartiles (cf. Tableau B.38) mène à des taux de classification relativement plus faibles, cela n'est pas le cas pour des histogrammes. Dans ce dernier cas, les classifications sont, en moyenne, meilleures que celles trouvées sans l'utilisation des filtres thermiques (cf. Tableau 5.41).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	75	79	83	77	78	83	80	81	85
Moyen	42	51	48	44	50	53	48	53	57
Minimal	22	18	17	24	20	22	24	29	22

TABLEAU 5.41 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les états « avant », « pendant » et « après » (régression non linéaire ; valeurs en %).

5.6.3 Troisième étude

L'objectif de la troisième étude se limite aux essais des états « avant » et « après ». En effet, ce raisonnement a été déjà considéré pour les méthodes de clustering au paragraphe 5.5.4 et des résultats relativement meilleurs ont été obtenus. Étant donné que ces états sont plus « éloignés », il est espéré que les méthodes de classification soient capables de mieux les distinguer.

5.6.3.1 Signaux

Les premières simulations considèrent l'utilisation des signaux. Les résultats obtenus en utilisant les signaux transformés en histogrammes sont détaillés dans le Tableau 5.42. Dans ce cas, pour toutes les méthodes de classification, les taux de classification sont significativement plus élevés. En général, les RN et SVM arrivent toujours à classer 100% des essais correctement dans une (ou plus) des simulations réalisées. Les meilleurs taux sont obtenus par les RN, suivis des SVM. Ces méthodes ont déjà montré être plus performantes que les ADB pour la classification d'essais. Par ailleurs, il est également noté que les pourcentages de classification correcte augmentent avec le nombre d'essais dans le groupe d'entraînement.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	90	100	100	91	98	100	93	100	100
Moyen	79	88	81	83	92	86	80	92	89
Minimal	41	49	24	45	25	29	33	37	34

TABLEAU 5.42 – Taux de classification correcte utilisant les signaux transformés en histogrammes pour les états « avant » et « après » (valeurs en %).

5.6.3.2 Fréquences propres

Par ailleurs, les méthodes de classification sont appliquées aux fréquences propres. Le Tableau 5.43 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en histogrammes est considérée. Dans ce cas, les trois méthodes de classification supervisée permettent d'atteindre des résultats pertinents, ayant des taux moyens de classification correcte de

l'ordre de 90%. De plus, toutes les méthodes ont réussi à classer, au moins une fois, tous les essais correctement, quel que soit le nombre d'essais utilisés dans la phase l'entraînement.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	100	100	100	100	100	100	100	100	100
Moyen	91	95	92	91	95	92	93	96	93
Minimal	45	33	8	47	28	15	44	44	21

TABLEAU 5.43 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour les états « avant » et « après » (valeurs en %).

Le Tableau 5.44 regroupe les taux de classification correcte en considérant le filtre linéaire. En général, les méthodes de classification supervisée permettent d'atteindre des résultats moins adéquats par rapport à ceux obtenus sans le filtre thermique (cf. Tableau 5.43).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	95	95	95	95	95	95	95	95	95
Moyen	90	94	91	90	94	91	92	95	92
Minimal	43	32	8	45	27	14	42	42	20

TABLEAU 5.44 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour les états « avant » et « après » (régression linéaire ; valeurs en %).

Les simulations sont également menées en considérant le filtre non linéaire (cf. Tableau 5.45). Dans ce cas, les taux de classification correcte sont légèrement plus importants que ceux du filtre linéaire. Toutefois, ces résultats ne montrent pas une amélioration nette comparés à ceux sans correction thermique.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	94	94	94	94	94	94	94	94	94
Moyen	89	92	90	89	92	90	90	93	90
Minimal	44	32	8	45	27	15	43	43	20

TABLEAU 5.45 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour les états « avant » et « après » (régression non linéaire ; valeurs en %).

5.6.3.3 Modes propres

Les modes propres sont maintenant considérés. Comme remarqué précédemment, les taux de classification correcte sont très faibles. La transformation en histogrammes se montre légèrement plus satisfaisante pour les RN. Dans ce cas, les taux sont relativement plus importants, contrairement à ceux trouvés en utilisant les SVM (cf. Tableau B.43).

Le Tableau 5.47 regroupe les taux de classification correcte en considérant le filtre linéaire. Comme il a été remarqué dans le cas des fréquences propres, les méthodes de classification supervisée permettent d'atteindre des résultats moins adéquats par rapport à ceux obtenus sans le filtre thermique (cf. Tableau 5.46).

Les simulations sont également menées en considérant le filtre non linéaire (cf. Tableau 5.48). Dans ce cas, bien que les taux moyens de classification correcte soient sont légèrement plus

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	73	79	77	77	76	78	77	79	81
Moyen	55	53	52	55	52	52	56	58	56
Minimal	28	12	16	36	25	7	38	34	30

TABLEAU 5.46 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les états « avant » et « après » (valeurs en %).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	70	76	74	74	73	75	74	76	78
Moyen	54	52	51	54	51	51	55	57	55
Minimal	27	12	16	35	25	7	37	33	29

TABLEAU 5.47 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les états « avant » et « après » (régression linéaire ; valeurs en %).

importants que ceux du filtre linéaire, il est noté que les taux maximaux sont plus faibles. Dans tous les cas, l'utilisation des filtres n'apporte pas une amélioration sur les résultats.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	68	74	72	72	71	73	72	74	76
Moyen	55	53	52	55	52	52	56	57	56
Minimal	28	12	16	36	25	15	38	34	30

TABLEAU 5.48 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour les états « avant » et « après » (régression non linéaire ; valeurs en %).

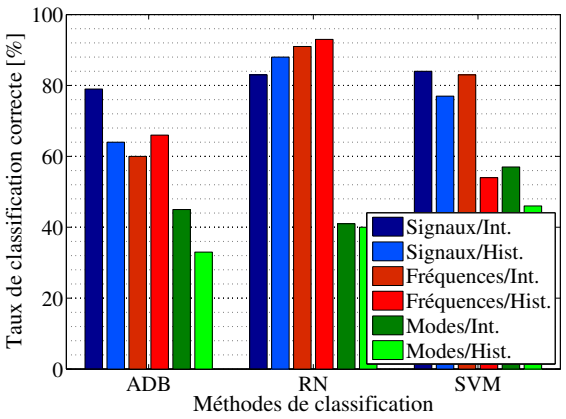
5.6.4 Bilan

Dans les paragraphes précédents, les trois méthodes de classification supervisée ont été appliquées aux données acquises sur les campagnes d'essais réalisées sur pont-rail PK 075+317. Dans un premier temps, une étude utilisant l'ensemble de 41 essais a été considérée. Dans cette étude, l'objectif a été de classer les essais enregistrés pendant les travaux de resserrage des bielles dans l'un des deux groupes : « avant » et « après » resserrage. Il a été considéré que les essais compris dans les phases 1 et 2 devraient être classés dans le premier groupe et ceux des phases 3 et 4, dans le deuxième. Sachant que les travaux de resserrage ont été réalisés de façon progressive, il est intéressant d'observer comment les méthodes de classification peuvent les classer. Les résultats ont montré que les trois méthodes utilisées permettaient de valider l'hypothèse suggérée *a priori*. De plus, il a été remarqué que les essais enregistrés aux limites des phases 2 et 3 présentent un comportement ambigu, pouvant être associés à des classifications parfois inversées. Ceci confirme donc le caractère progressif des travaux.

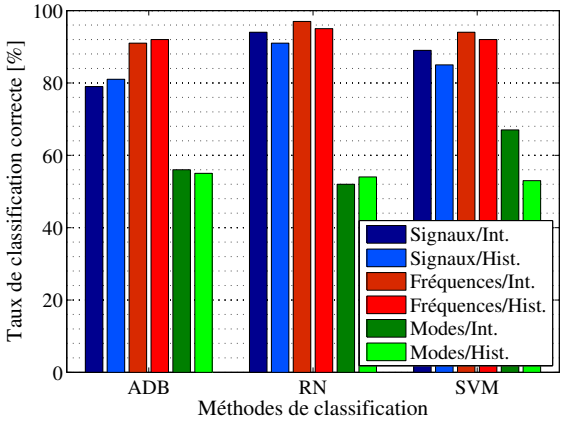
Dans un deuxième temps, une étude statistique similaire à celle menée en utilisant la poutre bi-appuyée a été présentée. Plusieurs simulations ont été réalisées en considérant différentes configurations pour les groupes d'entraînement, de validation et de test. Le but a été de classer correctement les essais enregistrés lors des trois campagnes de mesures en trois groupes distincts. En général, les méthodes des RN et SVM ont été les plus robustes, ayant des taux de classification correcte les plus élevés. Par ailleurs, les résultats obtenus en se basant sur les fréquences propres

ont été meilleurs que ceux basés sur les modes propres. L'utilisation des signaux a également montré être pertinente, notamment dans le cas des RN. Ce résultat est très important, car il montre que les réponses vibratoires peuvent aussi constituer un indicateur des changements structuraux. Les taux moyens globaux sont calculés pour chaque méthode et chaque type de transformation symbolique considérée (cf. Figure 5.37(a)). Il est toutefois remarqué que les résultats obtenus en utilisant les déformées modales ne sont pas entièrement satisfaisants. À l'exception des SVM, les taux de classification correcte ne dépassent pas 50%. Cela montre que la définition des seuils de séparation pour les modes propres est moins évidente que celle pour les signaux ou les fréquences propres.

Finalement, la troisième étude n'a considéré que les essais correspondant aux états structuraux « avant » et « après » resserrage. Étant donné que ces états sont relativement plus « éloignés », l'objectif a été d'observer si les résultats obtenus sont meilleurs que ceux de l'analyse précédente. Il a été observé que les taux de classification correcte ont augmenté sensiblement pour toutes les méthodes, notamment si les signaux et les fréquences propres sont considérés. La Figure 5.37(b) synthétise les taux moyens globaux calculés pour chacune des simulations réalisées. Les résultats obtenus en utilisant les signaux appliqués aux RN montrent que l'approche proposée est adéquate pour la discrimination d'états structuraux.



(a) Taux de classification pour les états « avant », « pendant » et « après »

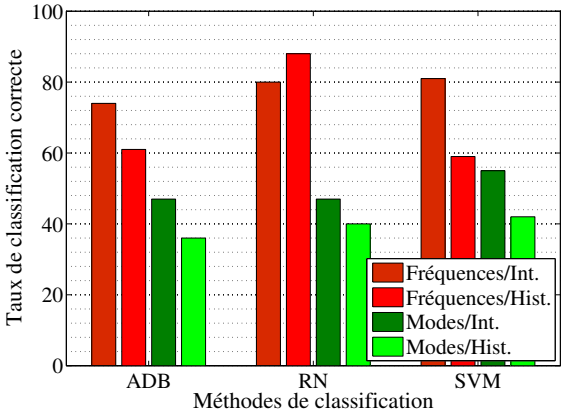


(b) Taux de classification pour les états « avant » et « après »

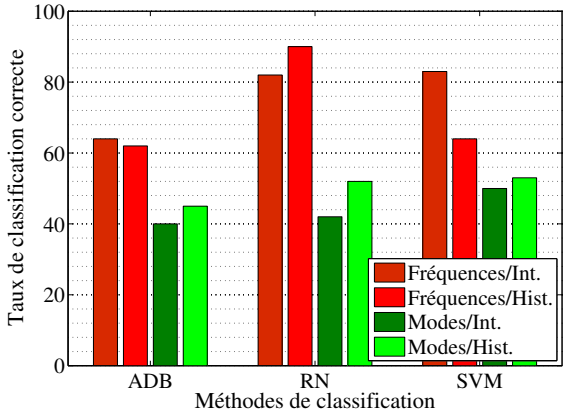
FIGURE 5.37 – Synthèse des résultats obtenus en utilisant les méthodes de classification supervisée.

Compte tenu des résultats obtenus par les méthodes de classification supervisée, une conclusion sur la prise en compte des effets thermiques peut être tirée. D'une part, il y a certains résultats qui sont relativement meilleurs (notamment dans le cas des ADB), ce qui pourrait confirmer le besoin de « filtrer » ces effets. D'autre part, il y en a d'autres qui se trouvent dégradés (comme pour les RN et les SVM, en général), ce qui pourrait indiquer que l'approche proposée n'est pas pertinente. Dans ce contexte, il est possible de dire que, pour cet ouvrage, les effets thermiques ne jouent pas un rôle important dans la discrimination des états structuraux. Les Figures 5.38(a) et 5.38(b) illustrent la synthèse des résultats obtenus pour les deux types de filtres thermique utilisés. Comme il a été observé pour les méthodes de partitionnement, les taux de classification correcte sont non seulement similaires pour les deux approches, mais aussi similaires avec ceux obtenus sans considérer l'effet de la variation des températures (Figure 5.37(a)).

Par ailleurs, l'étude complémentaire tenant compte uniquement des essais enregistrés dans les états « avant » et « après » resserrage des bielles a été considérée. Les analyses considèrent



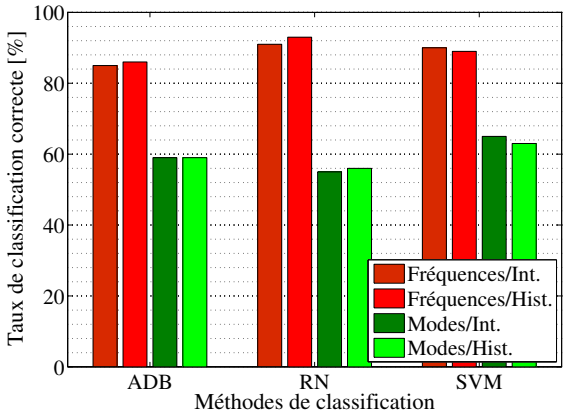
(a) Taux de classification obtenus en utilisant le filtre linéaire



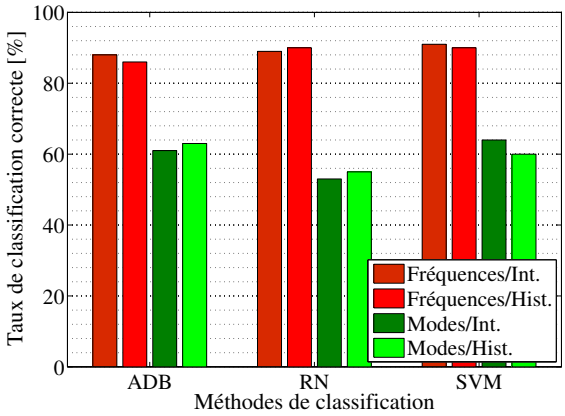
(b) Taux de classification obtenus en utilisant le filtre non linéaire

FIGURE 5.38 – Synthèse des résultats obtenus en utilisant les méthodes de classification supervisée (états avant, pendant et après).

les méthodes de classification appliquées aux paramètres modaux corrigés à partir des filtres thermiques proposés. La Figure 5.39(a) montre les taux moyens globaux évalués à partir de chaque méthode de classification considérée et du type de transformation symbolique utilisé pour le filtre linéaire. Il est noté que les pourcentages sont relativement plus faibles que ceux obtenus sans la prise en compte des effets thermiques (cf. Figure 5.37(a)). Toutefois, les classifications obtenues en considérant les déformées modales semblent être légèrement plus pertinentes. En général, les méthodes des RN et SVM sont les plus performantes, notamment lorsqu'elles sont appliquées aux fréquences propres. Par ailleurs, le filtre non linéaire est également considéré (Figure 5.39(b)). Dans ce cas, les résultats obtenus par les ADB sont relativement meilleurs que ceux avec le filtre linéaire. Toutefois, les taux de classification obtenus par les RN sont relativement plus faibles. Il est également noté que l'utilisation des essais « avant » et « après » mène à des résultats significativement meilleurs que ceux obtenus en considérant tous les états structuraux.



(a) Taux de classification obtenus en utilisant le filtre linéaire



(b) Taux de classification obtenus en utilisant le filtre non linéaire

FIGURE 5.39 – Synthèse des résultats obtenus en utilisant les méthodes de classification supervisée (états avant et après).

5.7 Affectation de nouveaux essais

Il a été mentionné précédemment qu'en 2003, SITES est intervenue avec la SNCF dans le projet national d'Évaluation Dynamique des Ponts et a effectué, dans ce contexte, trois campagnes de mesures dynamiques sur le pont-rail PK 075+317. En décembre 2004, la SNCF a souhaité procéder à une nouvelle campagne de mesure de même type sur cet ouvrage, pour suivre son comportement dynamique dans le temps. L'étude considérée dans ce paragraphe fait donc suite à cette campagne, identique à celles de juin 2003 présentées dans les paragraphes précédents. Au total, cinq nouvelles séries de mesures ont été réalisées :

- **première série :** réalisée en décembre 2004, avec 21 essais enregistrés,
- **deuxième série :** réalisée entre mai et juin 2005, avec 107 essais enregistrés,
- **troisième série :** réalisée entre juillet et octobre 2005, avec 254 essais enregistrés,
- **quatrième série :** réalisée entre novembre et décembre 2005, avec 52 essais enregistrés,
- **cinquième série :** réalisée entre janvier et mars 2006, avec 141 essais enregistrés.

Pour toutes les nouvelles séries de mesures, le plan d'instrumentation utilisé lors des campagnes réalisées en 2003 est resté le même (Fig 5.4). La seule différence, toutefois, est le capteur 1 qui n'est plus utilisé. Par ailleurs, la fréquence d'échantillonnage des signaux est passée à 1/500 s.

L'objectif de cette étude est d'appliquer les approches proposées dans le paragraphe 3.1.6.1 pour l'affectation des nouveaux essais aux clusters définis précédemment, qui représentent les différents états structuraux de l'ouvrage (i.e., « avant », « pendant » et « après »). De plus, cette démarche permet d'affecter des essais à un nouveau cluster représentant un nouveau comportement structural, différent de ceux existants. Le but est d'évaluer l'efficacité des travaux de resserrage menés en 2003 au cours des années suivantes.

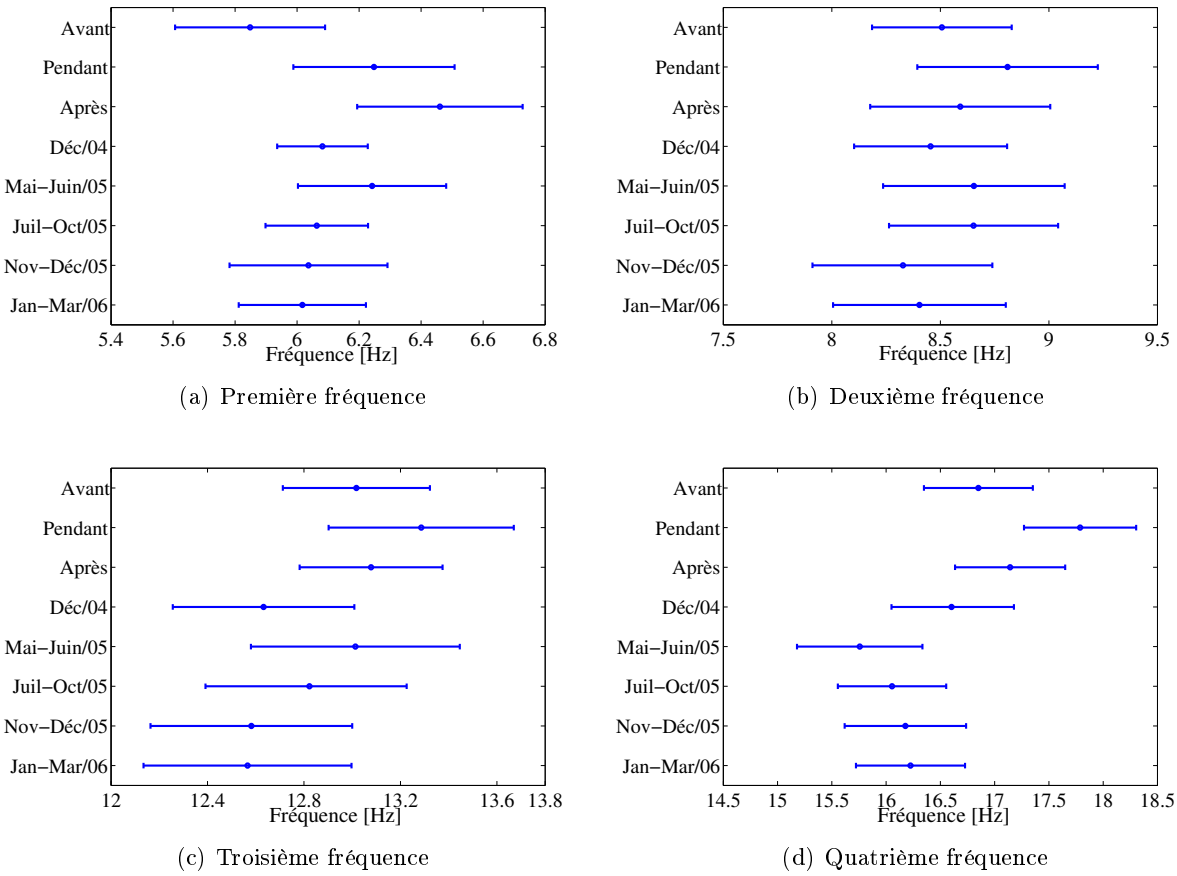
Préalablement à l'étude d'affectation des nouveaux essais par méthodes non supervisée et supervisée, une brève étude comparative des fréquences propres identifiées dans chaque campagne de mesure est proposée. La Figure 5.40 illustre la variation des quatre premières fréquences propres au cours du temps en considérant des intervalles de confiance à 95%. Pour toutes les fréquences, et en particulier pour la première, les valeurs moyennes augmentent juste après le resserrage des bielles. Toutefois, dans les campagnes réalisées après Décembre 2004, il est noté une diminution de ces valeurs, ce qui peut indiquer une possible perte d'efficacité de la procédure de resserrage. Il est également observé que l'état « pendant » a une variation importante, étant donné qu'il regroupe, dans un certain sens, les fréquences identifiées avant et après le resserrage. Une remarque importante à considérer est que les intervalles de confiance se superposent, ce qui ne permet pas de tirer une conclusion définitive.

5.7.1 Étude des effets thermiques sur les paramètres modaux

Les démarches proposées pour l'affectation de nouveaux essais sont également étudiées en considérant les corrections thermiques sur les paramètres modaux. Étant donné que les nouvelles mesures ont été réalisées à différentes périodes de l'année, il est important d'observer si la variation des températures pendant les campagnes d'acquisition peut influencer l'affectation des essais.

La Figure 5.41 illustre l'allure de la variation des températures au cours des nouveaux enregistrements [1]. Il est noté que l'étendue des variations thermiques est relativement large, oscillant entre -5 °C et 35 °C.

1. Dans la campagne d'essais réalisée en décembre 2004, les températures n'ont pas été mesurées.



(a) Première fréquence

(b) Deuxième fréquence

(c) Troisième fréquence

(d) Quatrième fréquence

FIGURE 5.40 – Intervalles de confiance évalués pour les quatre premières fréquences propres.

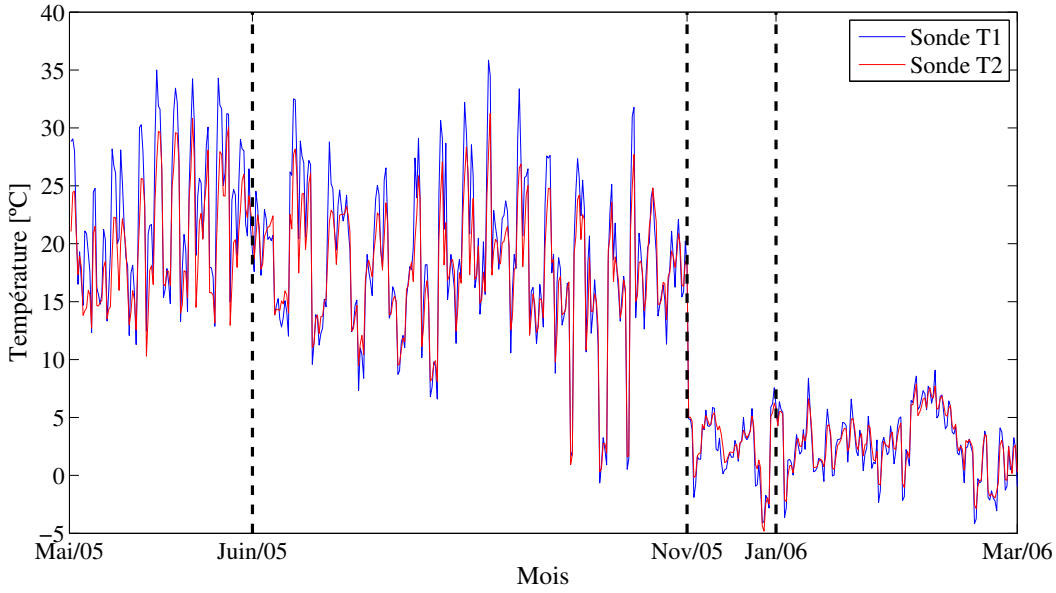


FIGURE 5.41 – Variation des températures au cours du temps pour les sondes T1 et T2.

La procédure utilisée dans ce paragraphe suit le raisonnement expliqué au paragraphe 5.3. Avant d'appliquer les techniques de régression, les valeurs des fréquences propres et des compo-

santes modales identifiées sont normalisées par rapport à une valeur de référence. Comme il a été détaillé dans le paragraphe 5.3, la température 23 °C mesurée par sonde la T2 est utilisée comme température de référence. Les fréquences propres (ainsi que chaque composante modale) sont normalisées par rapport à la valeur identifiée à cette température. Dans les analyses suivantes, les ratios des paramètres modaux identifiés à partir des 41 essais enregistrés lors des campagnes avant, pendant et après resserrage des bielles sont utilisés pour la construction des lois de correction.

Vraisemblablement, les températures enregistrées pendant la période comprise entre Novembre 2005 et Mars 2006 ont été significativement plus faibles que celles des mois précédents. En normalisant les paramètres modaux identifiés pendant cette période par ceux évalués à une température plus élevée (23 °C dans cette étude), les lois de correction peuvent éventuellement sur-pondérer l'effet des basses températures. Toutefois, corriger les paramètres modaux à une température plus faible (0 °C, par exemple), aurait pour conséquence de privilégier une plage de température où peu de mesures ont été effectuées. Une problématique similaire a été soulevée par Crémona (2007).

Le Tableau 5.49 regroupe les vecteurs des poids évalués pour chaque ratio des fréquences propres en utilisant des filtres linéaires. Dans ce tableau, β est le biais du modèle, w_{T1} et w_{T2} sont les poids qui multiplient les enregistrements de température des sondes T1 et T2, respectivement.

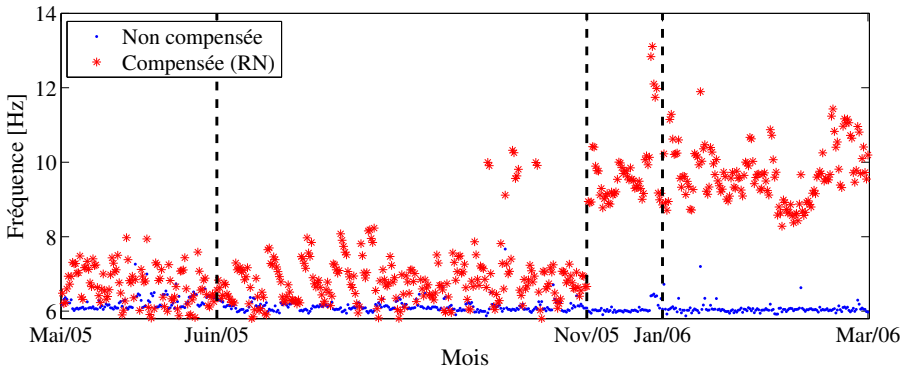
	Fréquence 1	Fréquence 2	Fréquence 3	Fréquence 4
β	0.5795	0.6595	1.1428	1.4539
w_{T1}	-0.0016	0.0079	0.0065	0.0198
w_{T2}	0.0199	0.0071	-0.0136	-0.0410

TABLEAU 5.49 – Vecteurs des poids évalués pour chaque loi de correction thermique (régression linéaire).

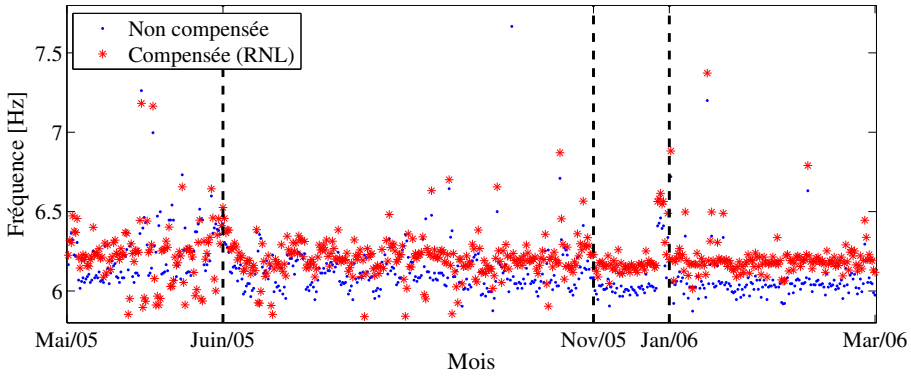
Les lois de correction basées sur la régression non linéaire sont déterminées en utilisant des réseaux de neurones. L'architecture utilisée est composée de 10 neurones dans la couche cachée et les fonctions d'activation utilisées sont la tangente hyperbolique (dans la couche cachée) et la sigmoïde (dans la couche de sortie). Les données d'entrée sont aussi les enregistrements de température des sondes T1 et T2 et les sorties, les ratios des paramètres modaux. L'entraînement du réseau est réalisé en utilisant 28 essais des trois états structuraux (avant, pendant et après) et la validation, les 13 autres essais (comprenant aussi les trois états). 1000 tirages aléatoires sont réalisés (pour différents groupes d'entraînement et de validation). Le modèle choisi est celui dont l'erreur commise lors de phase de validation est la plus faible. Étant donné que le nombre de poids du réseau calibré est conséquent (20 dans la couche cachée + 10 dans la couche de sorties + 11 biais), leurs valeurs ne sont pas détaillées.

Une fois les modèles de régression construits, les fréquences propres (et composantes modales) compensées peuvent être évaluées comme décrit dans le paragraphe 3.4. La Figure 5.42 donne un aperçu de la variation de la première fréquence propre compensée et non compensée des effets thermiques pour chaque état structural. Dans cette étude, il est observé que le filtre linéaire sur-estime les fréquences propres identifiées pendant la période de faibles températures. Les résultats obtenus à partir de la correction thermique utilisant le filtre linéaire sont alors « biaisés ». Inversement, le modèle non linéaire est capable de tenir compte de la tendance des variations au cours de tous les enregistrements. Ceci est l'un des avantages des réseaux de neurones. Une fois le modèle de régression non linéaire entraîné, il arrive à généraliser les valeurs des fréquences propres pour des observations jamais « rencontrés » (i.e., des observations enregistrées à des faibles températures).

Dans les paragraphes suivants, les deux approches proposées dans le paragraphe 3.1.6.1 sont



(a) Comparatif de la première fréquence propre compensée et non compensée (régression linéaire)



(b) Comparatif de la première fréquence propre compensée et non compensée (régression non linéaire)

FIGURE 5.42 – Allure de la variation de la première fréquence propre sans et avec correction des effets thermiques par régression linéaire (RL) et non linéaire (RNL).

appliquées aux signaux et aux paramètres modaux, afin de mieux analyser le phénomène mis en évidence. Pour ces deux approches, des lois de distribution sont ajustées aux vecteurs des distances calculées entre les essais et les prototypes de dans chaque cluster. Par ailleurs, des seuils de distances sont également évalués. Pour chaque application, la loi et le seuil de distance calculé sont détaillés. Une troisième et quatrième approches, complémentaires des deux premières, basées sur le calcul des distances entre les essais et sur les méthodes de classification supervisée, respectivement, sont également proposées. Pour les trois premières approches, la possibilité d'affectation d'un essai à un nouveau groupe différent de ceux existants (i.e., « avant », « pendant » et « après ») est considérée. Pour la dernière démarche, toutefois, seulement trois classes sont prises en compte. Par ailleurs, un comparatif entre les affectations réalisées utilisant le filtre linéaire et le non linéaire est donné. L'objectif est d'observer comment l'ajustement des lois de correction peut influencer l'affectation des essais, notamment dans le cas de la régression linéaire.

5.7.2 Affectation basée sur les divisions hiérarchiques

Dans ce paragraphe, les arbres construits à partir de la méthode des divisions hiérarchiques sont utilisés pour l'affectation des nouveaux essais. L'idée est de faire « parcourir » les nouveaux essais sur les arbres et, ensuite, de les affecter à l'un des clusters créés. Avant de procéder avec l'affectation des essais, les prototypes de chaque cluster sont évalués. Comme il a été décrit

précédemment, les prototypes peuvent être considérés comme une description de chaque groupe. Dans cette thèse, ils sont évalués en calculant la moyenne des représentations symboliques de tous les essais. Une fois les prototypes calculés, les vecteurs des distances entre les essais de chaque cluster et ses prototypes respectifs sont évalués. Cela permet donc, de créer une distribution de distances, d'en ajuster une loi de probabilités et finalement, d'évaluer un seuil à partir duquel un essai n'est plus considéré comme appartenant au cluster. Par conséquent, les 575 nouveaux essais sont utilisés comme données d'entrée pour les arbres construits et affectés aux clusters correspondant.

5.7.2.1 Signaux

Dans le cas des signaux, la Figure 5.14, illustre le partitionnement obtenu en utilisant les signaux transformés en histogrammes. Dans ce cas, il est observé que les capteurs 3 et 10 sont considérés comme les variables discriminantes avec les valeurs de coupures 0,0032 m$\cdot$ s^{-2} et 0,0116 m$\cdot$ s^{-2}, respectivement (Figure 5.14). Dans cet exemple, tous les essais sont affectés dans un nouveau cluster (par souci de clarté du texte, le tableau des résultats n'est pas présenté). Cela signifie qu'en analysant uniquement les réponses dynamiques, la structure a un comportement différent de ceux déjà connus (i.e., « avant », « pendant » et « après » resserrage). Dans le cas des histogrammes, la loi ajustée est la loi log-normale et le seuil est égal à 0,784.

5.7.2.2 Fréquences propres

Le Tableau 5.50 regroupe les taux de classification calculés en considérant l'analyse avec les fréquences propres non corrigées. Dans ce cas, les résultats sont significativement différents. En utilisant des histogrammes, il est observé que les trois premières campagnes sont déjà identifiées dans l'état « avant », tandis que les deux dernières constituent un nouvel état. La loi ajustée est la loi log-normale et la distance seuil est égale à 0,918.

Campagne	Avant	Pendant	Après	NC
Déc-04	**67**	14	0	19
Mai-Juin/05	**56**	24	0	20
Jul-Oct/05	**51**	19	0	30
Nov-Déc/05	27	10	0	**63**
Jan-Mar/06	31	19	0	**50**
Moyenne	**46**	18	0	36

TABLEAU 5.50 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux fréquences propres (histogrammes ; valeurs en %).

Les simulations suivantes sont basées sur les partitionnements obtenus en utilisant le filtre linéaire. L'affectation de nouveaux essais prend en compte les arbres construits à partir de la méthode des divisions hiérarchiques appliquée aux fréquences propres. Le Tableau 5.51 regroupe les taux de classification calculés. Dans ce cas, les résultats sont significativement différents de ceux obtenus sans la correction thermique (cf. Tableau 5.50). Ici, la plupart des essais ne sont jamais classés dans le groupe qui indique un nouveau comportement. Les essais des deux premières campagnes sont affectés au cluster de l'état « pendant », tandis que les essais des deux dernières campagnes sont affectés au groupe « avant ». La loi ajustée est aussi la loi log-normale et le seuil est égal à 0,337.

L'utilisation du filtre non linéaire est considérée ci-dessous. Pour rappel, toutes les séries de mesures sont considérées, à l'exception de la première, réalisée en décembre 2004, où les

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	25	**34**	21	20
Jul-Oct/05	27	**44**	21	8
Nov-Déc/05	**30**	29	18	23
Jan-Mar/06	**35**	26	15	24
Moyenne	29	**33**	19	19

TABLEAU 5.51 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux fréquences propres (régression linéaire ; valeurs en %).

températures n'ont pas été mesurées. Le Tableau 5.52 regroupe les pourcentages d'affectation calculés à partir des fréquences propres transformées en histogrammes. Il est noté que les résultats sont légèrement différents de ceux obtenus en utilisant le filtre linéaire. En effet, ces différences peuvent être expliquées par le mauvais ajustement obtenu par le filtre linéaire. En général, aucun essai n'est classé dans le groupe « après », mais plutôt dans celui correspondant à l'état « pendant ». La loi d'ajustement utilisée est la loi log-normale (seuil égal à 0,906).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	18	**36**	35	12
Jul-Oct/05	23	**31**	28	17
Nov-Déc/05	34	**39**	17	10
Jan-Mar/06	**37**	32	18	13
Moyenne	28	**34**	24	13

TABLEAU 5.52 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux fréquences propres (régression non linéaire ; valeurs en %).

5.7.2.3 Modes propres

Lorsque les modes propres non corrigés sont considérés (et comme il a été observé avec les signaux) tous les essais sont classés dans un nouveau cluster, ce qui indique une nouvelle condition structurale. La loi ajustée est aussi la loi log-normale et la distance seuil est égale à 0,257.

Les résultats obtenus en considérant les modes propres corrigés par le filtre linéaire sont détaillés dans le Tableau 5.53. En général, les essais sont regroupés selon les groupes « avant », « pendant » ou encore dans celui qui indique un état différent. La plupart des essais n'est pratiquement jamais classée dans le groupe « après », quel que soit le type de transformation symbolique utilisé. Dans le cas des histogrammes, la loi log-normale est utilisée pour la détermination du seuil (égal à 0,801).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	8	23	22	**47**
Jul-Oct/05	13	30	26	**31**
Nov-Déc/05	20	**34**	16	30
Jan-Mar/06	**31**	25	20	24
Moyenne	18	28	21	**33**

TABLEAU 5.53 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux modes propres (régression linéaire ; valeurs en %).

L'affectation obtenue par les divisions hiérarchiques appliquées aux modes propres corrigés (filtre non linéaire) est détaillée dans le Tableau 5.54. Comme il a été déjà remarqué, les résultats

obtenus ne permettent pas de conclure quant à l'impact de la correction thermique sur l'affectation des nouveaux essais. En considérant des intervalles, les essais de la deuxième campagne de mesures sont pour la plupart affectés dans l'état « après » (cf. Tableau B.51). Toutefois, cela n'est pas observé avec la transformation en histogrammes. Pour les campagnes suivantes, les essais sont classés soit dans le groupe correspondant à l'état « pendant », soit à celui « avant », en signalant toujours une perte d'efficacité du resserrage des bielles. La loi d'ajustement utilisée est la loi log-logistique (seuil égal à 0,255).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	11	24	23	**42**
Jul-Oct/05	19	**37**	25	19
Nov-Déc/05	29	**38**	17	16
Jan-Mar/06	**34**	21	22	23
Moyenne	23	**30**	22	25

TABLEAU 5.54 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux modes propres (régression non linéaire ; valeurs en %).

5.7.3 Affectation basée sur les nuées dynamiques

5.7.3.1 Signaux

La deuxième démarche pour l'affectation des nouveaux essais considère l'application de la méthode des nuées dynamiques. Dans cette analyse, il n'existe pas d'arbres de décision. Les prototypes sont donc évalués à partir des partitionnements obtenus en considérant les 41 essais enregistrés en 2003. Une fois les prototypes calculés, il est possible de déterminer les seuils à partir des distributions de distances évaluées. En utilisant les signaux, les résultats sont similaires à ceux obtenus par les divisions hiérarchiques : tous les essais sont affectés au cluster NC, ce qui indique un comportement différent de ceux existants. La loi ajustée est la loi log-normale et les seuil des distances est égal à 1,085.

5.7.3.2 Fréquences propres

Le Tableau 5.55 regroupe les taux d'affectation calculés et il est remarqué que la plupart des essais sont classés dans le cluster « avant » resserrage. Cela signale la perte d'efficacité de la procédure menée et montre que l'approche proposée peut fournir une analyse plus détaillée du phénomène (ce qui n'était pas possible en observant la Figure 5.40). En considérant les histogrammes, la loi log-logistique a été retenue pour un seuil égal à 0,717.

Campagne	Avant	Pendant	Après	NC
Déc-04	**52**	10	38	0
Mai-Juin/05	**71**	6	3	21
Jul-Oct/05	**56**	5	31	8
Nov-Déc/05	**33**	**33**	8	27
Jan-Mar/06	**38**	26	7	30
Moyenne	**50**	16	17	17

TABLEAU 5.55 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux fréquences propres (histogrammes ; valeurs en %).

Dans le cas des fréquences propres corrigées par le filtre linéaire, la tendance déjà observé précédemment est encore confirmée. Le Tableau 5.56 regroupe les taux d'affectation calculés et il est remarqué que la plupart des essais sont classés dans le cluster « pendant » le resserrage (contrairement à l'approche ne prenant pas en compte des effets thermiques, qui a indiqué le cluster « avant » resserrage ; cf. Tableau 5.55). Dans le cas des histogrammes, quelques essais sont affectés dans un cluster représentant un nouveau comportement. La loi log-normale est ajustée au vecteur des distances et la valeur seuil est égale à 0,667.

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	19	22	25	**34**
Jul-Oct/05	22	25	24	**29**
Nov-Déc/05	29	**32**	16	23
Jan-Mar/06	27	**30**	15	28
Moyenne	24	27	20	**29**

TABLEAU 5.56 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux fréquences propres (régression linéaire ; valeurs en %).

Les fréquences propres corrigées par le filtre non linéaire sont maintenant analysées. Les essais sont affectés d'une part dans le cluster « pendant » et d'autre part, dans le cluster « avant » (cf. Tableau 5.57). Vraisemblablement, il est observé que l'utilisation du filtre non linéaire change les résultats obtenus par rapport à ceux évalués sans la correction thermique (cf. Tableau 5.55). De plus, les résultats sont aussi différents de ceux obtenus par la correction linéaire. La loi ajustée est la loi log-normale et la distance seuil est égal à 0,433.

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	25	28	**31**	16
Jul-Oct/05	20	**37**	28	15
Nov-Déc/05	**29**	22	22	27
Jan-Mar/06	**35**	30	18	17
Moyenne	27	**29**	25	19

TABLEAU 5.57 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux fréquences propres (régression non linéaire ; valeurs en %).

5.7.3.3 Modes propres

La même procédure est appliquée aux modes propres non corrigés, transformés en histogrammes (cf. Tableau 5.58). Comme il a été observé avec les fréquences propres, la majorité des essais est classée dans le groupe correspondant à l'état « avant » de l'ouvrage. Dans ce cas, aucun essai n'est affecté au cluster « après ». Dans le cas de la transformation en histogrammes, la loi log-logistique a été ajustée aux distances avec un seuil égal à 0,491.

En utilisant les modes propres corrigés par le filtre linéaire, les résultats sont hétérogènes. Dans le cas de la transformation en intervalles interquartiles, il est indiqué que les essais sont associés à l'état « avant » resserrage, « pendant » ou encore à un nouvel état (cf. Tableau B.57). Toutefois, en utilisant des histogrammes, pratiquement tous les essais sont considérés comme appartenant à l'état « pendant » (cf. Tableau 5.59). Dans ce cas, la loi log-normale est ajustée au vecteur des distances (seuil égal à 0,852).

Le Tableau 5.60 regroupe les taux d'affectation calculés en considérant les modes propres compensés par le filtre non linéaire. De façon similaire aux fréquences propres, les résultats sont

Campagne	Avant	Pendant	Après	NC
Déc-04	**95**	0	0	5
Mai-Juin/05	**80**	0	0	20
Jul-Oct/05	**83**	0	0	17
Nov-Déc/05	**87**	0	0	13
Jan-Mar/06	**83**	0	0	17
Moyenne	**86**	0	0	14

TABLEAU 5.58 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux modes propres (histogrammes ; valeurs en %).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	18	28	23	**31**
Jul-Oct/05	20	**31**	26	22
Nov-Déc/05	26	**33**	17	24
Jan-Mar/06	**31**	28	16	25
Moyenne	24	**30**	20	26

TABLEAU 5.59 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux modes propres (régression linéaire ; valeurs en %).

différents de ceux évalués en utilisant le filtre linéaire. Ici, les essais sont classés dans trois des quatre clusters possibles. Dans le cas des histogrammes, par exemple, les essais enregistrés dans les deux premières campagnes sont affectés dans le cluster « après ». La loi d'ajustement utilisée est la loi log-normale (seuil égal à 0,554).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	16	14	**38**	32
Jul-Oct/05	25	24	**27**	24
Nov-Déc/05	29	**31**	14	26
Jan-Mar/06	**39**	34	24	3
Moyenne	**27**	26	26	21

TABLEAU 5.60 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux modes propres (régression non linéaire ; valeurs en %).

5.7.4 Affectation basée sur le calcul de distances

Dans le cadre de l'étude sur le pont-rail PK 075+317, il a été remarqué que les méthodes de clustering n'ont pas réussi à classer parfaitement tous les essais en trois groupes « purs ». Il est alors justifié d'utiliser une démarche alternative. Cette démarche considère connus les trois groupes, c'est-à-dire que chaque groupe regroupe tous les essais d'un type donné (avant, pendant, après). En ce sens, le calcul des prototypes et des lois de distribution des distances mènera à des nouveaux seuils d'affectation. L'objectif est de voir comment les erreurs commises lors des procédures de clustering peuvent influencer les résultats d'affectation des nouveaux essais.

5.7.4.1 Signaux

Dans un premier temps, les signaux transformés en données symboliques sont utilisés. Dans ce cas, pour les deux types de transformation, tous les essais sont affectés au nouveau cluster

NC. Ce résultat est en bon accord avec les classifications déjà obtenues en utilisant les divisions hiérarchiques et les nuées dynamiques. La loi ajustée est la loi log-normale (seuil égal à 0,886) pour les histogrammes.

5.7.4.2 Fréquences propres

Dans un deuxième temps, le calcul des distances est réalisé en considérant les fréquences propres non corrigées. Le Tableau 5.61 regroupe les pourcentages d'affectation pour chacun des états structuraux. Il est noté que l'affectation est toujours réalisée soit dans le cluster « pendant », lors des trois premières séries de mesures, soit dans le cluster « avant », pour les dernières campagnes. Cela montre, encore une fois, que les effets des travaux de resserrage ne sont plus vérifiés. La loi ajustée dans le cas des histogrammes est la loi log-normale (seuil égal à 0,913).

Campagne	Avant	Pendant	Après	NC
Déc-04	25	**35**	18	22
Mai-Juin/05	23	**33**	15	29
Jul-Oct/05	37	**40**	18	5
Nov-Déc/05	**43**	25	16	16
Jan-Mar/06	**37**	37	6	20
Moyenne	36	**40**	12	12

TABLEAU 5.61 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux fréquences propres (valeurs en %).

L'utilisation du filtre linéaire est maintenant considérée. Le Tableau 5.62 regroupe les pourcentages d'affectation pour chacun des états structuraux. Il est noté un accord entre les résultats de la transformation en intervalles (cf. Tableau B.60) et en histogrammes. Dans les deux cas, l'affectation est toujours réalisée soit dans le cluster « pendant » lors des trois premières séries de mesures, soit dans le cluster « avant », pour la dernière campagne. Ces résultats montrent que les effets du resserrage ne sont plus effectifs. La loi ajustée est la loi log-normale (seuil égal à 0,331).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	1	**59**	40	0
Jul-Oct/05	2	**57**	38	3
Nov-Déc/05	22	**35**	22	21
Jan-Mar/06	**44**	28	12	16
Moyenne	17	**45**	28	10

TABLEAU 5.62 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux fréquences propres (régression linéaire ; valeurs en %).

De plus, cette démarche est appliquée aux fréquences propres compensées par le filtre non linéaire. Dans ce cas, les taux d'affectation changent significativement. La plupart des essais sont classées dans le cluster « pendant » (dans le cas des intervalles ; cf. Tableau B.61) et dans le cluster « avant » (dans le cas des histogrammes), comme détaillé dans le Tableau 5.63. Bien qu'il n'existe pas un cluster prédominant, il est noté que les essais ne sont pratiquement jamais classés dans le groupe « après ». La loi d'ajustement utilisée est la loi log-normale (seuil égal à 0,837).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	**33**	25	25	17
Jul-Oct/05	37	**35**	14	14
Nov-Déc/05	**34**	32	18	16
Jan-Mar/06	**37**	25	29	9
Moyenne	**35**	29	22	14

TABLEAU 5.63 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux fréquences propres (régression non linéaire ; valeurs en %).

5.7.4.3 Modes propres

L'analyse avec les modes propres non corrigés est maintenant considérée. En utilisant la transformation en histogrammes, tous les essais sont considérés comme appartenant à un nouvel état structural. Dans ce cas, la loi log-normale est la mieux calée (seuil égal 0,659).

Pour une analyse considérant les déformées modales corrigées par le filtre linéaire, la plupart les essais sont considérés comme appartenant à un nouvel état structural, comme il a aussi été observé précédemment (cf. Tableau 5.64). La loi ajustée est la loi log-normale (seuil égal à 0,776).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	15	20	22	**43**
Jul-Oct/05	18	24	23	**36**
Nov-Déc/05	23	29	15	**32**
Jan-Mar/06	25	25	15	**34**
Moyenne	20	25	19	**36**

TABLEAU 5.64 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux modes propres (régression linéaire ; valeurs en %).

Par ailleurs, le Tableau 5.65 regroupe les résultats obtenus en utilisant les modes propres corrigés (filtre non linéaire). Pour la deuxième campagne d'essais et en considérant les deux types de transformations symboliques, la majorité des essais est affectée au groupe « pendant ». Ces résultats ne concordent pas totalement avec ceux obtenus en utilisant le filtre linéaire, ce qui remet en cause encore une fois la pertinence de la correction thermique dans l'analyse menée. La loi d'ajustement utilisée est la loi log-logistique (seuil égal à 0,552).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	25	25	15	**35**
Jul-Oct/05	33	**39**	13	15
Nov-Déc/05	18	**45**	20	17
Jan-Mar/06	**37**	28	19	16
Moyenne	28	**34**	17	21

TABLEAU 5.65 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux modes propres (régression non linéaire ; valeurs en %).

D'après ces résultats, il est observé que les erreurs éventuellement commises lors des procédures de clustering n'influencent pas significativement l'affectation des essais. De plus, pratiquement aucun essai n'est classé dans le groupe correspondant à l'état « après », ce qui indique la perte d'efficacité des travaux réalisés.

5.7.5 Affectation basée sur les méthodes de classification supervisée

À titre de complément, une approche alternative est considérée dans ce paragraphe. Cette démarche consiste à utiliser les méthodes de classification supervisée pour l'affectation des nouveaux essais à l'une des classes connues (i.e., « avant », « pendant » et « après » resserrage). Dans cette étude, les 41 essais enregistrés dans les campagnes de 2003 sont utilisés pour l'entraînement (et validation, dans le cas des RN et des SVM) des méthodes. Par conséquent, les 575 nouveaux essais sont utilisées dans la phase de test. La différence de cette analyse avec les trois précédentes est qu'elle ne considère pas la possibilité de ne pas classer un essai dans les trois classes connues. Bien que ceci ne soit pas toujours vrai, il est intéressant d'observer comment ces tests sont affectés.

Pour la méthode basée sur les RN, l'architecture du réseau utilisée est composée d'une couche cachée avec 10 neurones. Par ailleurs, les fonctions d'activation utilisées sont la fonction sigmoïde (cf. Équation (3.41)) dans la couche cachée et la fonction Heaviside (cf. Équation 3.40) dans la couche de sortie. Le taux d'apprentissage η est fixé à 0,01 et la constante de moment μ est fixé à 0,001. En ce qui concerne les SVM, la fonction noyau RBF est toujours utilisée et le paramètre σ est déterminé de manière itérative lors de la phase de validation.

5.7.5.1 Signaux

Le Tableau 5.66 regroupe les affectations obtenues pour chaque méthode en utilisant les signaux transformés en histogrammes. Dans ce cas, les méthodes basées sur les RN et les SVM indiquent que pratiquement tous les nouveaux essais sont affectés dans la classe correspondante à l'état « pendant » renforcement. Dans le cas des ADB, ceci est également observé, à l'exception de la troisième campagne de mesures où l'affectation est réalisée à la classe « avant » travaux.

Campagne	**ADB**			**RN**			**SVM**		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Déc/04	37	**44**	19	18	**80**	2	0	100	0
Mai-Juin/04	39	**43**	18	5	**79**	16	0	100	0
Jul-Oct/05	**45**	22	33	12	**70**	18	0	100	0
Nov-Déc/05	22	**56**	22	29	**56**	15	0	100	0
Jan-Mar/06	33	**47**	20	24	**61**	15	0	100	0
Moyenne	35	**42**	22	18	**69**	13	0	**100**	0

TABLEAU 5.66 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux signaux transformés en histogrammes (valeurs en %).

5.7.5.2 Fréquences propres

Cette démarche est maintenant appliquée aux fréquences propres non compensées. La tendance déjà observée en utilisant les signaux est encore vérifiée. Le Tableau 5.67 regroupe les pourcentages calculés en utilisant des histogrammes. Dans ce cas, les méthodes des RN et des SVM indiquent toujours la même classification. Les ADB, toutefois, indiquent l'état « après » pour la première série de mesures, l'état « avant » pour les deux dernières et « pendant » pour les intermédiaires.

Le Tableau 5.68 regroupe les affectations obtenues pour chaque méthode en utilisant les fréquences propres corrigées transformés en histogrammes. Dans cette analyse, toutes les méthodes indiquent une affectation soit à la classe « pendant », soit à classe « avant ».

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Déc/04	15	37	**48**	0	100	0	0	100	0
Mai-Juin/04	28	**44**	28	0	100	0	0	100	0
Jul-Oct/05	26	**39**	35	0	100	0	0	100	0
Nov-Déc/05	**38**	32	30	0	100	0	0	100	0
Jan-Mar/06	**45**	30	25	0	100	0	0	100	0
Moyenne	30	**36**	33	0	**100**	0	0	**100**	0

TABLEAU 5.67 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux fréquences propres transformées en histogrammes (valeurs en %).

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	**48**	29	23	0	100	0	**45**	34	21
Jul-Oct/05	29	**44**	27	0	100	0	**44**	21	35
Nov-Déc/05	31	**39**	30	0	100	0	**46**	30	24
Jan-Mar/06	**36**	33	31	0	100	0	**37**	33	30
Moyenne	36	**36**	28	0	100	0	**43**	30	28

TABLEAU 5.68 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux fréquences propres transformées en histogrammes (régression linéaire ; valeurs en %).

Le Tableau 5.69 présente les taux d'affectation évalués en utilisant les fréquences propres transformées en histogrammes, corrigées par le filtre non linéaire. Les RN indiquent toujours l'affectation de tous les essais dans le cluster « pendant ». Les ADB et les SVM, toutefois, tendent à partager les affectation entre les groupes « avant » et « pendant » resserrage. De plus, les ADB indiquent l'affectation dans le groupe « après » pour les essais enregistrés lors de la deuxième campagne d'essais.

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	16	39	**45**	0	100	0	33	**44**	23
Jul-Oct/05	25	**41**	34	0	100	0	**39**	30	31
Nov-Déc/05	**37**	28	35	0	100	0	**39**	25	36
Jan-Mar/06	**42**	27	31	0	100	0	**40**	26	34
Moyenne	30	34	**36**	0	**100**	0	**38**	31	31

TABLEAU 5.69 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux fréquences propres transformées en histogrammes (régression non linéaire ; valeurs en %).

5.7.5.3 Modes propres

Les modes propres sont considérés et l'affectation obtenue en utilisant des histogrammes est détaillé dans le Tableau 5.70. Ainsi, comme il a été observé avec les signaux, les résultats obtenus par les ADB, RN et SVM sont pratiquement les mêmes. Toutefois, seuls les RN affectent tous les essais dans la classe « pendant ». Les ADB et les SVM ne le font que pour les quatre dernières campagnes.

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Déc/04	20	39	**41**	0	100	0	8	45	**47**
Mai-Juin/04	25	**44**	31	0	100	0	18	**67**	15
Jul-Oct/05	12	**57**	31	0	100	0	22	**75**	3
Nov-Déc/05	17	**51**	32	0	100	0	15	**81**	4
Jan-Mar/06	29	**44**	27	0	100	0	12	**88**	0
Moyenne	21	**47**	32	0	**100**	0	15	**71**	14

TABLEAU 5.70 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux modes propres transformés en histogrammes (valeurs en %).

Lorsque les modes propres compensés sont utilisés, les pourcentages d'affectation des nouveaux essais sont détaillés dans le Tableau 5.71. À l'exception des ADB qui, pour la deuxième campagne, affectent les essais dans le groupe « après » resserrage, les autres résultats sont identiques à ceux vérifiés pour les intervalles (cf. Tableau B.70).

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	20	27	**53**	0	100	0	0	100	0
Jul-Oct/05	29	**38**	33	0	100	0	0	100	0
Nov-Déc/05	**37**	31	32	0	100	0	0	100	0
Jan-Mar/06	**40**	28	32	0	100	0	0	100	0
Moyenne	32	31	**38**	0	**100**	0	0	**100**	0

TABLEAU 5.71 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux modes propres transformés en histogrammes (régression linéaire ; valeurs en %).

Les résultats concernant les modes propres corrigés par le filtre non linéaire sont détaillés dans le Tableau 5.72. Dans ce cas, les affectations sont exactement les mêmes pour les RN et les SVM, à savoir la classification des essais sur dans le cluster « pendant ». Quelques différences sont notées par rapport aux résultats obtenus avec les ADB, qui affectent toujours les essais dans le groupe « pendant », ou dans le groupe « avant ».

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	13	35	**52**	0	100	0	0	100	0
Jul-Oct/05	28	**39**	33	0	100	0	0	100	0
Nov-Déc/05	30	**38**	32	0	100	0	0	100	0
Jan-Mar/06	**36**	30	34	0	100	0	0	100	0
Moyenne	27	36	**38**	0	**100**	0	0	**100**	0

TABLEAU 5.72 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux modes propres transformés en histogrammes (régression non linéaire ; valeurs en %).

5.7.6 Bilan

Dans le contexte d'une nouvelle série de campagnes de mesures dynamiques réalisée par SITES à la demande de la SNCF, plusieurs essais ont été enregistrés au cours des années. Les campagnes, ayant débuté un an après la procédure de resserrage menée en juin 2003, avaient pour

but d'observer l'évolution des caractéristiques modales du pont et aussi de suivre l'efficacité des travaux réalisés. Au total, 575 nouveaux essais ont été enregistrés sur 5 séries de mesures réalisées entre fin 2004 et début 2006.

Dans un premier temps, une analyse classique, considérant les fréquences propres identifiées lors des campagnes de 2003 et de celles de 2004 à 2005, a été menée. Cette analyse a consisté à observer la variation des paramètres modaux à partir de l'évaluation d'intervalles de confiance pour les valeurs moyennes. En observant les graphiques, il a été pratiquement impossible de tirer des conclusions sur le changement des quatre premières fréquences propres et, par conséquent, sur l'efficacité des travaux de resserrage. En effet, étant donné que les intervalles de confiance se superposent, il n'est pas possible d'affirmer avec 95% de confiance, que les valeurs moyennes des fréquences propres ont changé.

Quatre approches originales ont été utilisées pour l'affectation des nouveaux essais, afin de mieux appréhender la variation des effets de la procédure menée au cours des années. La première approche a consisté à utiliser les arbres de décision construits à partir de la méthode des divisions hiérarchiques pour affecter les nouveaux essais à l'un des clusters existants. Toutefois, il est également important de tenir compte de la possibilité qu'un essai n'appartienne à aucun des clusters existants. En cela, une analyse complémentaire a été proposée, basée sur l'évaluation de « seuils » à partir d'une distribution de distances. Cette approche a été appliquée aux signaux et paramètres modaux, transformés en intervalles interquartiles et en histogrammes. Même que les résultats aient été hétérogènes (dans le sens où il n'existe pas un cluster remarqué dans toutes les analyses), il a été montré que les essais ne sont pratiquement jamais classés dans le cluster correspondant à l'état « après ». Ce résultat donne de fortes indications sur la perte d'efficacité des procédures de resserrage menées.

De plus, une deuxième démarche basée sur les partitionnements obtenus à partir de la méthode des nuées dynamiques a été considérée. Dans ce cas, des prototypes représentatifs de chaque cluster ont été évalués et une procédure similaire à celle décrite précédemment a été menée. En général, les résultats ont été cohérents avec ceux de l'approche basée sur les divisions hiérarchiques. Bien que quelques essais de la campagne de 2004 aient été affectés dans le cluster « après », la plupart sont affectés dans les cluster « avant », « pendant » ou encore dans celui qui représente un état structural différent de ceux connus. Par ailleurs, une troisième approche basée sur le calcul direct des distances entre les essais a été considérée. Les résultats obtenus sont similaires aux précédents. Cela montre que, dans cette étude, les erreurs commises lors des procédures de clustering ne modifient pas les résultats des procédures d'affectation.

Une étude complémentaire tenant compte des méthodes de classification supervisée a été prise en compte. Dans cette étude, la possibilité d'affectation d'un essai à une nouvelle classe n'est plus considérée. L'objectif est plutôt d'observer comment ces méthodes peuvent extrapoler la connaissance acquise durant la phase d'entraînement. Ici, les résultats ont pratiquement tous indiqué le cluster « pendant » comme étant le comportement caractéristique des nouveaux essais.

Les quatre approches proposées pour l'affectation de nouveaux essais ont été également considérées en tenant compte des corrections thermiques appliquées aux paramètres modaux. Il a été montré qu'en général, les deux types de filtres considérés - linéaire et non linéaire - produisent des taux d'affectation différents. L'utilisation du filtre linéaire dans cette application ne semble pas donc pertinente. Étant donné que le modèle a été construit pour des températures relativement élevées, les paramètres modaux compensés pour des températures plus faibles ont été sur-estimés. Ceci n'est pas observé dans le cas de la régression non linéaire. Ce type de méthode permet une meilleure prédiction des données, même si l'ensemble d'observations utilisées dans l'entraînement est différent de ceux employé dans la phase de test. En général, l'affectation des essais a été pratiquement toujours la même dans les deux cas, soit dans le cluster de l'état « avant », soit dans celui de l'état « pendant ». Toutefois, il convient de remarquer que ces résultats sont significa-

tivement différents de ceux obtenus sans prise en compte des effets thermiques (cf. paragraphe 5.7). Cette observation met en évidence que pour des campagnes d'essais réalisées en différentes périodes de l'année, la prise en compte de la variation thermique sur les paramètres modaux doit être considérée.

Même que les affectations ne soient pas toujours les mêmes (i.e., n'indiquent pas un cluster « prédominant »), il est intéressant de noter que les résultats signalent systématiquement la perte d'efficacité des travaux. Ceci est notamment observé pour les campagnes réalisées au bout d'un an après le resserrage des bielles. Pratiquement toutes les approches d'affectation proposées indiquent alors que le comportement de l'ouvrage est revenu à l'état « pendant » (voire même « avant ») les travaux menés. Cette conclusion n'est possible d'être tirée qu'à partir de l'utilisation de l'ADS appliquée aux approches d'affectation proposées, mais aussi de son couplage avec les méthodes de classification supervisée.

5.8 Conclusions

Dans ce chapitre, une application expérimentale, basée sur des essais dynamiques enregistrés sur le pont-rail PK 075+317, a été proposée. Ce pont ferroviaire à poutrelles enrobées se situe sur la ligne à grande vitesse reliant Paris à Lyon. L'étude menée a été divisée en deux axes majeurs. En premier lieu, l'objectif a été de caractériser et de vérifier les améliorations apportées par une procédure de resserrage des bielles sur le comportement de l'ouvrage. Il a été constaté, lors des campagnes de mesures préliminaires, que la fréquence de passage des TGV était très proche de la première fréquence propre de vibration de l'ouvrage. La SNCF a décidé donc de mettre en œuvre un système d'encastrement mécanique des abouts du tablier sur les culées, afin de rigidifier la structure et, par conséquent, augmenter les fréquences propres du pont. En second lieu, le but a été d'évaluer l'impact des variations environnementales (liées à la température, dans le cas présent) sur les caractéristiques modales.

Afin de répondre au premier enjeu, diverses analyses classiques ont été réalisées. Dans un premier temps, une étude des intervalles de confiance appliqués aux fréquences propres a été menée. À l'exception de la première fréquence, tous les intervalles se chevauchent, ce qui rend difficile de conclure avec un niveau de confiance satisfaisant (95%) que les moyennes de ces paramètres ont changé après les travaux de resserrage. Dans une deuxième temps, des tests d'hypothèse ont été réalisés, le but étant de déterminer si les distributions des fréquences propres « avant » et « après » les travaux peuvent être considérées comme identiques. Les résultats ont montré que les première, troisième et quatrième fréquences propres échouent au test (pour un niveau d'acceptation de 5%), indiquant que les deux états peuvent être considérés comme différents. Ceci n'a pas été le cas pour la deuxième fréquence propre, pour laquelle les tests d'hypothèse n'ont pas permis de discriminer ces deux états. Par ailleurs, les coefficients MAC ont été évalués pour les déformées modales, en considérant également les états « avant » et « après » resserrage. Il a été montré que ces coefficients varient entre 99% et 100% pour les première et quatrième déformées modales. Pour les deuxième et troisième modes propres, les valeurs MAC ont été relativement plus faibles (oscillant entre 50% et 99%). En général, la sensibilité de ces types d'analyse ne permet pas de discriminer des modifications structurales avec un niveau de confiance suffisant.

Les approches basées sur le couplage de l'ADS aux méthodes de clustering - divisions hiérarchiques, agglomérations hiérarchiques et nuées dynamiques - et aux méthodes de classification supervisée - arbres de décision Bayésiens, réseaux de neurones et machines à vecteurs supports - ont été ensuite considérées afin d'améliorer la discrimination des états structuraux. Plusieurs études tenant compte de différents sous-ensembles d'essais ont alors été réalisées. Le but initial a été de discriminer les trois états structuraux (avant, pendant et après resserrage) en utilisant la

démarche proposée. Des analyses ont été menées en utilisant les signaux et paramètres modaux transformés en intervalles interquartiles et en histogrammes. En général, les résultats ont montré que ces approches sont capables de discriminer les modifications structurales avec un taux élevé de réussite dans certains cas, surtout en considérant les méthodes supervisées. De plus, comme il a été observé dans les simulations numériques au Chapitre 4, les méthodes de clustering basées sur les nuées dynamiques et les divisions hiérarchiques ont été plus efficaces pour la discrimination des états structuraux, notamment lorsque les fréquences propres sont utilisées. La méthode des agglomérations hiérarchiques, qui n'avait pas permis d'obtenir des résultats satisfaisants lors des simulations numériques, a encore montré une forte tendance à mélanger les essais, surtout dans les clusters associés à l'état intermédiaire (pendant resserrage). De façon similaire, les RN et les SVM ont été plus performants en séparant (dans certains cas, de manière parfaite) les trois états structuraux. Par ailleurs, une étude complémentaire a été menée en ne considérant que les deux états « extrêmes » (i.e., « avant » et « après » le resserrage). Il a été noté que les méthodes discriminent généralement mieux ces deux états avec des taux de classification relativement importants.

Par la suite, des analyses considérant l'évaluation des indices de partitionnement optimal appliqués aux méthodes de clustering ont été proposées. Comme pour les simulations numériques au Chapitre 4, les indices CH et Γ sont, en général, les plus pertinents. Ces indices ont pratiquement toujours indiqué le nombre attendu de clusters, notamment en étant basés sur les fréquences propres. L'indice C^* a toutefois montré être le plus sensible aux variations structurales, en signalant souvent un nombre plus élevé de groupes de comportement. Les indices sont aussi une façon de valider les partitionnements obtenus. Par exemple, si un procédé de clustering produit des clusters très hétérogènes, ce résultat est directement reflété sur le calcul des indices. En ce sens, il est probable qu'ils n'indiqueront pas le partitionnement obtenu précédemment comme optimal.

Par ailleurs, quatre approches originales pour l'affectation des nouveaux essais ont été proposées. Ces approches ont été appliquées dans un contexte de nouvelles campagnes de mesures réalisées par la SNCF lors des années 2004, 2005 et 2006. Ces séries d'essais ont été menées afin d'apprécier les effets des travaux de resserrage au cours du temps. Au total, 575 nouveaux essais ont été enregistrés. En premier lieu, une analyse classique simple a été considérée, en calculant les intervalles de confiance pour la moyenne des fréquences propres identifiées. Les résultats n'ont pas permis de tirer des conclusions définitives, ce qui a conduit à l'application des nouvelles démarches. La première démarche pour l'affectation de nouveaux essais a considéré les arbres de décision obtenus à partir de la méthode des divisions hiérarchiques. L'idée est de faire « parcourir » au nouvel essai l'arbre construit et d'observer où il est affecté. Toutefois, cette procédure peut mener à des résultats incohérents, étant donné qu'elle oblige la classification du nouvel essai dans un des groupes existants. Ainsi, une analyse complémentaire a été introduite en définissant le calcul des seuils de distances. Ces seuils sont utilisés pour déterminer si un essai *a priori* affecté dans un tel groupe, y appartient effectivement ou non. Par ailleurs, une deuxième approche basée sur la méthode des nuées dynamiques a été considérée, le but étant d'affecter les nouveaux essais par rapport aux prototypes évalués pour chaque cluster. La troisième méthode proposée ne considère aucune méthode de partitionnement. En réalité, l'existence de trois groupes « purs » est considéré (i.e. un groupe d'essais « avant » qui ne contient que les 15 essais de type « avant », un groupe « pendant » avec seuls les 13 essais de type « pendant » et un groupe « après » n'ayant que les 13 essais enregistrés « après » resserrage). L'objectif est d'observer comment les erreurs commises lors de la définition des clusters par les méthodes de clustering peuvent changer l'affectation des nouveaux essais. En général, il a été remarqué que cette analyse produit des résultats similaires à ceux obtenus avec les divisions hiérarchiques et les nuées dynamiques. Cela met en évidence, dans un certain sens, que le fait de ne pas avoir de clusters « purs » (i.e., n'ayant qu'un seul type d'essais) n'indique pas nécessairement que l'affectation sera fausse. Il faut considérer,

toutefois, qu'un petit nombre d'essais a été classé incorrectement avec ce type d'analyse. Cela signifie que les prototypes évalués lors des partitionnements obtenus sont similaires à ceux évalués à partir de clusters « purs ». Il est clair que si les procédés de clustering produisent des clusters très hétérogènes, ceci affectera la classification des nouveaux essais. La dernière approche n'a considéré que la possibilité de classer les nouveaux essais selon les trois classes existantes (« avant », « pendant » et « après »), en utilisant les trois méthodes de classification supervisée. Même si cette analyse est plus restreinte que les précédentes, il est également important d'observer comment les nouveaux essais sont affectés à l'un de ces états structuraux. Les résultats ont été relativement peu convaincants pour toutes les approches utilisées. Cette ambiguïté se reflète dans le sens où il n'y a pas eu un cluster « prédominant » sur toutes les affectations. Il est cependant possible de conclure que l'efficacité des travaux de resserrage n'est pas aussi marquée qu'au moment du renforcement. Ceci est surtout observé dans les campagnes de 2005 et 2006, où pratiquement toutes les approches d'affectation ont indiqué que le comportement de l'ouvrage est revenu à l'état « pendant » (voire « avant ») les travaux de renforcement menés.

Finalement, afin de répondre au deuxième enjeu de cette thèse, c'est-à-dire évaluer l'impact des effets thermiques sur les paramètres modaux, deux analyses de régression ont été présentées. La première est une régression de type linéaire multiple et la seconde est une régression non linéaire basée sur des réseaux de neurones. La procédure a été menée en considérant deux étapes : i) filtrage des effets de la variation de températures sur les fréquences et modes propres, à partir de l'utilisation des techniques de régression proposées ; ii) utilisation des paramètres modaux, « corrigés » des effets thermiques, dans des méthodes de clustering, de classification supervisée et des procédures d'affectation des nouveaux essais. Comme déjà mentionné, l'objectif a été d'observer comment les résultats obtenus précédemment pouvaient changer lorsque les effets thermiques sont considérés. Dans le cas des méthodes de clustering et de classification supervisée, il a été remarqué que les taux de réussite peuvent être améliorés ou dégradés. Dans ce cas particulier, la variation des températures ne semble pas causer un effet conséquent sur les paramètres modaux. Étant donné le caractère hétérogène des résultats, cela montre que, pour cet ouvrage et pendant la période analysée, il n'est pas pertinent de prendre en compte les effets de la variation de température. Il est important de noter que pratiquement tous les essais ont été enregistrés aux mêmes horaires et dans des plages de températures similaires. Par ailleurs, une deuxième étude concernant l'affectation des nouveaux essais a été considérée. Dans ce dernier cas, les analyses ont montré que la prise en compte des effets thermiques a modifié les résultats obtenus. En effet, les essais correspondant aux campagnes des mesures ont été enregistrés à différentes périodes de l'année (donc, à des températures différentes). Par ailleurs, il a été remarqué que l'utilisation du filtre linéaire dans cette application n'est pas pertinente. Étant donné que le modèle a été construit pour des températures plus élevées, les paramètres modaux compensés pour des températures plus faibles ont été sur-estimés. Ceci n'est pas observé dans le cas de la régression non linéaire. Bien que la conclusion soit la même - perte d'efficacité des travaux de resserrage - l'affectation des essais aux différents groupes change significativement.

Chapitre 6

Étude du pont PI-57 sur l'Oise

Dans ce chapitre, l'étude sur la surveillance dynamique du pont routier PI-57 reliant Paris à Lille et franchissant l'Oise est exploitée (Cury & Crémona 2010*a*). En premier lieu, la procédure d'identification modale menée sur l'ouvrage est détaillée. En deuxième lieu, les analyses de clustering et de classification supervisée couplées à l'ADS sont mises en œuvre. Par ailleurs, de la même façon que l'étude du pont-rail au chapitre précédent, une analyse des effets des variations thermiques sur les paramètres modaux est menée. Pour cela, les techniques de régression linéaire et non linéaire, introduites au paragraphe 3.4, sont utilisées. L'approche basée sur l'Analyse en Composantes Principales (ACP) décrite dans le paragraphe 3.5 est également considérée.

6.1 Présentation du cas d'étude

Dans le cadre d'un contrat de collaboration scientifique entre le LCPC, le LRPC de Bordeaux et la SANEF (Société des Autoroutes du Nord et de l'Est de la France), le suivi dynamique d'un pont routier a été considéré avec un double objectif : d'une part caractériser et quantifier les effets d'une procédure de renforcement sur le comportement de l'ouvrage, et d'autre part évaluer l'impact des variations environnementales sur les caractéristiques modales à partir de mesures continues de la température au cours des années.

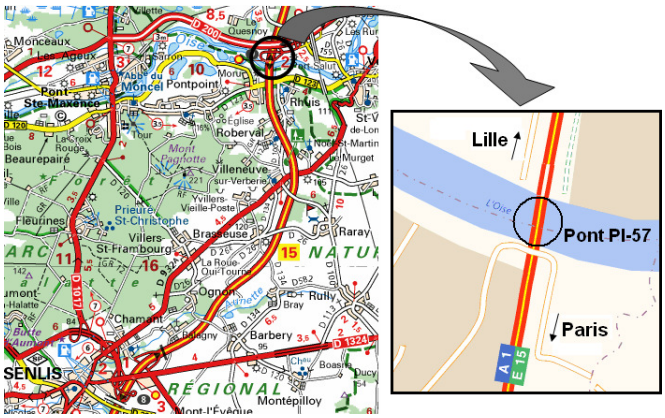


FIGURE 6.1 – Localisation du pont PI-57.

L'ouvrage concerné est le PI-57 situé dans la commune de Moru dans le département de l'Oise, permettant à l'autoroute A1 (Paris - Lille) de franchir l'Oise (Figure 6.1). Ce pont est

un ouvrage en béton précontraint qui a été construit dans les années 1964-1965 (Figure 6.2). Sa structure comporte des poutres caissons précontraintes à hauteur variable (de 1,60 m à 4,30 m), ayant 3 âmes et 2 hourdis (supérieur et inférieur) d'une longueur totale de 116,50 m (dont deux travées de 18 m et une travée intermédiaire de 80,5 m). Les travées centrales de chaque tablier ont été construites par encorbellement, pour mise en place de voussoirs coulés sur place. Le clavetage de la clef est assuré par des câbles de précontrainte placés dans les hourdis inférieurs. Les travées de rives sont des travées courtes (16,25 m et 18,00 m) et massives. Elles ont servi de contre-poids et de massifs d'ancrage de la précontrainte de fléau lors de la construction des tabliers (Figure 6.3).

FIGURE 6.2 – Vue de l'ensemble du pont PI-57.

(a) Vue de la cullée du pont

(b) Vue du caisson de la travée centrale

FIGURE 6.3 – Vue extérieure et intérieure du pont.

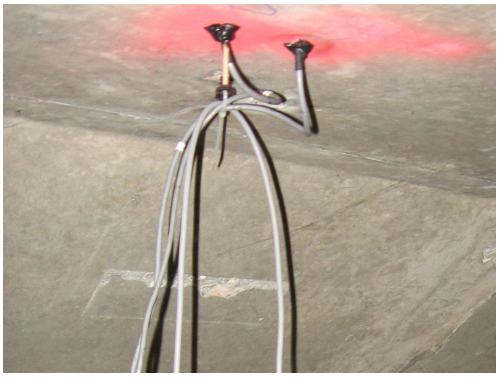
La surveillance du pont PI-57 a été menée à partir d'essais dynamiques réalisés sous excitation ambiante (trafic routier). Le système d'acquisition est composé de 16 accéléromètres positionnés dans les caissons du tablier Lille-Paris. Les accéléromètres utilisés sont de type piezo-électrique 4507B-005 de Bruël & Kjaër (technologie IEPE, -54 à 100 °C) de sensibilité 1V/g, ± 5g et ont été positionnés avec des équerres de fixation en aluminium collées, comme montré dans la Figure 6.4(a). Les mesures accélérométriques sont acquises sur la plage de fréquence 0-30 Hz. L'échantillonnage des signaux est fixée à 1/250 s sur une durée de 5 min toutes les 3 heures. Parallèlement, afin d'étudier l'influence thermique, des mesures de température sont réalisées. L'échantillonnage se fait toutes les heures. Les sondes de température sont de type Pt100 classe B (Figure 6.4(b)). Pour l'acquisition des accélérations, le système d'acquisition se compose :

– d'un contrôleur de données Gantner de type E-pac DL sur lequel est branché une clé USB

de 4 Go permettant le stockage des données,
- des modules A21 munis d'une voie d'acquisition accélérométrique IEPE,
- d'un modem TCP/IP permettant la récupération des données et la configuration du système à distance via un protocole FTP,

Le contrôleur E-pac, la clé USB et l'alimentation 24V ont été placés dans une armoire de type électrique à l'entrée du caisson à proximité du routeur internet mis en place par le service communication de la SANEF. Dans chaque caisson, une liaison série (Profibus + alimentation 24V) a été établie entre les modules d'acquisitions et le contrôleur E-pac.

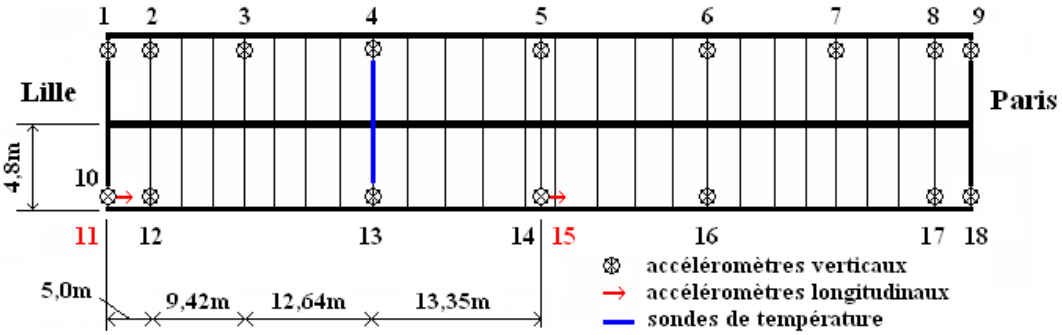
(a) Accéléromètres placés dans le caisson du pont



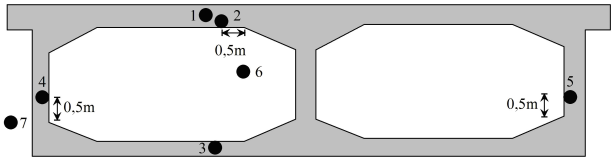
(b) Sonde de température

FIGURE 6.4 – Vue des capteurs instrumentés.

Le plan d'instrumentation des capteurs accélérométriques est montré à la Figure 6.5(a). Ce plan d'instrumentation permet de caractériser essentiellement les modes propres de flexion verticale et de torsion. L'accéléromètre 13 est utilisé comme capteur de référence pour la normalisation des déformées modales. Par ailleurs, 7 sondes de températures ont été positionnées sur un profil du caisson (Figure 6.5(b)). Le profil instrumenté se situe à 27 m de l'about côté Lille de l'ouvrage.



(a) Plan d'instrumentation des capteurs accélérométriques



(b) Plan d'instrumentation des sondes de température

FIGURE 6.5 – Implantation des accéléromètres et des sondes de température.

Cet ouvrage a subi divers problèmes pendant et après sa construction, provoquant l'apparition de fissures localisées et l'augmentation de la déflexion à mi-travée. Ces problèmes sont

principalement dus à l'insuffisance de précontrainte, causée par des effets de retrait et fluage et à la non prise en compte des contraintes thermiques lors du dimensionnement et de la construction (les connaissances sur ce type de contraintes étaient limitées à l'époque de la construction de ce pont). En raison du risque potentiel de fissuration du tablier, des études numériques ont montré que l'intégrité de l'ouvrage à long terme pouvait être affectée si des mesures correctives n'étaient pas prises rapidement. À partir de ces évaluations techniques et compte tenu de l'importance de la structure, la SANEF a décidé de renforcer les deux tabliers.

Pendant l'été 2009, cet ouvrage a subi des travaux de renforcement par l'ajout de câbles de précontrainte. La procédure de renforcement a consisté à réduire les contraintes de traction sous charges mobiles à 1,5 MPa dans la partie inférieure des sections droites du pont. Les contraintes de traction peuvent atteindre jusqu'à 5,10 MPa dans la section droite située à mi-portée. Des calculs effectués par le bureau d'études ACOGEC sous la supervision de la SANEF ont évalué la force totale de précontrainte nécessaire à 32000 kN, ce qui correspond à huit câbles 19T15S. Avec ce renforcement, le déplacement à mi-portée devrait être réduit de -2,44 cm à -0,69 cm sous l'effet du poids propre et des charges mobiles dans l'état limite de service. Par ailleurs, un petit déplacement longitudinal avait été prévu (4,02 mm). Toutefois, aucun déplacement supplémentaire ne devrait se produire au niveau des piliers.

Afin d'apprécier d'une part l'efficacité et la pérennité à court terme du renforcement, et d'évaluer d'autre part l'impact des conditions environnementales (surtout celles liées aux variations thermiques) sur le comportement dynamique du pont, deux campagnes d'acquisition des mesures se sont déroulées sur le long terme : la première phase, entre octobre 2008 et avril 2009, et la deuxième phase, entre octobre 2009 et avril 2010.

Le paragraphe suivant présente les résultats obtenus à partir des procédures d'identification modale menées lors de ces deux campagnes de mesures.

6.2 Identification modale

Préalablement à l'analyse temporelle des signaux, une brève étude des densités spectrales de puissance est proposée afin d'estimer grossièrement les fréquences excitées pendant les essais. La Figure 6.6 présente l'allure temporelle des signaux et la densité spectrale de l'accéléromètre 13.

Les fréquences indiquées par l'algorithme de la transformée de Fourier sont essentiellement celles qui ont été identifiées par la boite à outils Sys-Ident. Contrairement à l'identification modale réalisée pour le pont-rail PK 075-317, l'approche utilisée pour la détermination des paramètres modaux du pont sur l'Oise considère le couplage entre la méthode des Réalisations Stochastiques (MRS) et la méthode du Décrément Aléatoire (MDA).

Cette approche est nécessaire du fait du caractère particulier de la campagne de mesures. De façon générale, la plupart des procédures d'identification modale demandent une détermination préalable des bandes de fréquences pour qu'une procédure préliminaire de filtrage des mesures soit menée. Cette démarche est souvent utilisée pour des campagnes d'essais sont réalisées pendant quelques jours (ou semaines) pour lesquelles le suivi de la variation des fréquences propres peut être facilement effectué. Toutefois, pour des campagnes de mesures avec un grand nombre d'essais enregistrés, le suivi du comportement dynamique de l'ouvrage devient plus délicat. En effet, la définition des bandes de filtrage inadéquates peut non seulement conduire à la non détection de petits décalages fréquentiels liés à des effets extérieurs (température, humidité, vent, etc.), mais aussi à une identification erronée des paramètres modaux faute de déterminer précisément leurs plages de variation.

Compte tenu de cette problématique, une approche originale basée sur le couplage des mé-

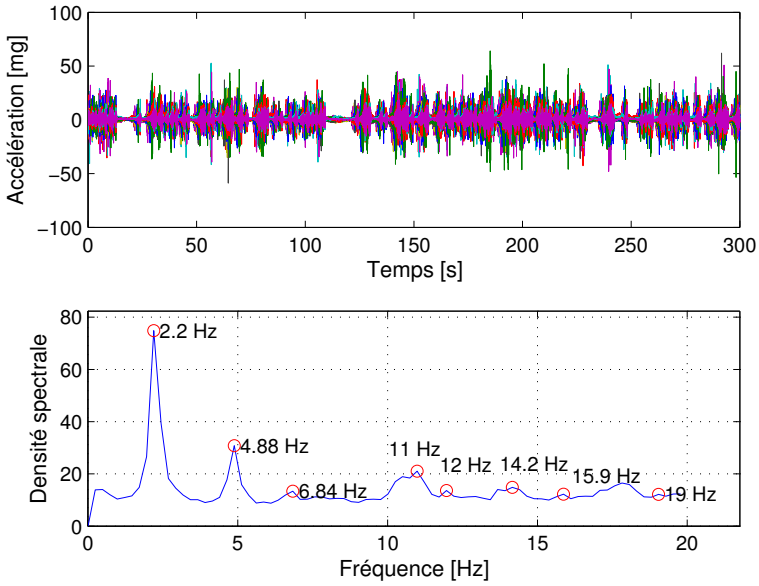


FIGURE 6.6 – Fréquences propres identifiées à partir de la densité spectrale de l'accéléromètre 13.

thodes MRS et MDA a été développée dans le cadre de cette thèse. La démarche consiste, dans un premier temps, à identifier les fréquences propres de manière « grossière » à partir de la MRS. Cette méthode est appliquée seulement au capteur de référence (accéléromètre 13) afin de permettre l'identification automatique des fenêtres spectrales autour de chaque mode. Le point de référence est le point de mesure pour lequel la normalisation des modes propres est réalisée : son amplitude modale est fixée à l'unité. Ce point est choisi de telle sorte que les modes propres les plus intéressants pour l'analyse puissent être identifiés à partir des réponses dynamiques. Bien que la MRS puisse être appliquée pour identifier l'ensemble des caractéristiques modales, cette technique peut devenir très coûteuse en termes de temps de calcul. En effet, la MRS nécessite de calculer les matrices de corrélation, ce qui devient un problème lorsque la taille des matrices augmente significativement. En outre, le calage de certains paramètres de la méthode (tel que l'ordre du modèle, par exemple) peut devenir prohibitif en termes de temps de calcul. Toutefois, la stabilité et la qualité des résultats constituent un avantage certain pour la MDA dans la mesure où un nombre limité de voies de mesure est considéré (dans cette étude, uniquement le capteur de référence). La MRS permet d'obtenir des diagrammes de stabilisation pour chacune des fréquences propres identifiées. À partir des diagrammes, il est possible de déterminer une valeur moyenne pour chaque fréquence propre et d'en définir des bandes de filtrage ($\pm$ 10%, dans cette étude). Dans un deuxième temps, les fonctions de décrément, pour chaque voie de mesure filtrée, sont estimées. À partir de ces fonctions, les paramètres modaux (fréquences propres, coefficients d'amortissement et déformées modales) de la structure sont évalués en utilisant la technique du décrément logarithmique. Dans cette analyse, chaque pas du décrément est considéré comme une réalisation. Dans cette étude, 10 réalisations sont considérées pour chaque essai. En ce qui concerne les déformées modales, les réalisations ayant un coefficient MAC inférieur à 0,95 comparé au mode propre moyen sont éliminées.

Ce procédé permet donc d'utiliser un schéma d'identification automatique tenant compte des éventuelles variations des paramètres modaux au cours du temps. Un bref aperçu du développement mathématique de ces deux méthodes est présenté dans l'Annexe D.

Le Tableau 6.1 regroupe les valeurs moyennes et les écarts-types respectifs des fréquences

propres et des coefficients d'amortissement identifiés lors de la première campagne de mesures. Dans cette phase, 1174 fichiers ont été enregistrés.

Mode	$\overline{f}$ [Hz]	σ_f [Hz]	$\overline{\xi}$ (%)	σ_ξ (%)
1	2,23	0,053	2,27	0,033
2	4,89	0,173	1,65	0,029
3	6,84	0,111	3,05	0,040
4	8,48	0,173	2,74	0,042
5	11,00	0,165	3,10	0,041
6	14,05	0,198	3,16	0,035
7	16,69	0,213	3,14	0,062

TABLEAU 6.1 – Valeurs moyennes et écarts-types des fréquences propres et amortissements (avant renforcement).

La procédure d'identification modale réalisée pendant la première phase (avant renforcement) est également appliquée aux mesures de la deuxième phase (après renforcement), le nombre de capteurs et le plan d'instrumentation restant inchangés. Le capteur de référence pour la normalisation des modes propres est toujours l'accéléromètre 13. Pendant cette phase, 1316 fichiers ont été enregistrés. Bien que la période d'acquisition des mesures soit la même pour les deux phases, quelques problèmes techniques lors de la première campagne ont empêché l'enregistrement de certains fichiers. Le Tableau 6.2 regroupe les valeurs moyennes et les écarts-types respectifs des fréquences propres et des coefficients d'amortissement identifiés lors de la deuxième campagne de mesures.

Mode	$\overline{f}$ [Hz]	σ_f [Hz]	$\overline{\xi}$ (%)	σ_ξ (%)
1	2,29	0,067	2,18	0,051
2	4,95	0,088	1,71	0,097
3	6,93	0,161	2,80	0,034
4	8,51	0,135	3,12	0,047
5	11,08	0,176	3,09	0,027
6	14,20	0,203	2,87	0,045
7	16,91	0,187	3,43	0,054

TABLEAU 6.2 – Valeurs moyennes et écarts-types des fréquences propres et amortissements (après renforcement).

En complément, un modèle 3D aux éléments finis a été développé au LCPC sous le logiciel CESAR/LCPC® pour permettre une corrélation qualitative des résultats identifiés avec un comportement de référence. La Figure 6.7 présente les modes propres calculés par le modèle aux éléments finis (colonne de gauche), et ceux identifiés par Sys-Ident lors de la première phase (colonne au milieu) et de la deuxième phase (colonne de droite). Il est possible de constater un bon accord entre les résultats numériques et expérimentaux. Le maillage des capteurs amont n'est pas symétrique car il existe des points non instrumentés. Les amplitudes de ces derniers sont alors fixées conventionnellement à 0 (représentées par des cercles). Enfin, il convient de signaler un comportement non symétrique rive droite/rive gauche mis en évidence nettement en torsion (représenté par les carrés). La zone identifiée présente donc un comportement mécanique anormal caractérisé par une dissymétrie du mode propre. Après analyse, il est apparu que la zone mise en évidence est proche d'une région fortement fissurée.

Le Tableau 6.3 présente le comparatif entre les fréquences propres des deux phases. Il est observé, notamment pour la première fréquence, une augmentation sensible après les travaux de renforcement, ce qui est en accord avec le type de procédure menée. Par ailleurs, la Figure 6.8

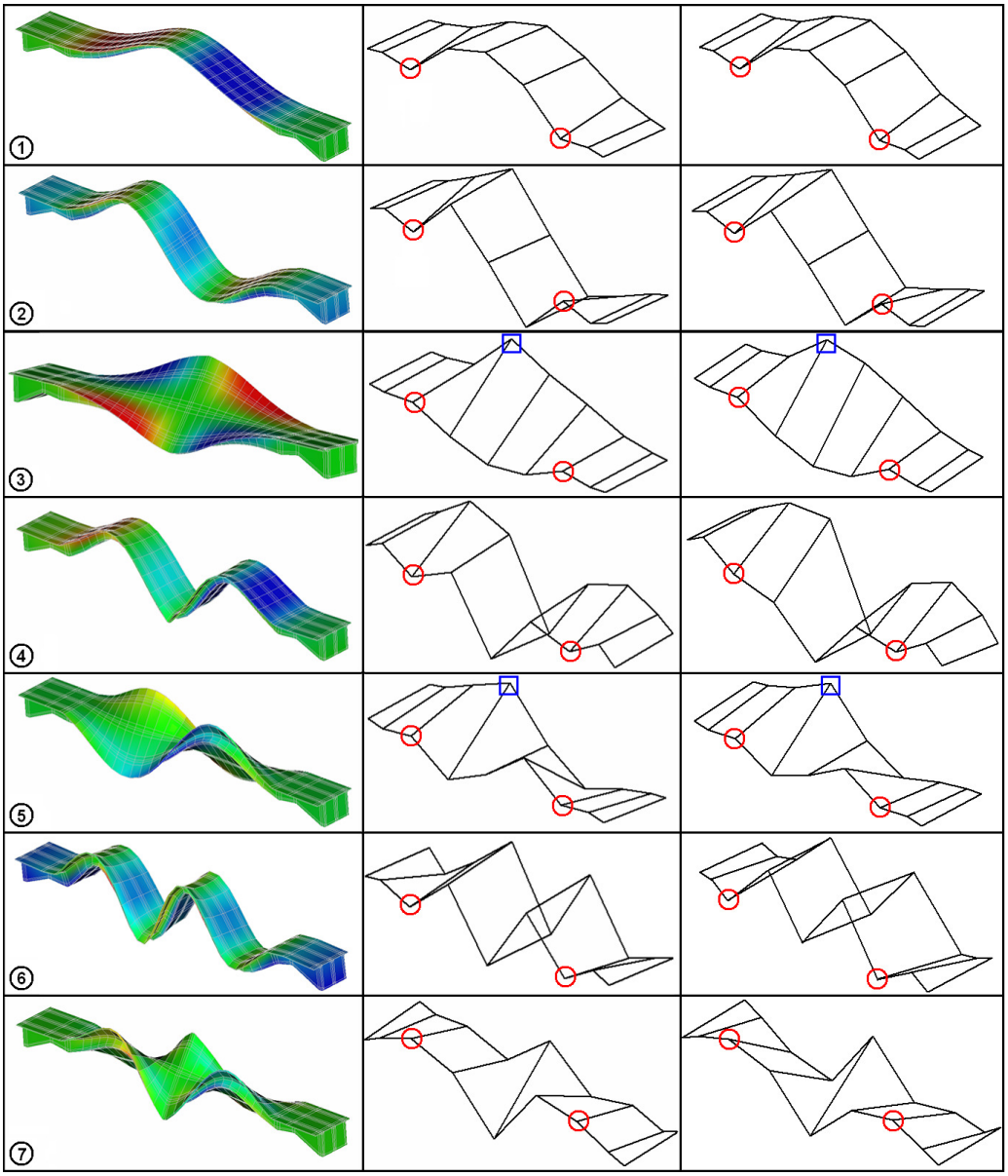


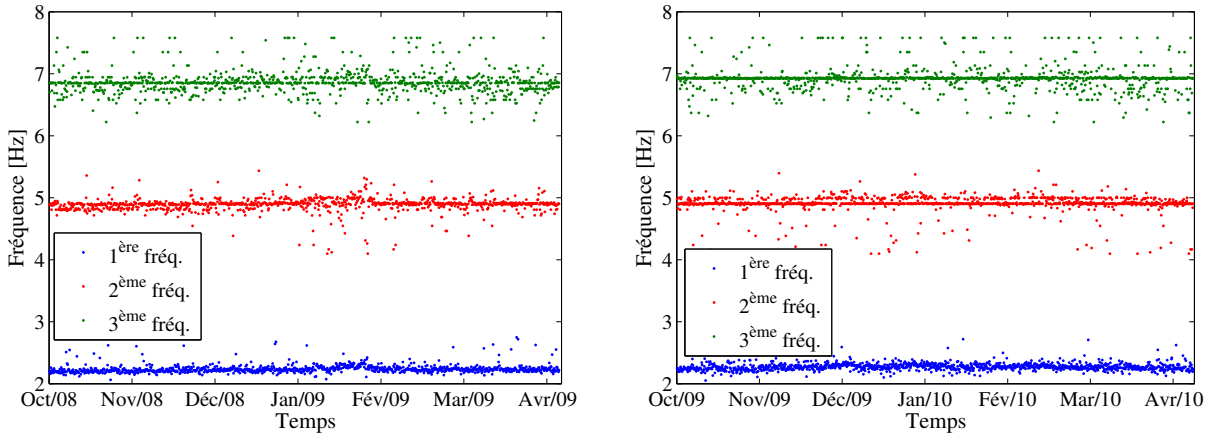
FIGURE 6.7 – Allure des modes propres du pont PI-57 (numérique, phase 1 et phase 2).

illustre la variation des trois premières fréquences, avant et après resserrage, au cours du temps. Il est noté que les fréquences propres identifiées lors de la deuxième phase sont plus élevées que celles de la première phase. Quelques variations sont observées au cours des mois. Ces fluctuations peuvent être liées à des phénomènes extérieurs comme des changements de température, par exemple. Dans les paragraphes suivants, une étude des effets thermiques sur les paramètres modaux est menée.

Comme dans le chapitre précédent, un premier objectif majeur de cette étude est d'utiliser l'ADS couplée aux méthodes de classification non supervisée et supervisée afin de discriminer les deux états structuraux du pont sur l'Oise (i.e., avant et après le renforcement). Un second objectif est d'étudier les effets de la variation des températures sur les paramètres modaux et d'appréhender comment ils peuvent influencer les classifications de comportements structuraux. Pour cela, une analyse basée sur les techniques de régression linéaire et non linéaire et l'ACP est proposée.

Mode	Fréquence [Hz] Phase 1	Fréquence [Hz] Phase 2	Différence (%) Phase 1/Phase2
1	2,23	2,29	2,69
2	4,89	4,95	1,23
3	6,84	6,93	1,32
4	8,48	8,51	0,35
5	11,00	11,08	0,72
6	14,05	14,20	1,11
7	16,69	16,91	1,32

TABLEAU 6.3 – Valeurs moyennes et écarts-types des fréquences propres identifiées avant (phase 1) et après (phase 2) renforcement.



(a) Évolution des fréquences propres dans la phase 1

(b) Évolution des fréquences propres dans la phase 2

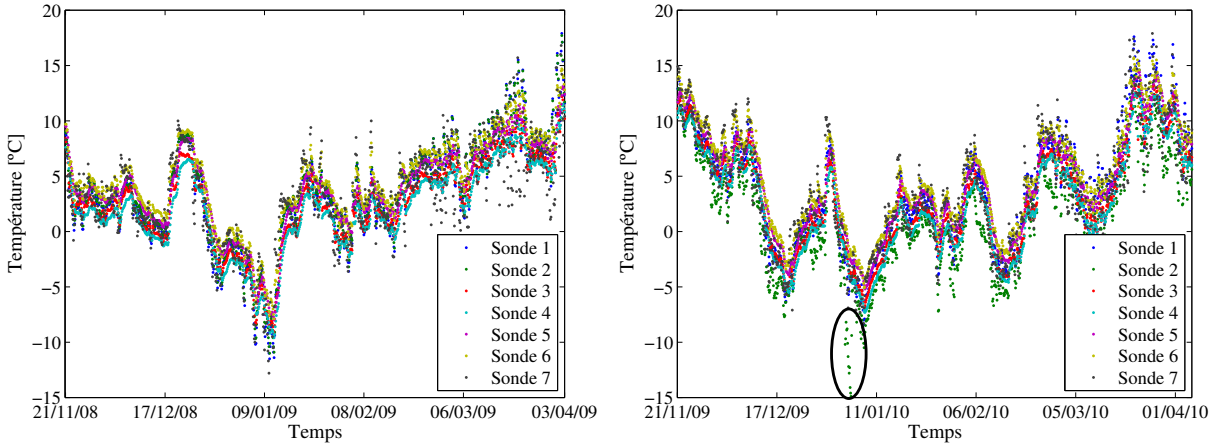
FIGURE 6.8 – Évolution des trois premières fréquences propres en fonction du temps.

6.3 Étude des effets thermiques sur les paramètres modaux

Comme il a été proposé au chapitre 5 avec l'application expérimentale sur le pont-rail, une étude des effets thermiques sur les paramètres modaux est maintenant proposée. Comme décrit au paragraphe 6.1, sept sondes ont été placées à l'intérieur et l'extérieur du caisson du pont afin d'enregistrer les températures durant les deux campagnes de mesures. Dans ce paragraphe, deux études sont proposées. Dans un premier temps, les techniques de régression sont utilisées pour la définition de modèles de prédiction. L'objectif est de détecter des variations anormales sur les fréquences propres (i.e., qui ne sont pas liées aux changements de température) et de vérifier, d'une part, les modifications sur les déformées modales et d'autre part, d'observer comment les essais correspondant aux variations anormales se retrouvent par rapport à ceux considérés « normaux ». Dans un deuxième temps, la procédure de « correction » des paramètres modaux est mise en œuvre. Pour cela, des filtres linéaires et non linéaires sont créés et les méthodes de clustering et de classification supervisée sont appliquées aux fréquences et modes propres « corrigés ». Le but est d'observer comment les taux de classification correcte sont modifiés ou non en considérant cette approche.

La Figure 6.9 illustre la variation des températures enregistrées par les sept sondes au cours des enregistrements des deux campagnes de mesures réalisées. Il est important de remarquer que dans la première phase, il n'y a pas d'enregistrement thermique pendant le mois d'octobre et le début du mois de novembre. Ainsi, les analyses menées par la suite ne considèrent que les essais

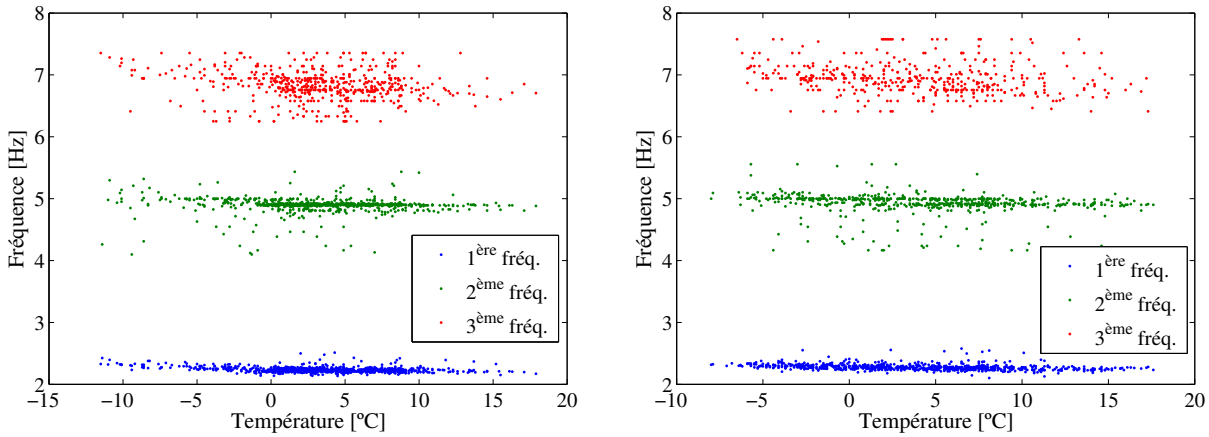
acquis entre le 21 novembre et le 3 avril (pour les deux phases). Pour la première phase, 972 essais sont utilisés, alors que 1164 essais le sont pour la deuxième phase. Il est rappelé que, à l'exception de la sonde 5, toutes les sondes sont placées à l'intérieur du caisson du pont.



(a) Phase 1 (b) Phase 2

FIGURE 6.9 – Variation des températures pendant les phases 1 et 2 d'instrumentation.

Il est noté que la plage de variation des températures pour les deux phases de mesures est comprise entre −15 °C et 20 °C. Toutefois, il est remarqué un écart thermique important entre ces deux campagnes, surtout durant les mois de novembre et janvier de la phase 2 qui ont été relativement plus chauds que ceux de l'année précédente. En analysant la Figure 6.9(b), il est possible d'observer que les températures enregistrées par la sonde 2 au début du mois de janvier ont des valeurs aberrantes par rapport à celles enregistrées par les autres sondes. En effet, il a été constaté un écart supérieur à 10 °C entre leurs enregistrements. La sonde 2 n'est donc pas considérée dans les analyses suivantes. La Figure 6.10 montre la variation des trois premières fréquences propres par rapport à la sonde 1, située à l'intérieur du caisson. Il est possible d'observer que les fréquences propres ont un comportement similaire par rapport à la variation de température. De plus, pour les deux phases d'instrumentation, les fréquences tendent à diminuer à mesure que la température augmente.



(a) Phase 1 (b) Phase 2

FIGURE 6.10 – Variation des trois premières fréquences propres que la sonde 1 (phases 1 et 2).

Cette analyse peut toutefois masquer certaines variations temporelles. En effet, en analysant

de manière simultanée la variation de la première fréquence propre et des températures enregistrées par la sonde T1 au cours du temps (pendant la phase 1), il est remarqué qu'une chute de température de près de 10 °C au mois de janvier a produit, pratiquement instantanément, une augmentation de 10% environ de la fréquence propre (Figure 6.11). Le phénomène a également été mis en évidence avec les autres fréquences, mais dans des proportions moins importantes. Ceci montre que, dans le cadre de cette application, les effets thermiques sur paramètres modaux ne sont pas négligeables.

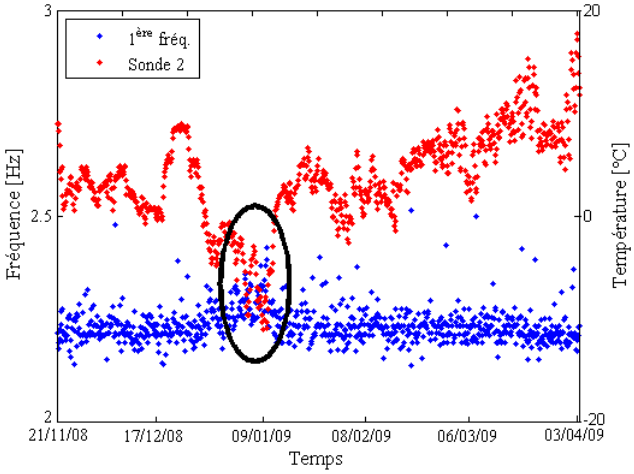


FIGURE 6.11 – Variation de la première fréquence et de la température au cours du temps.

Comme il a été décrit dans le paragraphe 3.4, avant d'appliquer les techniques de régression, les valeurs des fréquences propres et des composantes modales identifiées sont normalisées par rapport à une valeur de référence. Dans cette étude, la température 0 °C mesurée par la sonde T1 est utilisée comme température de référence, étant donné que cette valeur correspond à la moyenne de la plage de variations des températures enregistrées dans chaque campagne. Ainsi, les fréquences sont toujours normalisées par rapport à la valeur identifiée à cette température (la même démarche est appliquée pour chaque composante modale. Cette procédure est apparue la plus adéquate, en obtenant un meilleur ajustement des données analysées (Cury & Crémona 2010*d*). Il est noté que toutes les sondes de températures sont prises en compte pour la détermination des lois de correction (à l'exception des sondes T2 et T5). Dans les analyses suivantes, les températures mesurées et les ratios des paramètres modaux identifiés lors de la première phase de mesures (avant renforcement) sont utilisés pour la construction des lois de correction.

Le Tableau 6.4 regroupe les vecteurs des poids évalués pour chaque ratio des fréquences propres en utilisant des filtres linéaires. Dans ce tableau, β est le biais du modèle, $w_{T1}, \ldots, w_{T7}$ sont les poids qui multiplient les enregistrements de chaque sonde de température, respectivement.

	Fréquence 1	Fréquence 2	Fréquence 3
β	1,0062	1,0038	0,9944
w_{T1}	-0,0012	0,0007	-0,0006
w_{T3}	0,0024	0,0033	-0,0068
w_{T4}	-0,0017	-0,0012	-0,0004
w_{T6}	0,0005	-0,0027	0,0080
w_{T7}	-0,0010	-0,0005	-0,0008

TABLEAU 6.4 – Vecteurs des poids évalués pour chaque loi de correction thermique (régression linéaire ; phase 1).

Les lois de correction basées sur la régression non linéaire sont déterminées en utilisant des réseaux de neurones. L'architecture utilisée est composée de 10 neurones dans la couche cachée et les fonctions d'activation utilisées sont les fonctions tangente hyperbolique (dans la couche cachée) et sigmoïde (dans la couche de sortie). Les données d'entrée et de sortie sont les enregistrements de température des sondes thermiques et les ratios des paramètres modaux, respectivement. L'entraînement du réseau est réalisé en utilisant 680 essais de la phase 1 et pour la validation, 292 (également de la phase 1). Ainsi, 1000 tirages aléatoires sont réalisés (pour différents groupes d'entraînement et de validation). Le modèle choisi est celui dont l'erreur commise lors de phase de validation est la plus faible. Étant donné que le nombre de poids du réseau calibré est conséquent (50 dans la couche cachée + 10 dans la couche de sorties + 11 biais), les valeurs ne sont pas données.

Vraisemblablement, deux lois de correction indépendantes peuvent être créées. La première, utilisée pour corriger les paramètres modaux de la phase 1 et la deuxième, pour compenser les effets des paramètres modaux de la phase 2. Toutefois, en normalisant toutes les fréquences propres et composantes modales par rapport à une valeur de référence, cette démarche peut être évitée. En réalité, il faut que la loi de correction définie à partir des observations de la phase 1 soit pertinente vis-à-vis des observations de la phase 2. Ainsi, il convient de « tester » les modèles de correction avant de les mettre en œuvre. Dans cette étude, huit essais ont été enregistrés pendant la phase 2 à la température de 0 °C. L'idée est alors de comparer les prédictions (pour les huit essais) déterminées à partir de la loi de la phase 1 (cf. Tableau 6.4) avec celles obtenues à partir de la loi de la phase 2 (cf. Tableau 6.5).

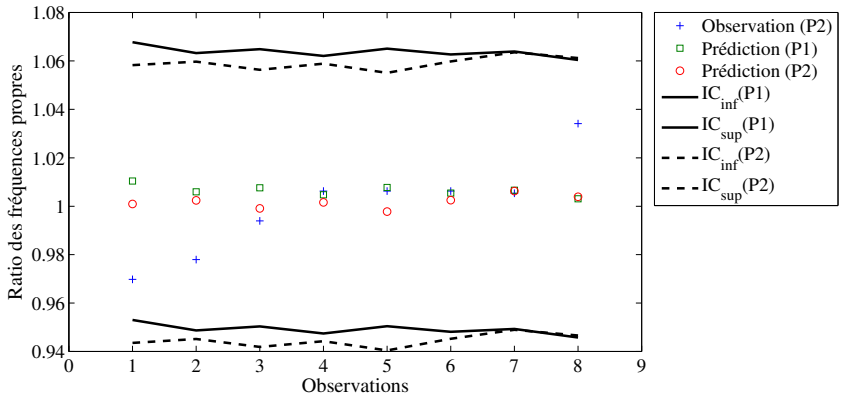
	Fréquence 1	Fréquence 2	Fréquence 3
β	0.9973	0.9926	0.9881
w_{T1}	-0.0025	-0.0004	-0.0015
w_{T3}	0.0048	-0.0100	-0.0049
w_{T4}	-0.0044	0.0025	0.0022
w_{T6}	0.0013	0.0082	0.0038
w_{T7}	0.0008	-0.0005	0.0006

TABLEAU 6.5 – Vecteurs des poids évalués pour chaque loi de correction thermique (régression linéaire ; phase 2).

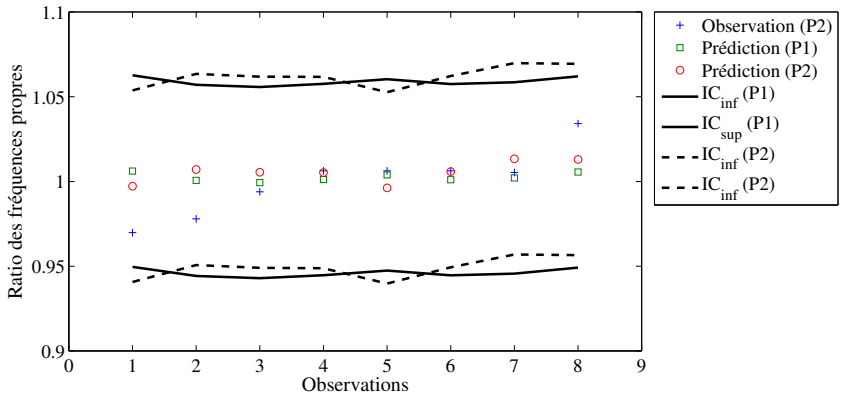
La Figure 6.12(a) illustre un comparatif des prédictions réalisées pour les ratios des premières fréquences propres identifiées lors de la phase 2, à la température de référence. Il est noté que les prédictions des deux modèles linéaires sont significativement proches situées dans les bandes de confiance. Une analyse similaire est menée en considérant le modèle non linéaire (cf. Figure 6.12(b)). De façon similaire à ce qui a été observé précédemment, les prédictions sont assez proches et se retrouvent toutes à l'intérieur des intervalles de confiance évalués.

Ces analyses sont également menées en considérant les deuxième et troisième fréquences propres. Des résultats identiques sont observés (cf. Figures C.1 et C.2). Étant donné que l'utilisation d'une seule loi de correction peut être considérée de manière pertinente, les simulations suivantes ne considèrent que les températures mesurées et les ratios des paramètres modaux identifiés lors de la première phase de mesures (avant renforcement) pour la construction des lois de correction.

La Figure 6.13 illustre l'évolution des trois premières fréquences propres corrigées par le filtre linéaire par rapport au temps. Il est difficile d'apprécier les différences entre ces fréquences propres et celles sans correction thermiques à partir d'une observation simple sur leurs évolutions. Dans les paragraphes suivantes, des analyses classiques et symboliques sont menées afin de mieux distinguer les essais des phases 1 et 2. Dans l'Annexe C, l'évolution des fréquences propres



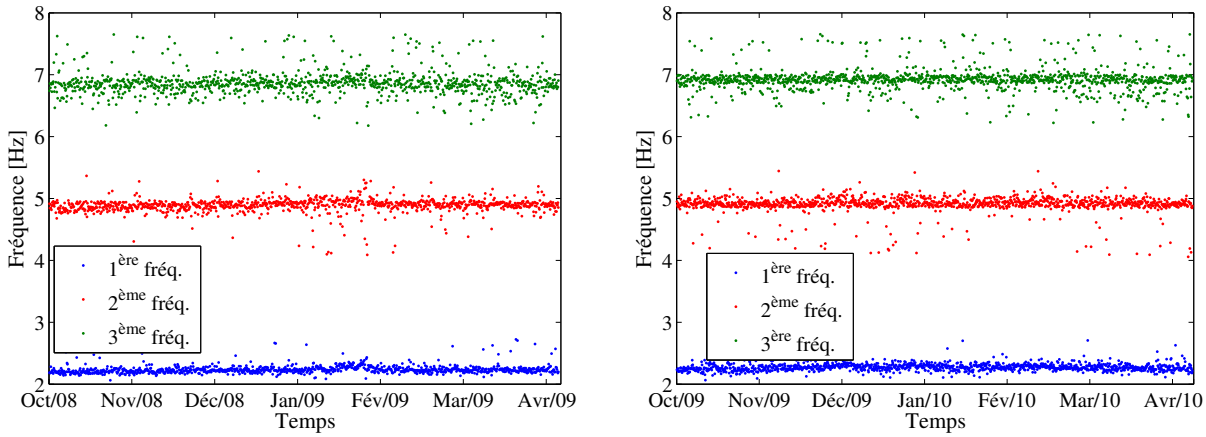
(a) Comparatif des prédictions (modèles linéaires)



(b) Comparatif des prédictions (modèles non linéaires)

FIGURE 6.12 – Comparatif des prédictions pour la première fréquence propre évaluées par des modèles de régression linéaires et non linéaires.

corrigées par le filtre non linéaire est montrée.



(a) Évolution des fréquences propres dans la phase 1 (b) Évolution des fréquences propres dans la phase 2

FIGURE 6.13 – Évolution des trois premières fréquences propres en fonction du temps (régression linéaire).

6.3.1 Étude des modèles de prédiction

Dans ce paragraphe, une approche originale basée sur des modèles de régression et l'ACP (Analyse en Composantes Principales) couplés à l'ADS est proposée. L'objectif principal est de pouvoir détecter des modifications qui peuvent représenter un comportement anormal de l'ouvrage, en éliminant les effets liés aux variations de température. Dans un premier temps, les techniques de régression linéaire et non linéaire sont considérées afin de construire des lois qui expliquent la variation des fréquences propres par rapport aux températures.

Cette démarche est mise en œuvre en deux temps. Dans un premier temps, les fréquences propres identifiées et les températures mesurées lors de la première phase d'instrumentation sont utilisées pour l'entraînement des modèles de régression. Dans ce cas, l'ajustement des lois (linéaires ou non linéaires) est réalisé en considérant toutes les sondes de températures, sauf les sondes T2, pour des raisons citées au paragraphe précédent, et T5, qui se situe en dehors du caisson où les accéléromètres sont placés. Une fois le modèle ajusté, les poids optimaux sont évalués (comme détaillé dans le paragraphe 3.4) et appliqués aux températures mesurées pendant la deuxième phase d'instrumentation. Cette procédure mène alors à des valeurs prédites pour les fréquences propres identifiées pendant la deuxième campagne de mesures. Par ailleurs, des intervalles de confiance autour des valeurs prédites sont déterminés en utilisant les températures de la phase 1 (cf. Équation (3.72)). En théorie, toutes les observations situées en dehors de cet intervalle devraient être considérées comme non représentatives de la « respiration naturelle » de l'ouvrage et donc, transcrire un possible comportement « anormal » de l'ouvrage. Toutefois, le comportement décrit ici comme « anormal » est en réalité causé par les travaux de renforcement réalisés. Ainsi, toutes les observations qui se retrouvent en dehors de la bande de confiance sont celles pour lesquelles les effets du renforcement sont marqués. Théoriquement, si le modèle de régression est bien ajusté, c'est-à-dire, s'il arrive à caractériser la variation des fréquences propres uniquement en fonction de la température, toutes les observations prédites devraient être caractérisées comme « anormales ». Si ceci n'est pas le cas, alors l'effet de renforcement est couvert par l'effet thermique et ne peut être distingué. Il faut aussi considérer la possibilité d'erreurs de détection. En effet, il se peut qu'une observation considérée normale (à l'intérieur de la bande de confiance) par les modèles de régression, ne représente pas nécessairement la même condition structurale, mais un événement parfois causé par le trafic, l'humidité, le vent, etc. ou encore une erreur de prédiction du modèle. La deuxième étape de l'approche proposée consiste alors à utiliser l'analyse d'affectation des essais à un cluster donné.

L'objectif est de détecter des essais qui ont changé de comportement après les travaux, mais qui n'ont pas été identifiés par les modèles de régression. La démarche utilisée ici est basée sur le calcul des distances entre les essais qui se situent à l'intérieur de la bande de confiance calculée et le prototype (noyau) du cluster contenant les essais de la phase 1. Si cette distance est plus importante que la distance seuil (évaluée préalablement), cet essai est considéré comme une « anomalie » (le modèle de régression n'est donc pas suffisant pour détecter un changement structural). Ceci indique donc une erreur de détection du modèle de régression utilisé. À l'inverse, si la distance est plus petite que le seuil, l'essai est considéré comme un comportement similaire à celui de l'état structural avant renforcement. Dans ce cas, la variation fréquentielle causée par les travaux ne peut pas être détectée par cette approche.

Par ailleurs, une démarche alternative utilisant l'ACP est proposée. L'avantage de l'ACP est qu'elle ne nécessite pas de mesure des températures. En réalité, les effets thermiques sont traités comme des variables inhérentes aux variations fréquentielles. L'idée est d'utiliser les fréquences propres identifiées lors des deux campagnes de mesures et, à partir du calcul des bandes de confiance, de déterminer les observations dites « anormales ». Théoriquement, les essais enregistrés pendant la phase 1 devraient être situés à l'intérieur des bandes de confiance. À l'inverse, les

essais acquis pendant la phase 2 (après travaux de renforcement) devraient se situer en dehors des intervalles de confiance. Le couplage avec l'approche d'affectation des essais est également proposé. Le but est de vérifier si les « anomalies » signalées par l'ACP représentent une modification structurale ou une fausse alarme.

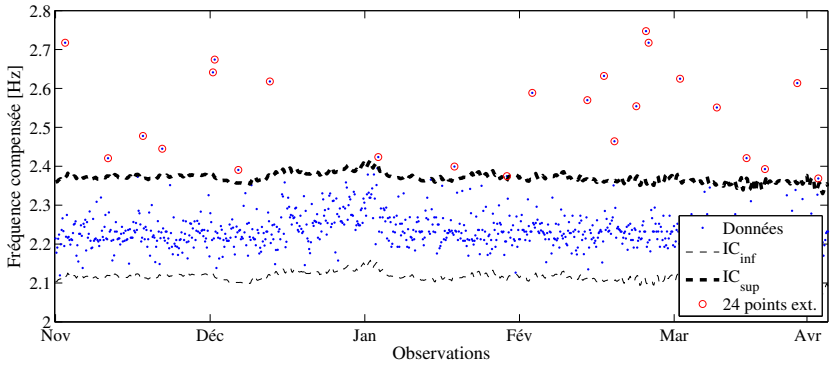
Comme il a été décrit précédemment, avant d'appliquer les techniques de régression, les valeurs des fréquences propres sont normalisées par rapport à une valeur de référence. La température 0 °C mesurée par sonde la T1 est utilisée comme température de référence. Ainsi, les fréquences propres sont normalisées par rapport à la valeur identifiée à cette température. Il est important de remarquer que toutes les sondes de températures sont prises en compte pour la détermination des lois de correction (à l'exception des sondes T2 et T5).

6.3.1.1 Modèles de régression linéaire

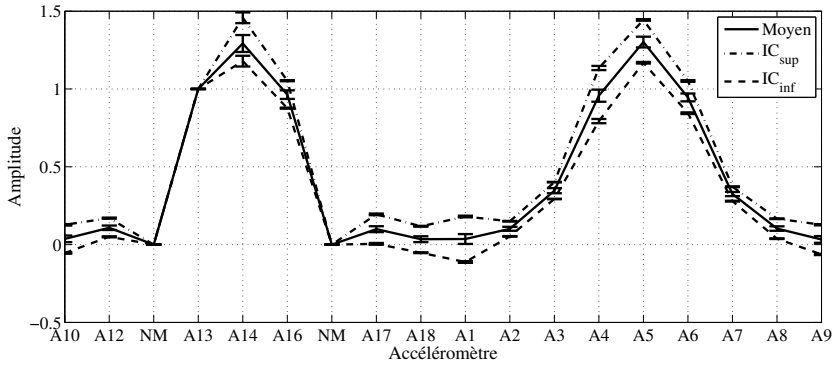
Les premières simulations correspondent aux analyses de régression appliquées à la première fréquence propre. Dans ce cas, l'ensemble des températures mesurées et des ratios des premières fréquences propres identifiées pendant la première phase d'instrumentation est utilisé pour l'ajustement des modèles de régression. Manifestement, ces fréquences propres peuvent également être prédites par le modèle calibré, ce qui est généralement appelé « simulation » du modèle. Cette procédure est souvent menée afin de vérifier la qualité de l'ajustement réalisé. La Figure 6.14(a) illustre les résultats de la simulation avec la première fréquence propre. Il est noté que la plupart des valeurs prédites se situent à l'intérieur de l'intervalle de confiance. Cela montre que le modèle est relativement bien ajusté par rapport aux données utilisées. La Figure 6.14(b) illustre les résultats obtenus par le modèle de régression appliqué aux ratios des composantes modales du premier mode propre. Dans cette figure, les valeurs moyennes (± 1 écart-type) pour chaque composante modale située à l'intérieur de l'intervalle de confiance sont montrées. Par ailleurs, les valeurs moyennes des bandes de confiance supérieure et inférieure (± 1 écart-type) pour chaque composante sont également illustrées. Il est noté que les composantes du mode moyen se situent généralement à l'intérieur des bandes de confiance.

Dans les analyses suivantes, l'ensemble des enregistrements de température de la deuxième phase est considéré dans l'étape de prédiction. La Figure 6.15(a) montre les résultats obtenus pour la prédiction de la première fréquence propre en utilisant le modèle linéaire. Les intervalles de confiance sont également montrés. Dans cette simulation, il est noté que seuls 28 essais sont considérés comme « anormaux » d'après le modèle de régression linéaire. En d'autres termes, le modèle de régression a pu différencier les effets des travaux de renforcement de ceux liés aux effets thermiques dans 28 essais. Par ailleurs, la prédiction des composantes des déformées modales sont présentées à la Figure 6.15(b). Dans cette figure, le mode moyen (± 1 écart-type) les moyennes des bandes de confiance (± 1 écart-type) sont également illustrés.

Afin de mieux analyser les essais dits normaux, l'approche d'affectation d'essais basée sur le calcul des distances entre un essai donné et un prototype est utilisée (détaillé dans le paragraphe 3.1.6.1). L'idée est de repérer les essais qui se situent à l'intérieur des intervalles de confiance et de les affecter au cluster d'essais de la phase 1. Si cette distance est plus importante que la distance seuil (évaluée préalablement), cet essai est considéré comme une « anomalie » (donc le modèle de régression n'est pas suffisant pour détecter un changement structural). Dans cette première analyse, il y a 1136 (1164-28) essais à affecter et les 972 essais de la phase 1 sont considérés dans le cluster C(P1). Le prototype pour ce cluster est évalué, une loi de probabilités est ajustée aux distances entre essais intra-cluster, et finalement la distance seuil est définie. Étant donné que le modèle de régression est appliqué à la première fréquence propre seulement, cette procédure ne considère que cette fréquence (et la déformée modale associée) transformée(s) en données symboliques (intervalles interquartiles et en histogrammes). Dans le cas des fréquences propres, la



(a) Modèle linéaire appliqué à la première fréquence propre (simulation)



(b) Modèle linéaire appliqué au premier mode propre (simulation)

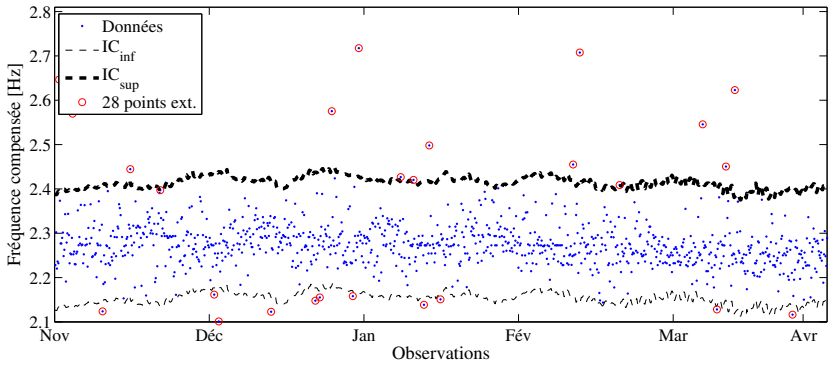
FIGURE 6.14 – Modèle de régression linéaire appliqué aux paramètres modaux (simulation).

loi ajustée est la loi log-normale (seuil égal à 0,894). Le Tableau 6.6 regroupe les taux d'affectation calculés. Quel que soit le paramètre modal considéré ou la transformation symbolique utilisée, les essais ne sont pas affectés à la phase 1. Dans ce cas, le modèle de régression n'est pas suffisant pour mettre en évidence une modification structurale d'un effet thermique. Par ailleurs, quelques essais sont classés dans le cluster correspondant à la phase 1. Cela indique que ni le modèle de régression, ni l'approche d'affectation ne sont capables de détecter les effets des travaux menés. Les taux d'affectation évalués pour les déformées modales sont pratiquement identiques à ceux obtenus en utilisant les fréquences propres. Il est intéressant de remarquer que l'affectation n'est pas nécessairement la même pour les fréquences propres et les modes propres. En effet, cette procédure dépend fortement de l'ajustement des lois de probabilité et de la variabilité de chaque paramètre modal. Dans le cas des modes propres, la loi ajustée est la loi log-normale (seuil égal 0,771).

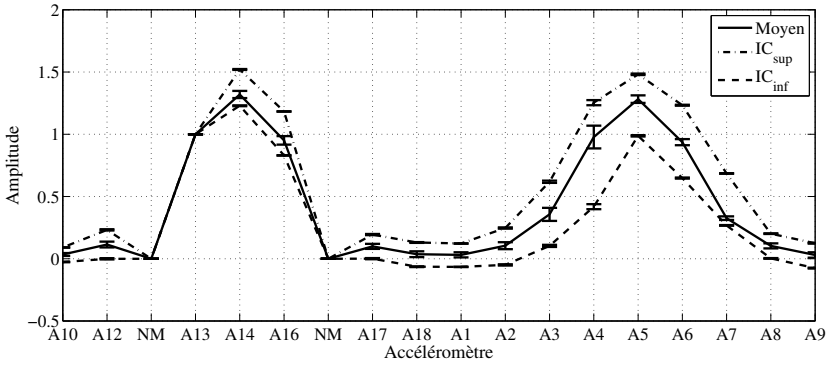
Cluster	**Fréquences propres**	**Modes propres**
C(P1)	35%	38%
Non classé	65%	62%

TABLEAU 6.6 – Affectation des essais en utilisant la première fréquence/mode propre transformés en histogrammes (modèle linéaire).

Un modèle de régression linéaire est maintenant appliqué à la deuxième fréquence propre. Dans un premier temps, les enregistrements de température et les fréquences propres identifiées durant la première phase sont utilisés pour calibrer le modèle de régression. Les résultats sont montrés à la Figure 6.16(a). Dans le cas des simulations avec les composantes des déformées



(a) Modèle linéaire appliqué à la première fréquence propre (prédiction)



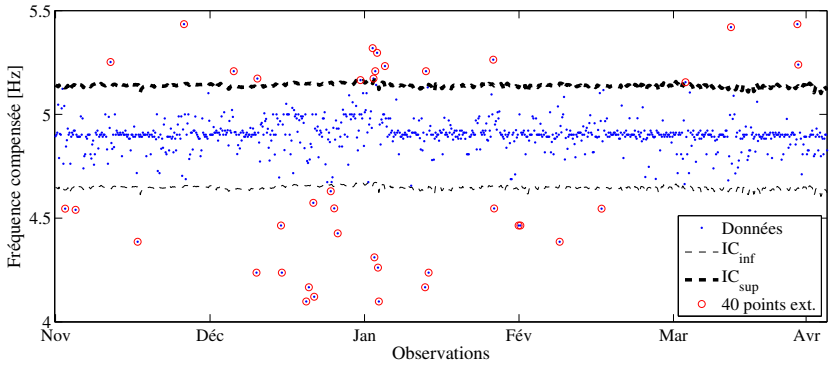
(b) Modèle linéaire appliqué au premier mode propre (prédiction)

FIGURE 6.15 – Modèle de régression linéaire appliqué aux paramètres modaux (prédiction).

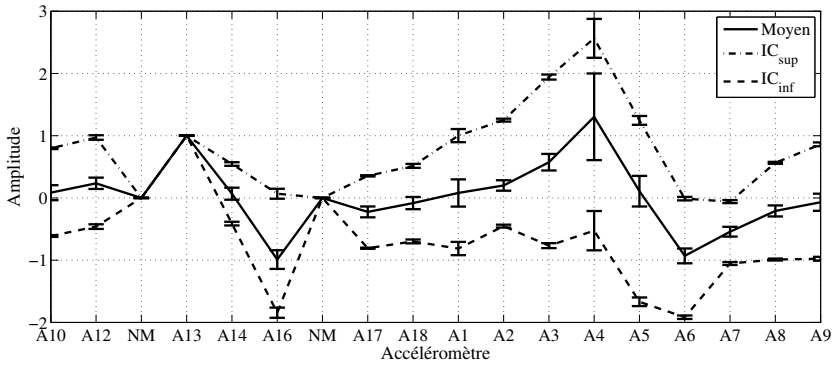
modales, les résultats sont illustrées à la Figure 6.16(b). Comme il a été observé avec le premier mode propre (cf. Figure 6.15(b)), la plupart des composantes modales prédites par le modèle de régression se situent à l'intérieur des bandes de confiance.

Dans un second temps, les températures mesurées lors de la deuxième phase sont considérées pour la prédiction de la deuxième fréquence propre. Les résultats obtenus sont montrés à la Figure 6.17(a). Dans ce cas, 142 essais se situent en dehors des intervalles de confiance. Cela indique que le modèle de régression construit pour cette fréquence propre peut, dans un certain sens, mieux discriminer les effets des travaux de renforcement de ceux dus à la température. La Figure 6.17(b) illustre la déformée modale moyenne correspondante aux observations situées à l'intérieur de l'intervalle de confiance. Les bandes de confiance sont également montrées.

L'approche pour l'affectation des essais est maintenant considérée. Le prototype du cluster C(P1) est évalué et les distances sont calculées en considérant les 1022 (1164−142) essais compris dans la bande de confiance, représentés par les paramètres modaux transformés en données symboliques. Étant donné que le modèle de régression ne prend en compte que la deuxième fréquence propre, seul ce paramètre modal (et le mode propre associé) sont considérés dans le calcul d'affectation. Les résultats sont détaillées dans le Tableau 6.7. En considérant les fréquences propres transformées en intervalles interquartiles, un peu plus de la moitié des essais n'est pas classé en phase 1. Cela indique que le modèle de régression construit pour la deuxième fréquence propre peut ne pas mettre en évidence la modification structurale. En général, le couplage entre les modèles de régression et l'approche d'affectation ne permet pas de différencier ces effets. La loi ajustée est la loi log-logistique (seuil égal à 0,778). Les résultats sont pratiquement du même ordre en considérant les déformées modales que ceux obtenus avec les fréquences propres. Dans ce cas, la loi log-logistique est ajustée aux distances entre essais (seuil égal à 0,507).



(a) Modèle linéaire appliqué à la deuxième fréquence propre (simulation)



(b) Modèle linéaire appliqué au deuxième mode propre (simulation)

FIGURE 6.16 – Modèle de régression linéaire appliqué aux paramètres modaux (simulation).

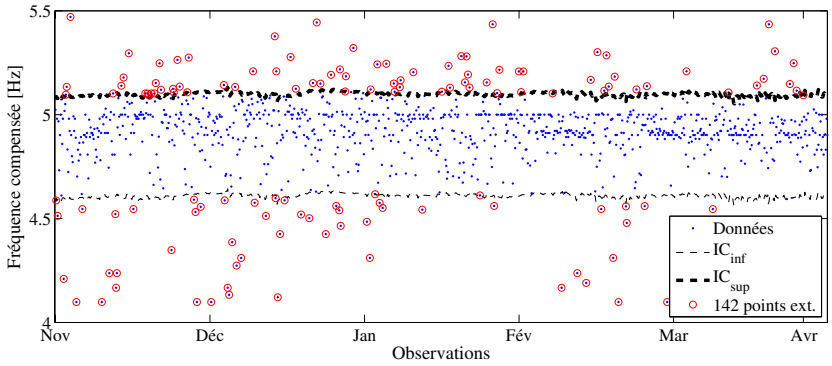
Cluster	**Fréquences propres**	**Modes propres**
C(P1)	48%	43%
Non classé	52%	57%

TABLEAU 6.7 – Affectation des essais en utilisant la deuxième fréquence/mode propre transformés en histogrammes (modèle linéaire

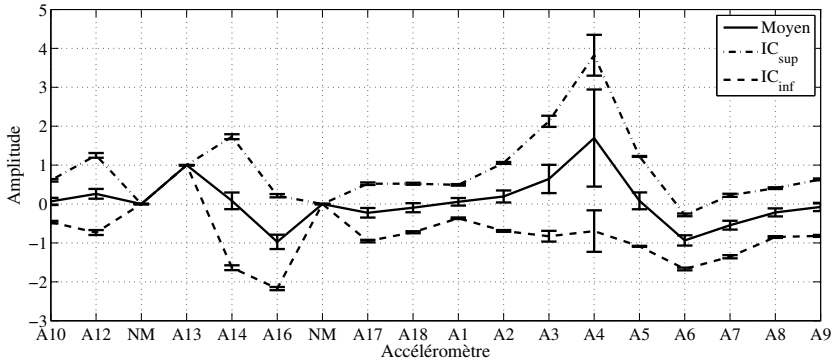
Un modèle de régression linéaire est construit en considérant la troisième fréquence propre. Le modèle est d'abord calibré en tenant compte des mesures de température et des fréquences propres identifiées lors de la première phase d'instrumentation (cf. Figure 6.18(a)). Les ratios des composantes modales sont ensuite considérés et un modèle de régression est construit. La déformée modale moyenne et les bandes de confiance sont montrées dans la Figure 6.18(b) avec les écarts-types respectifs (pour chaque composante modale).

Une fois le modèle calibré, les enregistrements de température de la deuxième phase sont utilisés pour la prédiction des fréquences propres. La Figure 6.19(a) illustre les résultats obtenus. Dans cette analyse, 128 observations se situent à l'extérieur des bandes de confiance (un nombre moins important par rapport à l'analyse précédente, mais supérieur à celui évalué en utilisant la première fréquence propre). Les déformées modales sont illustrées à la Figure 6.19(b).

Les essais qui se situent à l'intérieur des bandes de confiance sont maintenant analysés. Pour cela, l'approche d'affectation est mise en œuvre. Dans un premier temps, les distances entre les 1036 (1164 − 128) essais compris dans les intervalles de confiance et le prototype du cluster C(P1) sont évaluées. Les résultats sont détaillées dans le Tableau 6.8. Comme il a été observé pour la deuxième paire des paramètres modaux, pratiquement la moitié des essais n'est pas classé



(a) Modèle linéaire appliqué à la deuxième fréquence propre (prédiction)



(b) Modèle linéaire appliqué au deuxième mode propre (prédiction)

FIGURE 6.17 – Modèle de régression linéaire appliqué aux paramètres modaux (prédiction).

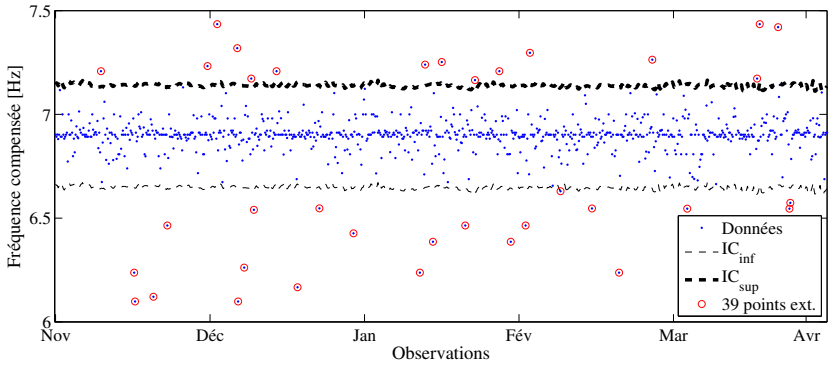
au cluster de la phase 1. Cela confirme que le modèle de régression linéaire n'est pas capable de discriminer correctement une modification structurale des effets thermiques. La loi ajustée dans le cas de la transformation en histogrammes est la loi log-normale (seuil égal à 0,624). Les résultats sont pratiquement du même ordre en considérant les déformées modales. Dans ce cas, la loi log-logistique est ajustée aux distances entre essais (seuil égal à 0,788).

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	55%	56%
Non classé	45%	44%

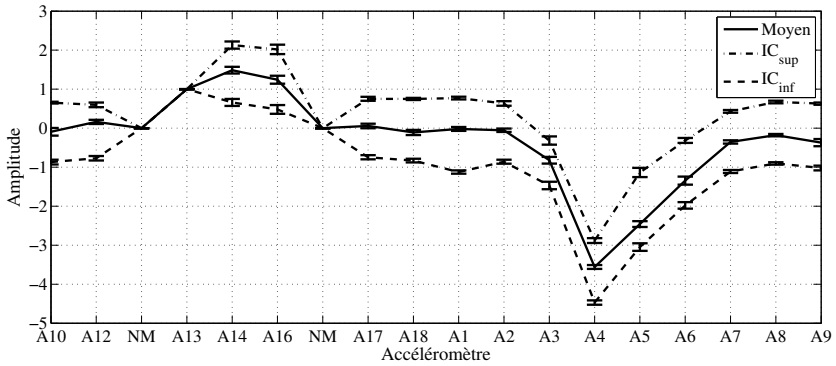
TABLEAU 6.8 – Affectation des essais en utilisant la troisième fréquence/mode propre transformés en histogrammes (modèle linéaire).

6.3.1.2 Modèles de régression non linéaire

Une démarche similaire est proposée en considérant des modèles de régression non linéaire. En premier lieu, un modèle est évalué pour l'ensemble des premières fréquences propres identifiées dans la phase 1. L'architecture utilisée est composée de 10 neurones dans la couche cachée et les fonctions d'activation utilisées sont la fonction tangente hyperbolique (dans la couche cachée) et la fonction sigmoïde (dans la couche de sortie). Les données d'entrée sont les enregistrements de température des sondes thermiques et les sorties, les ratios des paramètres modaux. L'entraînement du réseau est réalisé en utilisant 680 essais de la phase 1 et 292 pour la validation (aussi de la phase 1). Ainsi, 1000 tirages aléatoires sont réalisés (pour différents groupes d'entraînement



(a) Modèle linéaire appliqué à la troisième fréquence propre (simulation)



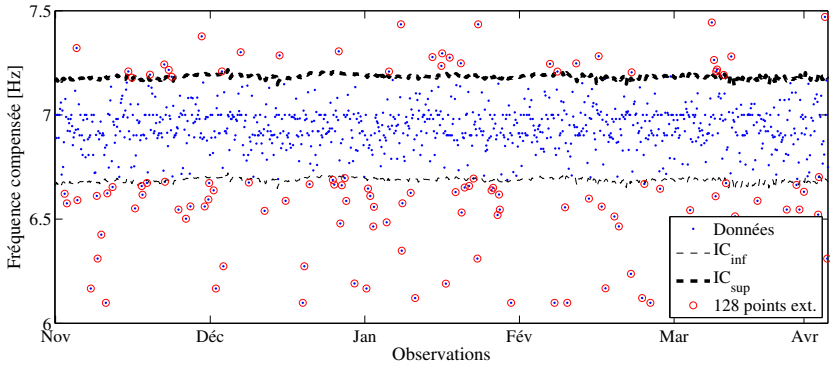
(b) Modèle linéaire appliqué au troisième mode propre (simulation)

FIGURE 6.18 – Modèle de régression linéaire appliqué aux paramètres modaux (simulation).

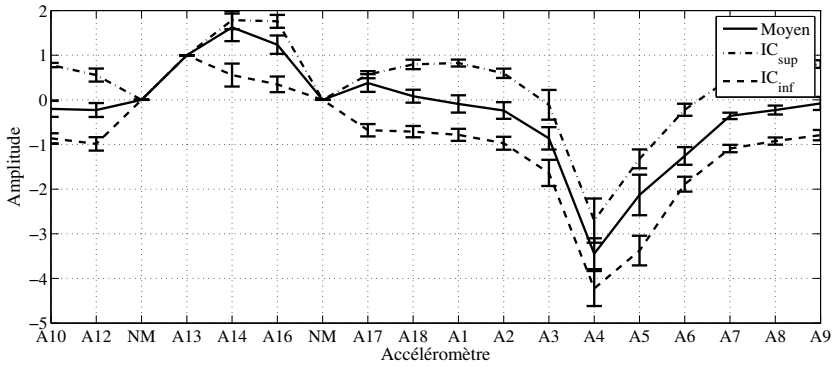
et de validation). Le modèle choisi est celui dont l'erreur commise lors de phase de validation est la plus faible. Étant donné que le nombre de poids du réseau calibré est conséquent (50 dans la couche cachée + 10 dans la couche de sorties + 11 biais), les valeurs ne sont pas données. La Figure 6.20(a) illustre les résultats obtenus lors des phases d'entraînement et de validation du modèle. Il est noté que quelques peu d'observations se situent en dehors des intervalles. Les déformées modales et les intervalles de confiance calculés pour chaque composante modale sont montrées dans la Figure 6.20(b).

La Figure 6.21(a) illustre les résultats obtenus pour la prédiction de la première fréquence propre de la deuxième phase d'instrumentation. Dans ce cas (ainsi comme il a été observé avec le modèle linéaire), peu d'observations se situent à l'extérieur des bandes de confiance. Ceci implique que le modèle construit n'est pas capable de distinguer les modifications structurales des effets thermiques. Dans cette analyse, seuls 33 essais sont considérés comme « anormaux ». La plupart de ces essais ont aussi été signalés par le modèle linéaire appliqué à la première fréquence propre. Les résultats en utilisant le modèle de régression non linéaire appliqué aux ratios des composantes modales sont présentés à la Figure 6.21(b).

La démarche pour l'affectation des essais est maintenant mise œuvre. Le prototype est évalué en considérant les essais de la phase 1 (du cluster C(P1)) et les 1131 (1164−33) essais compris dans les bandes de confiance sont considérés. Le Tableau 6.9 regroupe les pourcentages d'affectation calculés. Comme il a été observé précédemment, la plupart des essais ne sont pas affectés au cluster de la phase 1. Ceci implique que le modèle de régression thermique n'est pas suffisant pour discriminer les effets des travaux. En appliquant cette démarche aux modes propres, les résultats sont relativement similaires. Toutefois, il reste toujours des essais pour lesquels la procédure d'affectation n'est pas capable de les distinguer. Étant donné que le cluster C(P1) est toujours



(a) Modèle linéaire appliqué à la troisième fréquence propre (prédiction)



(b) Modèle linéaire appliqué au troisième mode propre (prédiction)

FIGURE 6.19 – Modèle de régression linéaire appliqué aux paramètres modaux (prédiction).

le même, les lois de probabilités ajustées (et valeurs seuils associées) sont identiques à celles déterminées pour le modèle de régression linéaire.

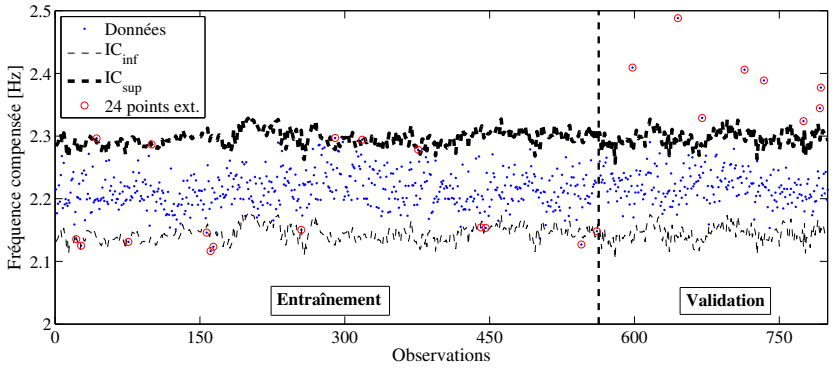
Cluster	**Fréquences propres**	**Modes propres**
C(P1)	37%	35%
Non classé	63%	65%

TABLEAU 6.9 – Affectation des essais en utilisant la première fréquence/mode propre transformés en histogrammes (modèle non linéaire).

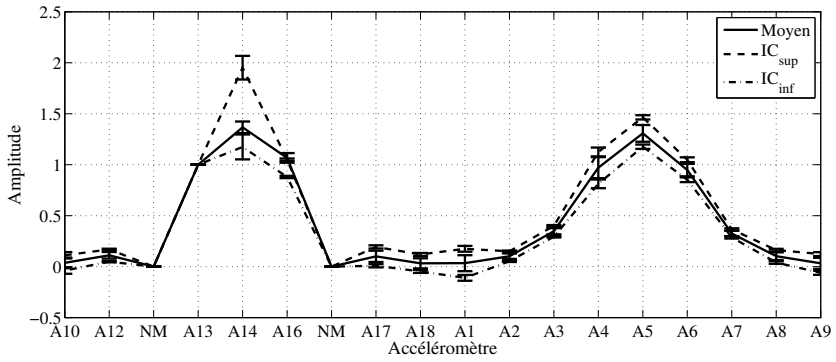
Le modèle de régression non linéaire basé sur les réseaux de neurones est maintenant appliqué à la deuxième fréquence propre. Dans un premier temps, les enregistrements de température et les fréquences propres identifiées durant la première phase sont utilisés pour calibrer le modèle de régression. Les résultats sont montrés à la Figure 6.20(a). Les composantes des déformées modales prédites par le modèle de régression et les bandes de confiance associées sont illustrées à la Figure 6.20(b).

Dans un deuxième temps, les nouveaux enregistrements de températures sont utilisés pour la prédiction des fréquences propres. Les résultats obtenus sont montrés à la Figure 6.23. Les résultats sont similaires à ceux obtenus avec le modèle linéaire. Ici, 139 observations se situent à l'extérieur des intervalles de confiance. Par ailleurs, les déformées modales sont présentées à la Figure 6.21(b).

Le Tableau 6.10 regroupe les taux d'affectation évalués pour chacun des clusters. Il est remarqué qu'un peu plus de la moitié des essais est affecté au cluster qui représente un nouvel état



(a) Modèle non linéaire appliqué à la première fréquence propre (simulation)



(b) Modèle non linéaire appliqué au premier mode propre (simulation)

FIGURE 6.20 – Modèle de régression non linéaire appliqué aux paramètres modaux (simulation).

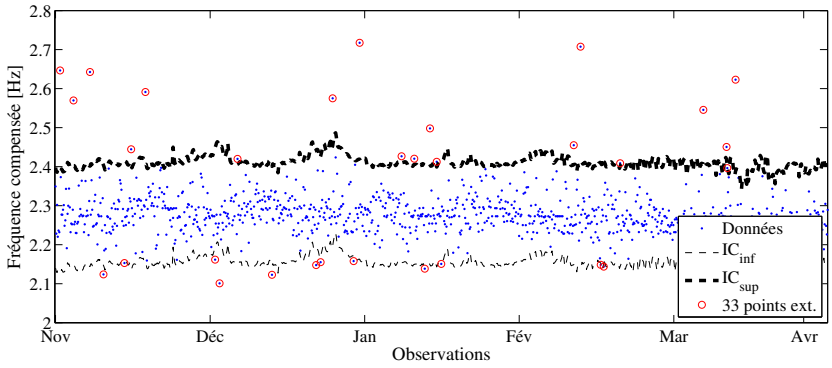
(i.e., différent de celui considéré comme l'état avant les travaux de renforcement). Cela montre que les modèles de régression linéaire (et non linéaire) appliqués à la deuxième fréquence propre peuvent mieux caractériser les modifications structurales que ceux construits en utilisant la première fréquence propre. Dans le cas des modes propres, près de 60% des essais ne sont pas classés au cluster de la phase 1. Cela montre que l'approche d'affectation peut servir de complément aux modèles de régression pour mettre en évidence les effets d'une modification structurale. Toutefois, il reste encore un nombre important d'essais pour lesquels les effets thermiques ne peuvent pas être distingués des ceux causés par les travaux.

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	48%	43%
Non classé	52%	57%

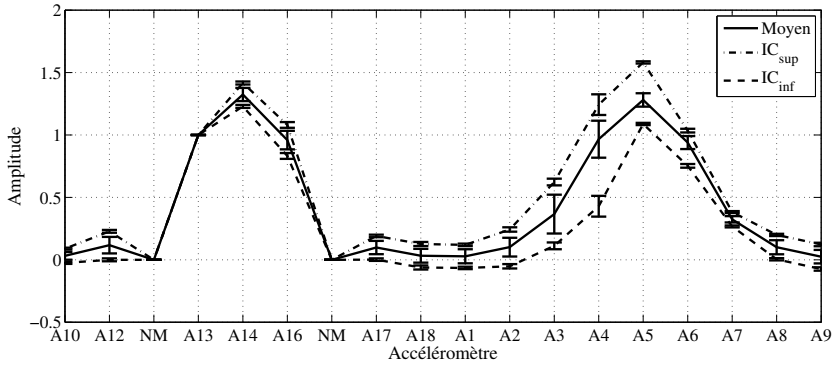
TABLEAU 6.10 – Affectation des essais en utilisant la deuxième fréquence/mode propre transformés en histogrammes (modèle non linéaire).

Finalement, un modèle non linéaire est construit pour la troisième fréquence propre. Le modèle est préalablement calibré en considérant les mesures de température et les fréquences propres identifiées lors de la première phase d'instrumentation (cf. Figure 6.24(a)). Les déformées modales et les bandes de confiance sont également montrées (cf. Figure 6.24(b)) avec les écarts-types respectifs (pour chaque composante modale).

Une fois le modèle calibré, les fréquences propres sont prédites à partir des températures mesurées pendant la phase 2. Dans cette analyse, 130 observations se situent dehors des intervalles de confiance (un résultat similaire a été obtenu en utilisant le modèle de régression linéaire). Les



(a) Modèle non linéaire appliqué à la première fréquence propre (prédiction)



(b) Modèle non linéaire appliqué au premier mode propre (prédiction)

FIGURE 6.21 – Modèle de régression non linéaire appliqué aux paramètres modaux (prédiction).

résultats de la prédiction pour les fréquences et modes propres sont montrés à la Figure 6.25.

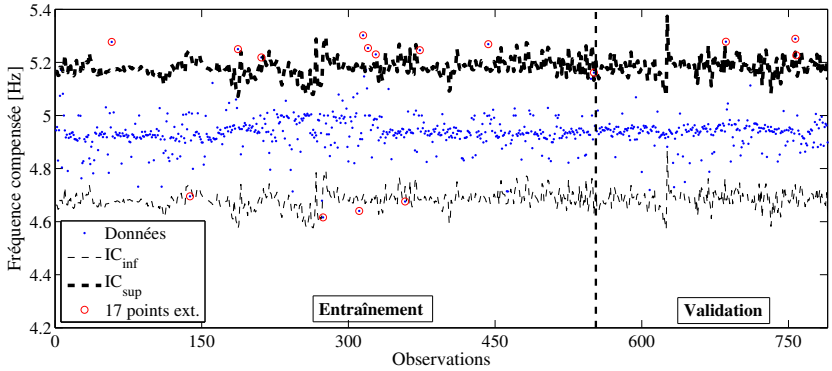
Les taux d'affectation évalués pour chacun des clusters (C(P1) et celui qui représente un nouvel état) sont regroupés dans le Tableau 6.11. Contrairement à ce qui été observé dans le cas du modèle de régression linéaire, un peu plus de la moitié des essais est affecté au cluster de la phase 1. Cela montre que l'utilisation de l'approche d'affectation n'apporte pas une meilleure discrimination des effets thermiques. Cette observation est également mise en évidence en considérant les modes propres.

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	50%	53%
Non classé	50%	47%

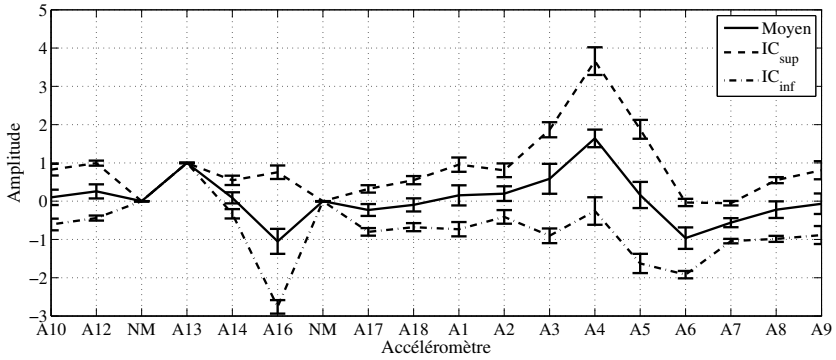
TABLEAU 6.11 – Affectation des essais en utilisant la troisième fréquence/mode propre transformés en histogrammes (modèle non linéaire).

6.3.1.3 Analyse en Composantes Principales

Les simulations suivantes considèrent l'approche basée sur l'ACP (Analyse en Composantes Principales) introduite à le paragraphe 3.5. Les trois premières fréquences propres identifiées lors des phases 1 et 2 d'instrumentation sont considérées pour le calcul de l'indice d'endommagement κ. Les simulations suivantes considèrent la transformation sur une ou deux composantes principales. L'avantage de cette approche est qu'elle ne nécessite pas de calibrer un modèle, comme dans le cas des techniques de régression. De cette façon, une étude sur la détection d'anomalies



(a) Modèle linéaire appliqué à la deuxième fréquence propre (simulation)



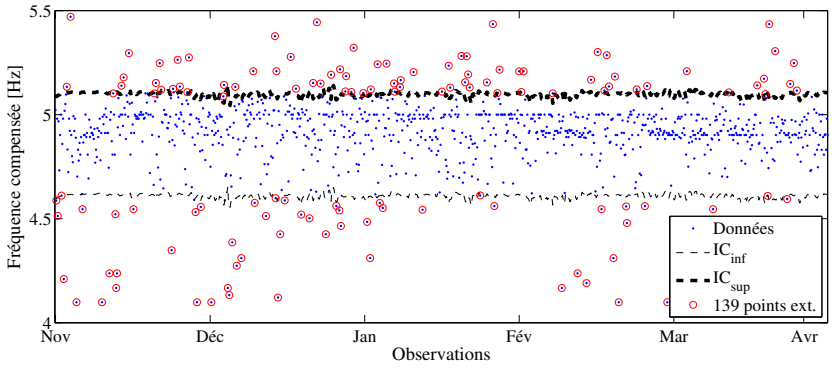
(b) Modèle non linéaire appliqué au deuxième mode propre (simulation)

FIGURE 6.22 – Modèle de régression non linéaire appliqué aux paramètres modaux (simulation).

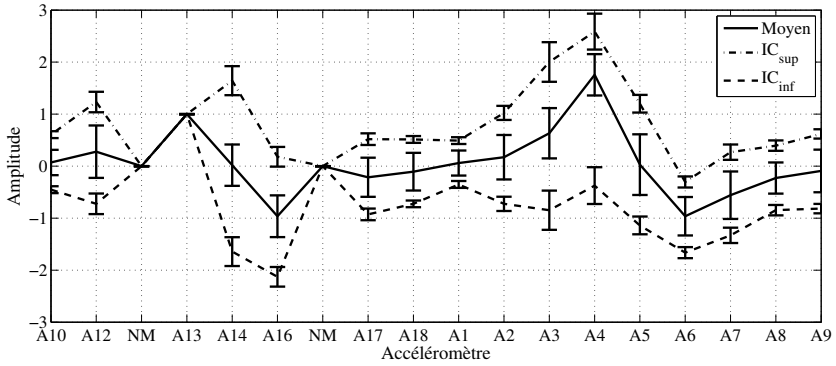
pendant la première campagne d'essais est également menée. De plus, les fréquences propres identifiées pendant le mois d'octobre (2008 et 2009) peuvent être aussi analysées, contrairement aux méthodes de régression qui nécessitaient de mesure des températures.

Dans un premier temps, la norme Euclidienne est utilisée. La Figure 6.26 illustre les résultats obtenus en utilisant une et deux composantes principales pour la re-projection des fréquences propres. Le trait vertical sépare les essais de la première phase de ceux de la deuxième. En cela, les indices d'endommagement sont calculés et les essais qui se situent à l'extérieur de l'intervalle de confiance (95%) sont considérés comme « anormaux ». Il est noté que l'utilisation d'une seule composante principale mène à la détection de plusieurs essais anormaux, ce qui est relativement moins observé dans le cas de l'analyse avec deux composantes principales. Ceci a également été remarqué par Yan et al. (2005) et peut être expliqué par le fait que la re-projection dans l'espace original, en ne considérant qu'une seule composante, produit des erreurs plus importantes. En cela, le nombre de « fausses alarmes » tend à être plus grand. De plus, il est difficile de noter un accord net entre les résultats obtenus par ces deux transformations.

L'approche d'affectation d'essais est maintenant considérée. Trois clusters sont utilisés : celui correspondant à la phase 1 (avant), à la phase 2 (après) et finalement celui correspondant à un nouvel état. Dans un premier temps, les 71 essais considérés comme anormaux pendant la phase 1 sont analysés. En cela, ils sont affectés soit au cluster de la phase 1 (C(P1)), soit dans un autre groupe. Dans un deuxième temps, la même procédure est appliquée aux 129 essais de la phase 2. Dans ce dernier cas, les essais sont affectés soit au cluster de la phase 2 (C(P2)), soit dans un groupe différent. Cette démarche est aussi mise en œuvre en utilisant les résultats obtenus par l'ACP sur 2 composantes principales. L'enjeu principal de cette analyse est d'apprécier la capacité de l'ACP à détecter un changement structural. Le Tableau 6.12 regroupe les taux d'affectation



(a) Modèle non linéaire appliqué à la deuxième fréquence propre (prédiction)



(b) Modèle non linéaire appliqué au deuxième mode propre (prédiction)

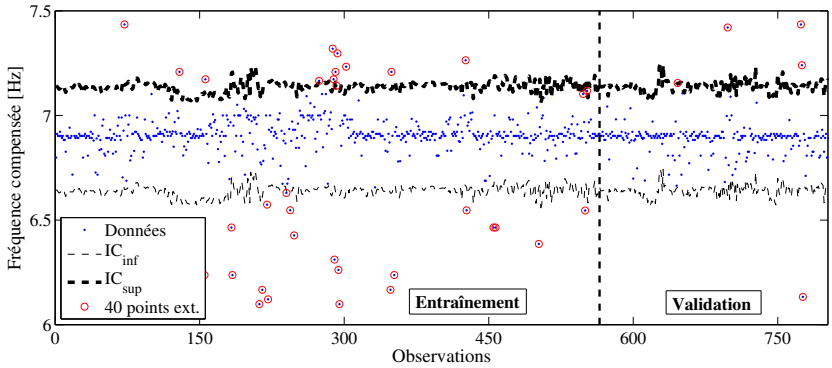
FIGURE 6.23 – Modèle de régression non linéaire appliqué aux paramètres modaux (prédiction).

calculés. Il est noté qu'un grand nombre d'essais sont affectés à leurs clusters originaux. Dans le cas des essais de la phase 1, par exemple, toutes les observations devraient si situer au dessous de la bande de confiance (étant donné qu'aucune modification structurale a été réalisée). Ceci montre que la démarche basée sur l'ACP suscite un nombre important de fausses alarmes. À l'inverse, dans le cas des essais de la phase 2, la plupart devraient se situer au dessus de la bande de confiance, ce qui n'est pas observée. Cela montre que l'ACP peut mener à des erreurs de détection. Pour toutes les simulations, pratiquement 90% des essais ne sont pas considérés comme « anormaux », d'après l'approche d'affectation des nouveaux essais.

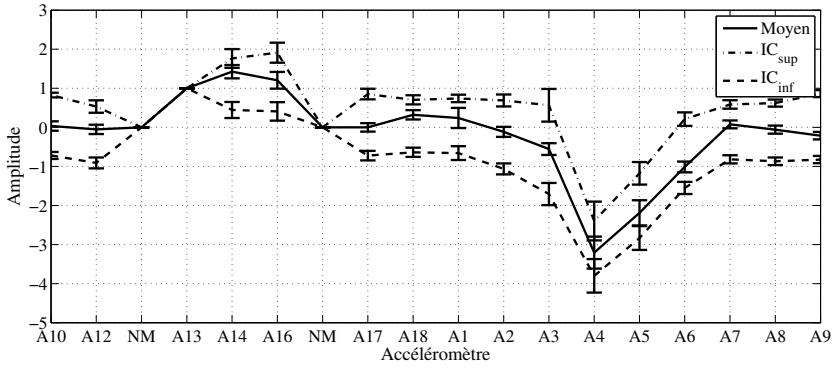
Cluster	**1 CP**	**2 CP**
C(P1)	88%	95%
Non classé	12%	5%
C(P2)	89%	90%
Non classé	11%	10%

TABLEAU 6.12 – Affectation des essais en utilisant les fréquences propres transformées en histogrammes (norme Euclidienne).

La norme de Mahalanobis est maintenant considérée (cf. Figure 6.27). Dans ce cas, il est remarqué que moins d'essais se situent à l'extérieur de l'intervalle de confiance pour les deux phases. Ces essais sont essentiellement les mêmes que ceux signalés par l'analyse avec la norme Euclidienne. En utilisant deux composantes principales, le nombre d'essais caractérisés comme « anormaux » diminue. Toutefois, il existe des essais qui n'ont pas été détectés en n'utilisant qu'une seule composante.



(a) Modèle linéaire appliqué à la troisième fréquence propre (simulation)



(b) Modèle non linéaire appliqué au troisième mode propre (simulation)

FIGURE 6.24 – Modèle de régression non linéaire appliqué aux paramètres modaux (simulation).

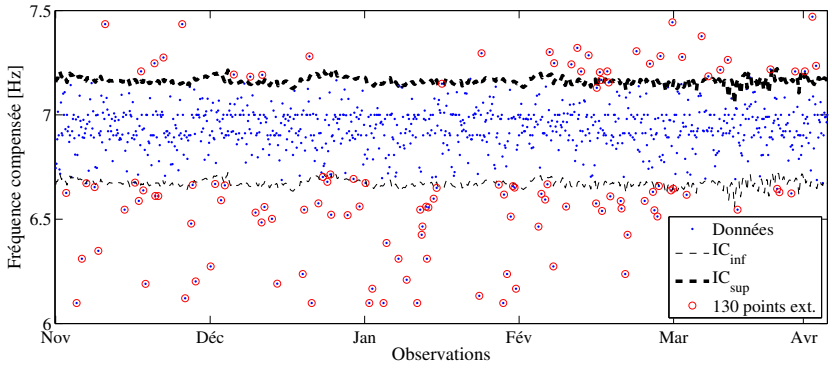
L'approche d'affectation est encore considérée, comme détaillée dans le Tableau 6.13. Comme observé en utilisant la norme Euclidienne, un nombre important d'essais n'est pas considéré comme étant anormal. Cela montre que l'utilisation de l'ACP peut ne pas être pertinente pour mettre en évidence les modifications structurales réalisées dans cet ouvrage.

Cluster	**1 CP**	**2 CP**
C(P1)	84%	87%
Non classé	16%	13%
C(P2)	95%	94%
Non classé	5%	6%

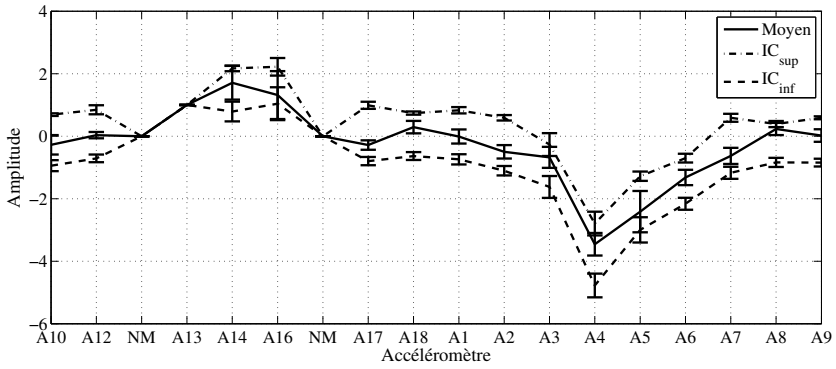
TABLEAU 6.13 – Affectation des essais en utilisant les fréquences propres transformées en histogrammes (norme Mahalanobis).

6.3.1.4 Bilan

Dans le paragraphe précédent, une approche originale basée sur le couplage des modèles de régression linéaire et non linéaire à l'ADS a été proposée. L'objectif de cette démarche est de pouvoir détecter des variations des fréquences propres n'étant pas liées aux effets de la variation thermique. Pour cela, un procédé réalisé en deux temps a été considéré. Dans un premier temps, des modèles de régression sont calibrés à partir de l'ensemble des températures mesurées et des fréquences propres identifiées pendant la première campagne d'essais. Une fois les modèles ajustés, des nouvelles valeurs de températures (mesurées lors de la deuxième phase d'instrumen-

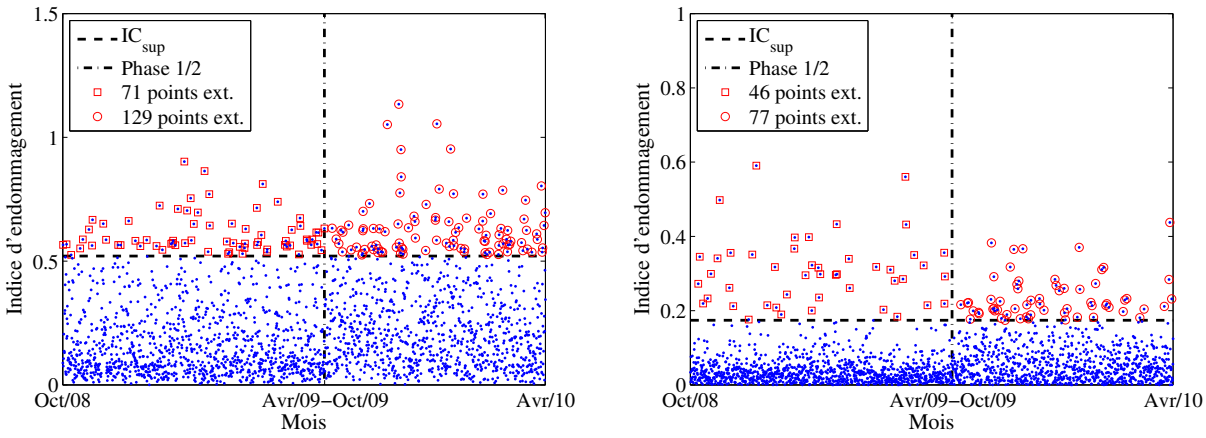


(a) Modèle non linéaire appliqué à la troisième fréquence propre (prédiction



(b) Modèle non linéaire appliqué au troisième mode propre (prédiction)

FIGURE 6.25 – Modèle de régression non linéaire appliqué aux paramètres modaux (prédiction).

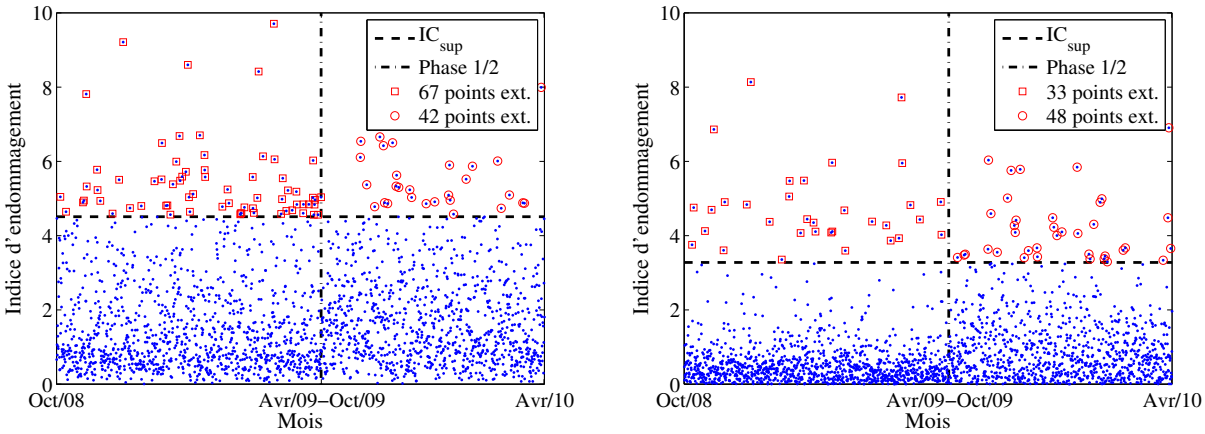


(a) ACP en utilisant une composante principale

(b) ACP en utilisant deux composantes principales

FIGURE 6.26 – ACP en considérant la norme Euclidienne.

tation) sont utilisées par ces modèles. Ainsi, des valeurs prédites pour les fréquences propres sont évaluées. À partir de la détermination des intervalles de confiance pour les valeurs prédites, il est possible de détecter des observations dites « anormales » (celles qui se situent en dehors des intervalles calculés). En réalité, ces observations représentent les effets causés par les travaux de renforcement réalisés. En d'autres termes, toutes les observations qui se situent en dehors de la bande de confiance sont celles pour lesquelles les effets du renforcement sont remarqués. Théoriquement, si le modèle de régression est bien ajusté, toutes les observations prédites doivent



(a) ACP en utilisant une composante principale

(b) ACP en utilisant deux composantes principales

FIGURE 6.27 – ACP en considérant la norme de Mahalanobis.

être caractérisées comme « anormales ». Il faut cependant considérer la possibilité d'erreurs de détection. Il se peut qu'une observation considérée normale (dans la bande de confiance) par les modèles de régression, ne représente pas nécessairement la même condition structurale, mais un événement parfois causé par le trafic, l'humidité, le vent, etc. ou encore une erreur de prédiction du modèle.

Afin d'apporter d'autres éléments de réponse, cette démarche est couplée avec l'ADS, et plus spécifiquement avec des méthodes d'affectation d'essais. La deuxième étape consiste alors à appliquer ces méthodes aux essais considérés comme « normaux » par les modèles de régression en les affectant soit au groupe d'essais de la phase 1, soit dans un autre groupe (correspondant à un état structural différent). Il est alors possible de déterminer si la détection par les modèles de régression est pertinente ou non. En effet, il a été montré que les modèles de régression n'ont pas signalé un nombre important d'essais comme étant des « anomalies ». Toutefois, en utilisant la deuxième étape d'affectation, il a été montré que la plupart n'appartenaient généralement pas au cluster d'essais de la phase 1. Ces résultats montrent qu'il ne suffit pas d'utiliser uniquement des modèles de régression pour discriminer les effets d'une modification structurale de ceux liés à la variation de la température. Certes, pour de certaines simulations, il a été remarqué que même en utilisant l'approche d'affectation des essais, il n'est pas possible de mettre en évidence les modifications structurales réalisés dans cet ouvrage.

Par ailleurs, une approche basée sur l'ACP a été proposée. L'avantage est qu'elle ne nécessite pas des valeurs mesurées par les sondes de température, une fois que les effets thermiques sont considérés comme étant inhérents aux fréquences propres identifiées. Les résultats obtenus par l'ACP ne donnent généralement pas suffisamment des éléments de conclusion pour distinguer une modification structurale d'un changement lié aux effets thermiques. Il a été montré qu'un nombre élevé de fausses alarmes a été détecté par l'approche d'affectation des essais.

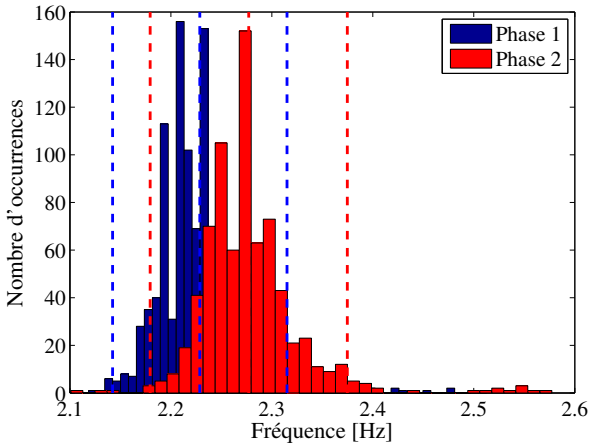
En général, ces démarches permettent d'accompagner, en temps réel d'acquisition de données, des modifications qui sortent de la « respiration naturelle » de l'ouvrage. Bien qu'elles ne permettent pas de tirer des conclusions définitives sur le caractère de ces modifications, elles représentent un nouvel élément pour la détection de changements structuraux.

6.4 Analyse classique

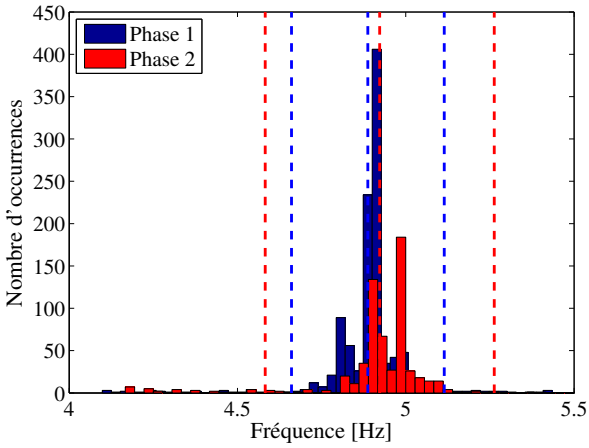
6.4.1 Sans prise en compte des effets thermiques

Avant de mettre en œuvre les démarches utilisant l'ADS, une brève analyse classique est menée en tenant compte des outils statistiques dits classiques et de quelques méthodes citées dans le Chapitre 1, pour la discrimination de comportements structuraux.

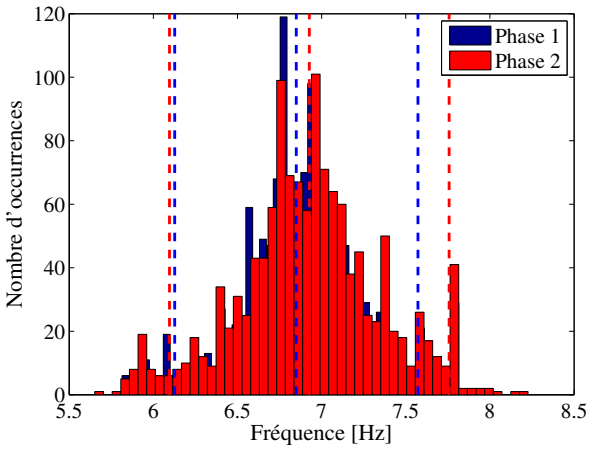
Les histogrammes des trois premières fréquences propres identifiées pendant les deux campagnes de mesures sont montrés dans la Figure 6.28. Dans cette figure, les traits verticaux pointillés représentent les intervalles de confiance à 95% pour les valeurs moyennes des fréquences propres. Il est remarqué que les traits correspondant à chaque état se chevauchent, ce qui ne permet pas de conclure, avec un seuil de confiance à 95%, que les moyennes des fréquences propres ont changé et donc que les états structuraux sont différents.



(a) Première fréquence propre



(b) Deuxième fréquence propre



(c) Troisième fréquence propre

FIGURE 6.28 – Comparatif des histogrammes pour les trois premières fréquences propres (phases 1 et 2).

Des tests d'hypothèse sont également effectués en utilisant les fréquences propres identifiées avant et après renforcement. Comme dans le chapitre précédent, l'hypothèse nulle du test est que les fréquences sont issues d'une même distribution (ce qui implique à la non détection des effets des travaux réalisés). L'hypothèse alternative est donc que les fréquences propres sont

issues d'une distribution différente. Le niveau d'acceptation est toujours fixé à 5%. Dans cette étude, les tests sont réalisés en considérant les trois premières fréquences propres. Les résultats obtenus signalent la non acceptation de l'hypothèse nulle pour les trois tests menés, ce qui met en évidence que les distributions avant et après renforcement sont différentes. Les valeurs p^* sont essentiellement égales à zéro (de l'ordre de 10^{-15} à 10^{-34}) pour tous les tests.

Par ailleurs, une analyse comparative des déformées modales est menée. Les modes propres sont tracés pour chaque ligne de capteurs, en ne considérant que les accéléromètres verticaux (voir Figure 6.5). Les coefficients MAC sont également calculés. La Figure 6.29 illustre les trois premières déformées modales et les coefficients MAC respectifs (NM dénote les points de mesures sans accéléromètre). Il est noté que la variation des modes est relativement faible, à l'exception du mode 3 qui semble être le plus sensible aux modifications structurales apportées par les travaux de renforcement. Pour les deux premières déformées, le coefficient MAC est supérieur à 99%. Il est noté qu'il ne dépasse pas 85% pour le troisième mode propre. Les effets des travaux menés sont notamment observés au niveau des points de mesure A5 et A14 (localisés à mi-travée). À ce niveau, une diminution sensible des amplitudes modales est remarquée. En effet, pour le premier mode, les amplitudes ont diminué de l'ordre de 5% (pour A14). Le même résultat est observé pour le deuxième mode propre, avec une réduction de près de 11% de l'amplitude (pour A4).

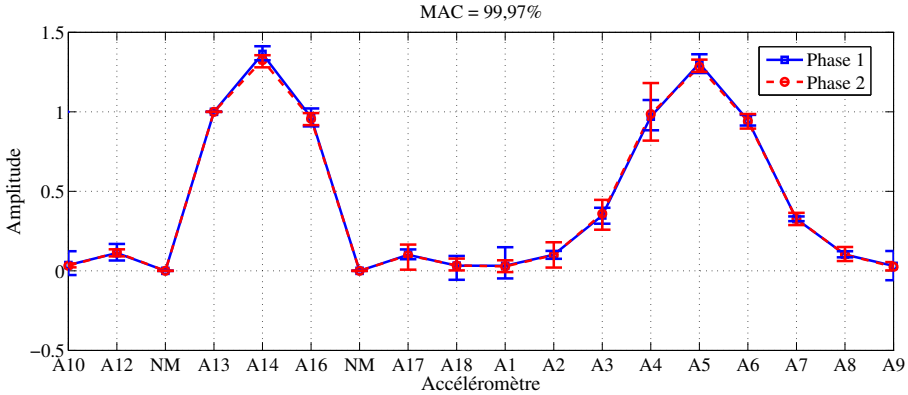
6.4.2 Avec prise en compte des effets thermiques

Avant de passer à l'étude symbolique des paramètres modaux corrigés, il est important de reprendre les analyses classiques exposées précédemment. La Figure 6.30 montre les histogrammes des trois premières fréquences propres corrigées par le filtre linéaire, en comparant les phases 1 et 2. Il est noté que tous les intervalles de confiance se chevauchent, ne permettant pas de distinguer clairement les deux états structuraux. Une analyse similaire est menée en considérant la correction par des filtres non linéaires (cf. Figure C.4). Il est également remarqué que les intervalles de confiance se superposent. Toutefois, il est noté que les distributions des fréquences sont plus écartées après les corrections thermiques (pour les deux filtres). Il est alors espéré que les analyses de clustering et de classification puissent mieux discriminer les états structuraux. Par ailleurs, il est à remarquer que le filtre non linéaire arrive à mieux séparer les distributions de la troisième fréquence propre.

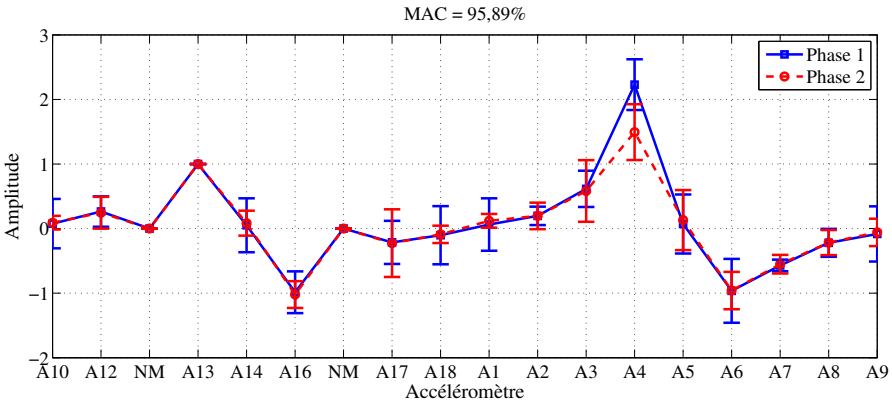
Les tests d'hypothèse ont été également réalisés. Pour les deux types de filtres, les résultats sont identiques. Quelle que soit la fréquence propre considérée, tous les tests rejettent l'hypothèse nulle, ce qui implique la détection des effets des travaux réalisés. Les déformées propres (coefficients MAC associés) sont fournies dans les Figures 6.31 (régression linéaire) et C.5 (régression non linéaire). Il est remarqué une sensible diminution de ces indices. Cela montre que les filtres ont, dans un certain sens, écarté les déformées modales. Par ailleurs, il est noté une meilleure symétrie du troisième mode propre corrigé (points A13 et A16), ce qui peut être éventuellement lié à l'utilisation des procédures de filtrage thermique.

Bien que les analyses classiques permettent d'apprécier, de façon préliminaire, les différences entre les états structuraux, elles ne sont généralement pas suffisamment sensibles pour discriminer des modifications structurales. En effet, si une approche statistique est considérée (comme des intervalles de confiance, par exemple), il est délicat de déterminer, pour un niveau de confiance donné, si les changements observés sur les paramètres modaux sont dus aux travaux réalisés ou à d'autres effets. Il est donc justifié d'appliquer des analyses complémentaires afin de permettre une meilleure discrimination de ces états structuraux.

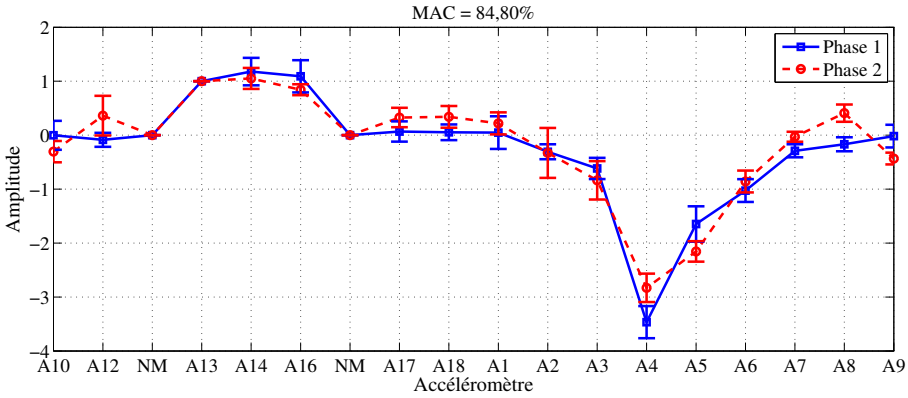
Les paramètres modaux sont rarement identifiés dans leur totalité pour chaque essai enregistré pour des campagnes de surveillance dynamique sur le long terme (comme dans le cas présent). Dans les simulations de clustering et de classification supervisée considérées, seules les trois



(a) Premier mode propre



(b) Deuxième mode propre



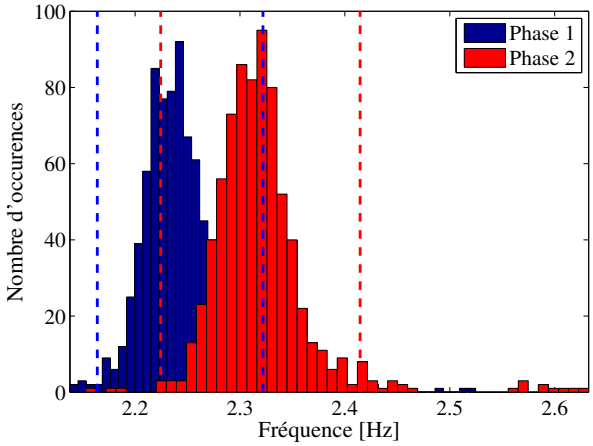
(c) Troisième mode propre

FIGURE 6.29 – Comparatif des trois premiers modes propres et calcul du coefficient MAC (entre les phases 1 et 2).

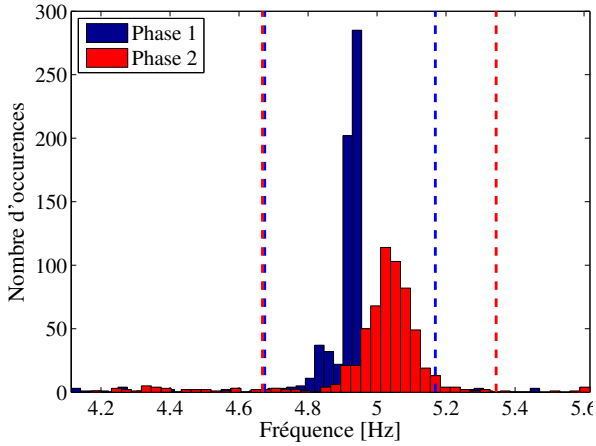
premières fréquences propres et les trois déformées modales associées sont donc utilisées. En effet, ces paramètres modaux sont ceux qui ont été identifiés dans la plupart des essais enregistrés.

6.5 ADS couplée aux méthodes de clustering

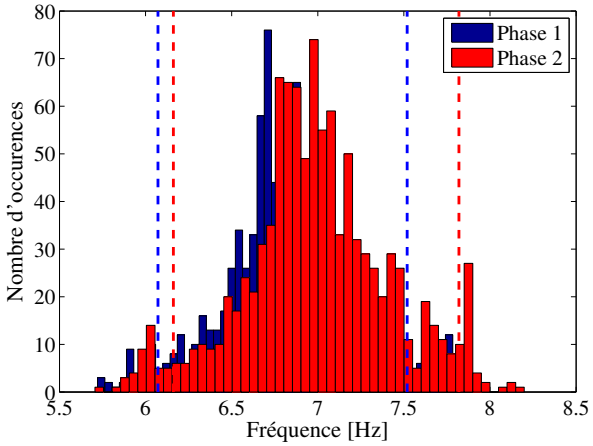
De même que dans les chapitres 4 et 5, l'ADS est couplée aux méthodes de partitionnement de données pour discriminer les deux états structuraux du pont PI-57. L'objectif de cette



(a) Première fréquence propre



(b) Deuxième fréquence propre



(c) Troisième fréquence propre

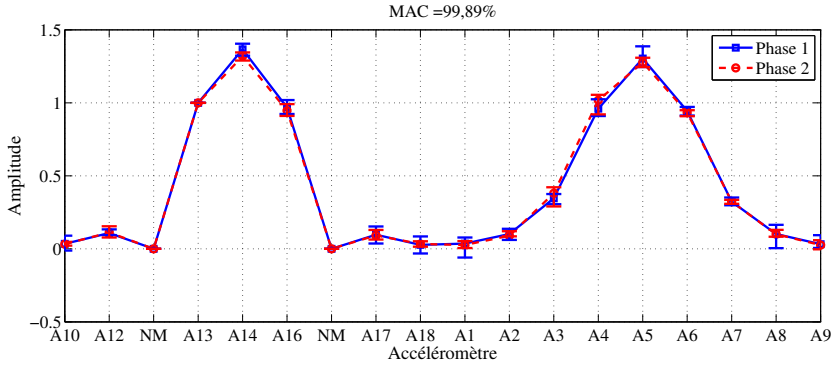
FIGURE 6.30 – Comparatif des histogrammes pour les trois premières fréquences propres corrigées par le filtre linéaire (phases 1 et 2).

étude est d'appliquer les approches proposées dans le cadre de cette thèse pour discriminer les essais enregistrés dans chacune des phases en deux groupes distincts. Les méthodes de clustering sont appliquées aux signaux et aux trois premières paires des paramètres modaux transformés en données symboliques. Dans ce chapitre, deux représentations symboliques sont utilisées. La première, basée sur des intervalles interquartiles et la deuxième, sur des histogrammes à 20 catégories. Dans les analyses suivantes, seuls les résultats obtenus à partir de la transformation en histogrammes sont donnés. Dans le cas des intervalles, les résultats sont montrés dans l'Annexe C de ce mémoire.

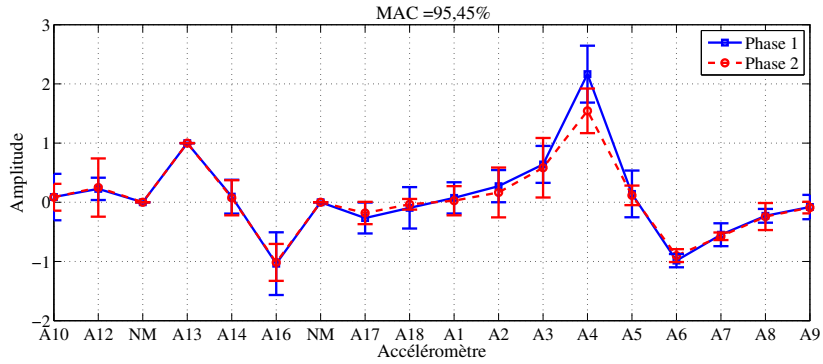
6.5.1 Analyse sur l'ensemble des essais

Comme mentionné précédemment, deux campagnes de mesures ont été réalisées. La première, entre le 14 octobre 2008 et le 4 avril 2009, a permis d'enregistrer un total de 1174 fichiers. La deuxième, durant les années de 2009 et 2010, a permis d'acquérir 1316 fichiers de mesure. Dans les paragraphes suivants, les analyses de clustering appliquées aux signaux considèrent l'ensemble des 2490 essais.

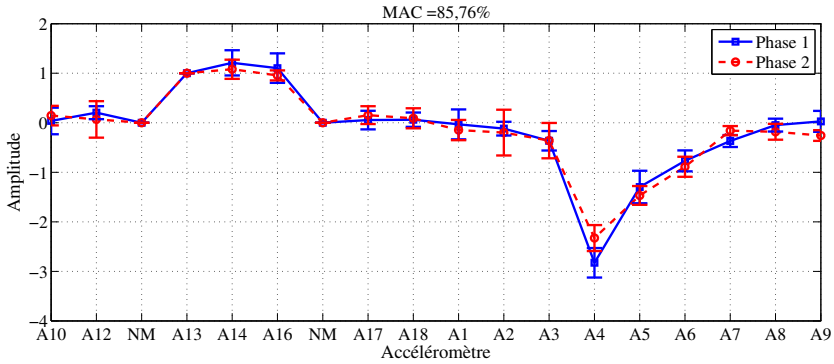
Toutefois, il est important de rappeler que dans la première phase, il n'y a pas d'enregis-



(a) Premier mode propre



(b) Deuxième mode propre



(c) Troisième mode propre

FIGURE 6.31 – Comparatif des trois premiers modes propres corrigés par le filtre linéaire et calcul du coefficient MAC (entre les phases 1 et 2).

trements thermiques pendant le mois d'octobre et début novembre. Ainsi, les analyses menées par la suite appliquées aux paramètres modaux ne considèrent que les essais acquis entre le 21 novembre et le 3 avril (pour les deux phases).

6.5.1.1 Signaux

La méthode des divisions hiérarchiques est appliquée aux signaux transformés en histogrammes. Il est évident que le partitionnement optimal est celui capable de séparer l'ensemble des essais en deux groupes parfaitement distincts, chacun ne contenant que des essais du même type. Étant donné le nombre conséquent d'essais utilisés dans cette simulation, les sorties graphiques

ne sont pas présentées. Les résultats obtenus en utilisant la transformation en histogrammes sont légèrement moins pertinents que ceux obtenus en considérant des intervalles (cf. Annexe C). Dans ce dernier cas, 53% des essais sont classés correctement dans le cluster associé à la première phase et 52% dans celui associé à la phase 2. La variable la plus discriminante est le capteur 2. Il est intéressant de noter que c'est un capteur qui enregistre des accélérations parmi les plus faibles (contrairement à ceux situés à mi-travée, par exemple). Cela montre que le critère de coupure des groupes n'est pas nécessairement lié à l'amplitude des vibrations.

L'utilisation de la méthode des agglomérations hiérarchiques tend à mélanger plus fréquemment les différents états structuraux, avec des taux de classification correcte moins importants que ceux obtenus avec les autres méthodes. Dans cette analyse, et en tenant compte de la transformation en histogrammes, 45% des essais seulement sont correctement classés dans le groupe correspondant à la première phase d'acquisition et 43% dans celui de la deuxième phase.

Lorsque la méthode des nuées dynamiques est considérée et de même que pour les divisions hiérarchiques, les taux de classification correcte sont légèrement plus importants : 54% pour les deux clusters. Toutefois, il faut rester prudent et tenir compte que les taux de réussite sont relativement faibles, ce qui empêche une caractérisation adéquate des effets des travaux de renforcement menés.

6.5.1.2 Fréquences propres

Les trois premières fréquences propres transformées en histogrammes sont considérées dans ce paragraphe. En premier lieu, la méthode des divisions hiérarchiques est utilisée. Dans ce cas, les pourcentages de classification correcte sont plus importants que ceux obtenus avec les signaux, étant donné que 71% des essais sont bien classés dans un cluster et 74% dans l'autre. La variable discriminante est la première fréquence propre.

Les résultats obtenus en utilisant les agglomérations hiérarchiques sont, en général, moins satisfaisants. En considérant des histogrammes, seulement 68% des essais sont correctement classés dans le cluster de la phase 1, alors que 70% le sont pour la phase 2.

En ce qui concerne la méthode basée sur les nuées dynamiques, les taux de classification sont plus élevés que ceux avec la méthode précédente, et du même ordre si comparés à ceux obtenus avec les divisions hiérarchiques. En utilisant la transformation en histogrammes, les pourcentages de classification correcte sont 77% pour le premier groupe (phase 1) et 72% pour le deuxième (phase 2).

6.5.1.3 Modes propres

La même démarche est maintenant appliquée aux trois premiers modes propres. Les résultats de la méthode des divisions hiérarchiques sont en général meilleurs si comparés à ceux obtenus en utilisant les signaux, mais moins pertinents que ceux évalués avec les fréquences propres. Cette tendance a été vérifiée pour l'application numérique au chapitre 4 et expérimentale (pont-rail) au chapitre 5. Dans ce dernier cas 70% des essais sont correctement classés dans le groupe correspondant à la phase 1 et à la phase 2, lorsque des histogrammes sont utilisés. La variable la plus discriminante est le deuxième mode propre.

Lorsque les agglomérations hiérarchiques sont considérées, les taux sont généralement plus faibles que ceux obtenus avec les autres méthodes de partitionnement. Dans ce cas, les pourcentages de classification correcte sont de 62% dans le cluster de la phase 1 et 61% dans celui de la phase 2. En comparant les deux types de transformation en données symboliques, il est noté que les classifications sont relativement meilleures dans le cluster de la phase 1 en considérant

des histogrammes et, à l'inverse, les taux pour le cluster de la phase 2 sont plus importants en utilisant des intervalles (cf. Annexe C). Ceci met en évidence la nécessité de croiser les résultats issus des différentes transformations en données symboliques.

En ce qui concerne la méthode des nuées dynamiques appliquée aux modes propres, les taux sont plus élevés que ceux obtenus avec les méthodes précédentes. En effet, 74% des essais correspondant à la phase 1 d'acquisition sont correctement classés et 71% de ceux correspondant à la phase 2 le sont dans le cluster associé à la phase « après » renforcement.

Les résultats de cette première analyse laissent penser que les effets environnementaux et/ou les variations extérieures (e.g., trafic routier) rendent difficile l'identification des effets du renforcement si les essais sont considérés tous ensemble. Comparer les mêmes périodes de l'année avant et après renforcement peut être une solution pour mettre en évidence une meilleure discrimination des états structuraux. Cette démarche est testée dans le paragraphe suivant.

6.5.2 Application aux paramètres modaux corrigés

Comme il a été proposé dans l'application expérimentale sur le pont-rail au chapitre 5, la deuxième partie de cette étude consiste à appliquer les trois premières paires de paramètres modaux « compensés » des effets thermiques aux méthodes de clustering. Dans un premier temps, les analyses de clustering considèrent l'ensemble des essais. Il est important de noter qu'il n'y a pas eu d'enregistrements de températures pendant le mois d'octobre 2008. Les analyses suivantes ne considèrent donc que les paramètres modaux identifiés entre les mois de novembre et avril (pour les deux phases d'instrumentation). En cela, seuls les fréquences et modes propres identifiés pendant cette période sont utilisés. En premier lieu, les paramètres modaux sont corrigés par les filtres linéaire et non linéaire. En second lieu, ces paramètres sont utilisés par méthodes de clustering. Les analyses suivantes considèrent toujours les trois premières fréquences propres et déformées modales associées.

6.5.2.1 Analyse sur l'ensemble des essais

Le Tableau 6.14 regroupe les taux de classification correcte obtenus par les divisions hiérarchiques appliquées aux paramètres modaux. En considérant les fréquences propres représentées par des histogrammes, il est noté que les résultats sont relativement plus pertinents que ceux obtenus sans la correction thermique. Les pourcentages obtenus en utilisant les modes propres sont similaires, étant donnée une légère augmentation des taux de classification correcte. En général, les résultats montrent que la correction des paramètres modaux est pertinente, étant donné que les taux de classification sont plus importants (cf. Tableau 6.14) que ceux obtenus sans l'utilisation des filtres thermiques. Dans cette analyse, la deuxième fréquence propre et le deuxième mode propre sont les paramètres modaux utilisés pour la discrimination des groupes structuraux. Par la suite, C(P1) dénote le cluster associé à la phase 1 et C(P2) celui associé à la phase 2.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	75%	71%	71%	70%
C(P2)	74%	73%	74%	70%

TABLEAU 6.14 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (histogrammes ; filtre linéaire).

Par ailleurs, les agglomérations hiérarchiques sont considérées. En considérant les fréquences et modes propres, des légères augmentations sur les taux de classification correcte sont remarquées. Dans le cas des fréquences propres, 70% des essais sont correctement classés dans le cluster de la phase 1 (contre 68% de l'analyse précédente) et 69% (contre 70%) le sont pour la phase 2. Ce dernier taux est légèrement inférieur aux résultats précédents. En considérant les déformées modales, les pourcentages sont aussi relativement plus importants, bien qu'ils sont moins marqués que ceux obtenus en considérant les fréquences propres (cf. Tableau 6.15).

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	70%	65%	68%	62%
C(P2)	69%	62%	70%	61%

TABLEAU 6.15 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (histogrammes ; filtre linéaire).

La méthode basée sur les nuées dynamiques est appliquée aux paramètres modaux. L'augmentation des taux de classification est également observée dans les deux cas. Pour les fréquences propres transformées en histogrammes, par exemple, 76% des essais sont correctement classés dans le cluster de la phase 1 (contre 75% sans prise en compte des effets thermiques) et 74% dans le cluster de la phase 2 (contre 72%). Pour les modes propres, les résultats sont relativement moins pertinents (73% pour la phase 1 et 70% pour la phase 2, contre 74% et 71%, respectivement). Le Tableau 6.16 regroupe la totalité des résultats.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	76%	73%	77%	74%
C(P2)	74%	70%	72%	71%

TABLEAU 6.16 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (histogrammes ; filtre linéaire).

Une deuxième technique, basée sur la régression non linéaire à partir des réseaux de neurones, est proposée. Pour les simulations suivantes, la boîte à outils Netlab a été utilisée. L'architecture du réseau consiste en une couche cachée avec 10 neurones et la fonction d'activation est la tangente hyperbolique. L'objectif est non seulement de corriger les effets issus de la variation de températures, mais aussi de comparer les résultats obtenus en utilisant le filtre linéaire. En effet, lorsque il s'agit d'une application où de nombreuses observations (mesures) existent, le filtre non linéaire peut être plus adéquat qu'un filtre linéaire.

La méthode des divisions hiérarchiques est appliquée aux fréquences et modes propres corrigés par le filtre non linéaire. Le Tableau 6.17 regroupe les taux de classification obtenus. Dans les deux cas, mais plus nettement pour les fréquences propres, les pourcentages sont significativement plus élevés que ceux obtenus avec le filtre linéaire. Cela montre que l'utilisation du filtre non linéaire est plus pertinente et que la prise en compte des corrections thermiques est essentielle dans cette étude. Les résultats obtenus en utilisant les déformées modales sont relativement meilleurs que ceux sans correction thermique, bien que la différence ici soit plus faible que celle observée avec le filtre linéaire. De plus, la transformation en intervalles interquartiles est relativement plus pertinente que celle en histogrammes (cf. Tableau C.12). Dans cette analyse, la première fréquence propre et le premier mode propre sont les variables discriminantes lorsque des histogrammes sont utilisés.

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	77%	73%	71%	70%
C(P2)	76%	70%	74%	70%

TABLEAU 6.17 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (histogrammes ; filtre non linéaire).

Le Tableau 6.18 présente les résultats obtenus à partir de l'application des agglomérations hiérarchiques aux paramètres modaux compensés par le filtre non linéaire. Contrairement à ce qui a été observé avec la méthode des divisions hiérarchiques, les pourcentages de classification correcte n'ont pas changé significativement. En général, les classifications obtenues ici sont légèrement meilleures que celles obtenues sans la considération des effets thermiques.

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	70%	62%	68%	62%
C(P2)	70%	66%	70%	61%

TABLEAU 6.18 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (histogrammes ; filtre non linéaire).

Finalement, les résultats obtenus par la méthode des nuées dynamiques sont détaillés dans le Tableau 6.19. Comme il a été remarqué pour les divisions hiérarchiques, les taux de classification correcte sont relativement plus importants que ceux obtenus avec le filtre linéaire (ou encore sans l'utilisation des filtres). Dans le cas des fréquences propres, près de 80% des essais sont correctement classés quel que soit le type de transformation symbolique utilisée (cf. Tableau C.14, dans le cas des intervalles). Par ailleurs, les pourcentages obtenus en utilisant les déformées modales sont plus élevées (75% pour les deux clusters).

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	77%	75%	77%	74%
C(P2)	79%	75%	72%	71%

TABLEAU 6.19 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (histogrammes ; filtre non linéaire).

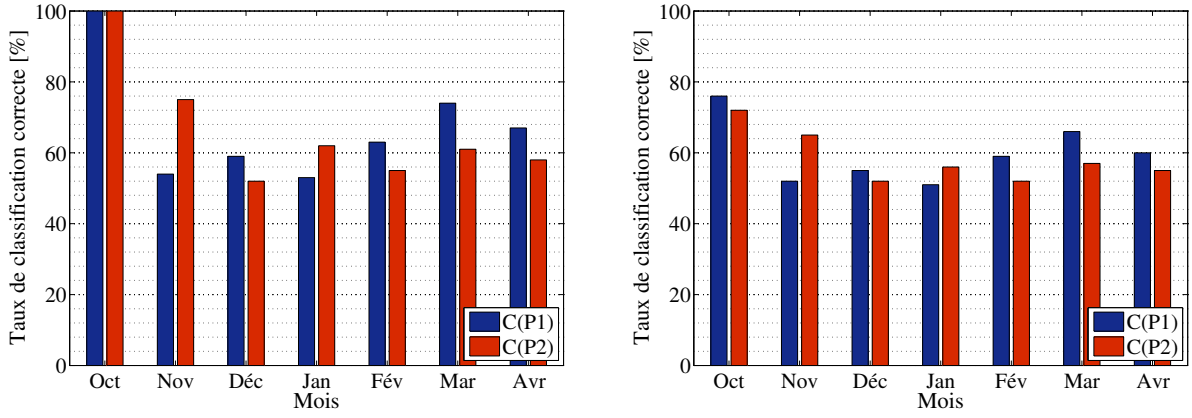
6.5.3 Comparaison des phases 1 et 2 mois par mois

Une analyse détaillée est proposée dans ce paragraphe. À l'inverse d'une étude globale où les essais des phases 1 et 2 sont considérés tous en même temps, un traitement de données mois par mois est proposé. Le but est de comparer les essais enregistrés dans un mois de 2008 avec ceux du même mois en 2009 (pour les mois de janvier, février, mars et avril, les résultats correspondent aux années 2009 et 2010). L'objectif est de diminuer, d'une certaine façon, les incertitudes liées aux changements de température et aux variations du trafic. Pour le mois d'octobre 2008 et 2009, 182 essais ont été enregistrés ; 444 essais en novembre 2008/2009 ; 458 en décembre 2008/2009 ;

455 en janvier 2009/2010 ; 386 en février 2009/2010 ; 439 en mars 2009/2010 et 126 en avril 2009/2010.

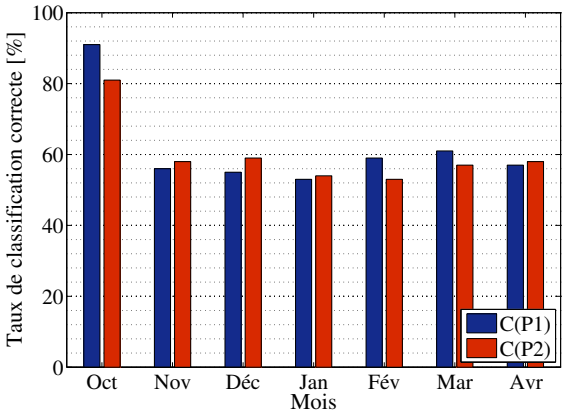
6.5.3.1 Signaux

La Figure 6.32 illustre les résultats obtenus en utilisant les méthodes de partitionnement appliquées aux signaux transformés en histogrammes. Dans cette figure, les barres verticales représentent les taux de classification correcte évalués pour chaque mois sont illustrés. Il est remarqué que pour la méthode des divisions hiérarchiques appliquées aux signaux, 100% de classification correcte est obtenu pour le mois d'octobre. Ce résultat est important, car il montre qu'il est possible d'extraire des informations pertinentes sur le comportement de l'ouvrage directement à partir des mesures effectuées *in situ*. Bien que cette classification ne soit pas atteinte avec les autres méthodes (ou encore pour les autres mois de l'année), les pourcentages de classification correcte ont augmenté sensiblement par rapport aux résultats obtenus dans le paragraphe précédent. Cela met en évidence que la procédure utilisée au paragraphe 6.4.1, où plusieurs essais, enregistrés dans différentes périodes de l'année, sont considérés simultanément, peut ne pas être la plus adéquate pour la détection des modifications structurales.



(a) Divisions hiérarchiques

(b) Agglomérations hiérarchiques



(c) Nuées dynamiques

FIGURE 6.32 – Méthodes de partitionnement appliquées aux signaux transformés en histogrammes.

La Figure 6.32 permet également d'observer que les méthodes des divisions hiérarchiques et des nuées dynamiques sont celles capables de classer les essais avec les meilleurs taux de

réussite, bien qu'ils restent relativement faibles (de l'ordre de 60% en moyenne). Par ailleurs, les agglomérations hiérarchiques présentent des résultats moins pertinents, ayant un taux moyen de 50% de classification correcte. Le Tableau 6.20 regroupe les variables discriminantes déterminées dans chaque analyse en considérant les divisions hiérarchiques. Il est noté que la plupart des capteurs choisis sont ceux situés proche des appuis.

6.5.3.2 Fréquences propres

Les résultats obtenus en utilisant les trois premières fréquences propres sont présentés dans la Figure 6.33. Dans ce cas, les taux de classification correcte sont plus élevés que ceux obtenus en utilisant les signaux. Ainsi, comme il a été remarqué précédemment, la méthode des divisions hiérarchiques appliquée à des histogrammes permet de classer correctement tous les essais enregistrés pendant le mois d'octobre. De plus, les taux de classification correcte sont aussi plus importants que ceux obtenus en tenant compte de tous les essais simultanément. Il est noté que les pourcentages de classification correcte sont de l'ordre de 80%, en moyenne, pour les méthodes des divisions hiérarchiques et des nuées dynamiques, et 70% pour les agglomérations hiérarchiques. Il est également remarqué que la représentation en histogrammes semble être plus adaptée pour cette analyse, étant donné que les taux de classification correcte sont significativement plus élevés que ceux évalués en considérant des intervalles interquartiles (cf. Figure C.7). Le Tableau 6.20 regroupe les variables discriminantes déterminées dans chaque analyse en considérant les divisions hiérarchiques. Dans la plupart des cas, la première fréquence propre a été choisie comme celle la plus discriminante.

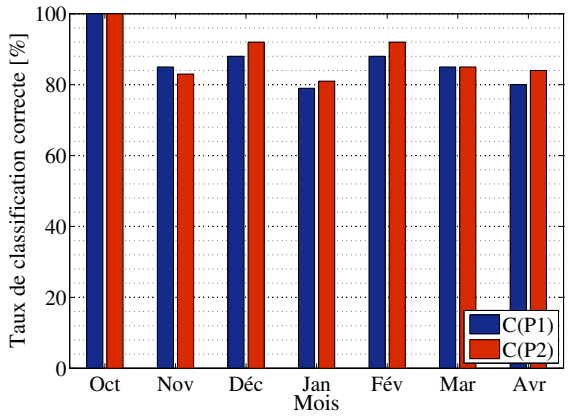
Ici, la démarche proposée pour la correction des paramètres modaux est également considérée et les procédés de clustering sont à nouveau considérés. L'objectif est de vérifier si l'utilisation des filtres thermiques peut améliorer les taux de classification obtenus auparavant. La Figure 6.34 présente les résultats obtenus avec la méthode des divisions hiérarchiques appliquées aux fréquences propres corrigées par le filtre linéaire et non linéaire, respectivement. Il est noté que pour la plupart des mois considérés, les taux de classification correcte sont plus importants en utilisant le filtre non linéaire. Ceci est notamment observé pour le mois de décembre où les pourcentages sont particulièrement plus élevés que ceux obtenus sans la correction thermique (Figure 6.33(a)).

La méthode des agglomérations hiérarchiques, considérée à la Figure 6.35, permet d'atteindre des taux de classification correcte légèrement plus importants en utilisant des filtres thermiques. Lorsque que le filtre non linéaire est utilisé, les résultats sont encore améliorés. Dans ce dernier cas, le taux moyen pour tous les mois est de l'ordre de 80% (contre 75% pour le filtre linéaire).

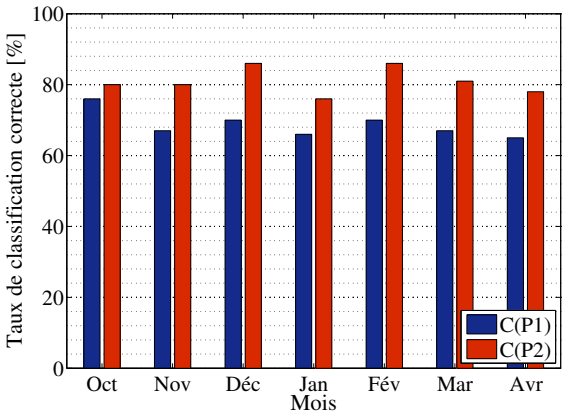
Les classifications obtenues en utilisant les nuées dynamiques sont présentées à la Figure 6.36. Pour cette méthode, les résultats sont significativement meilleurs en considérant les deux filtres thermiques. L'utilisation du filtre non linéaire permet là encore d'obtenir les meilleurs résultats. Ceci est vérifié surtout pour le mois de février où pratiquement tous les essais correspondant à la phase 2 sont classés correctement.

6.5.3.3 Modes propres

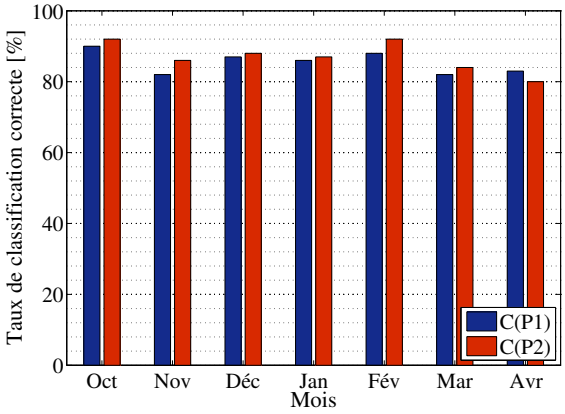
La Figure 6.37 illustre les résultats obtenus en utilisant les trois premiers modes propres. Des taux relativement plus élevés de classification correcte sont également observés dans ce cas (par rapport à l'analyse de l'ensemble d'essais). En particulier, 75% des essais sont, en moyenne, classés correctement en utilisant la méthode des divisions hiérarchiques. Le taux est relativement plus faible lorsque les agglomérations hiérarchiques sont considérées, avec une moyenne de 70%. Pour cette méthode, il est noté que la représentation en intervalles interquartiles (cf. Figure C.11)



(a) Divisions hiérarchiques

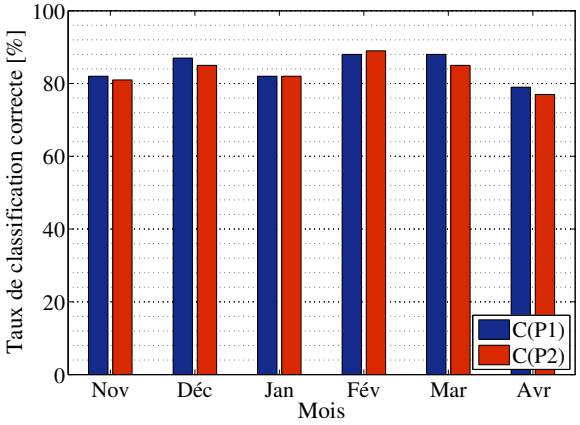


(b) Agglomérations hiérarchiques

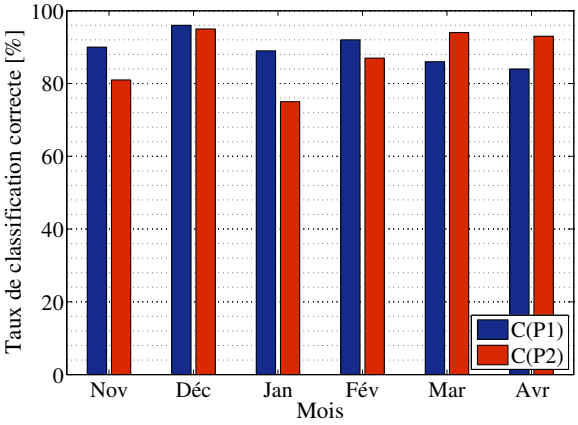


(c) Nuées dynamiques

FIGURE 6.33 – Méthodes de partitionnement appliquées aux fréquences propres transformées en histogrammes.



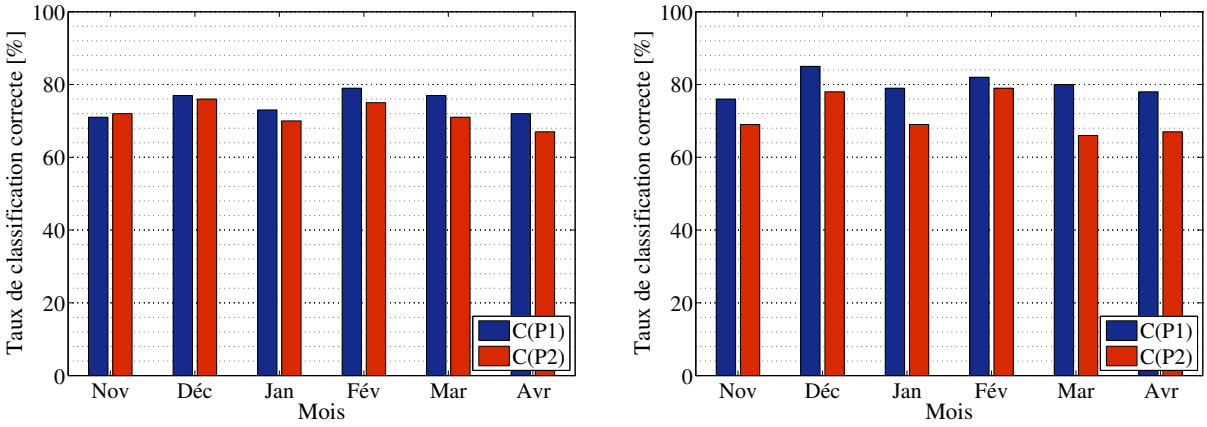
(a) Filtre linéaire



(b) Filtre non linéaire

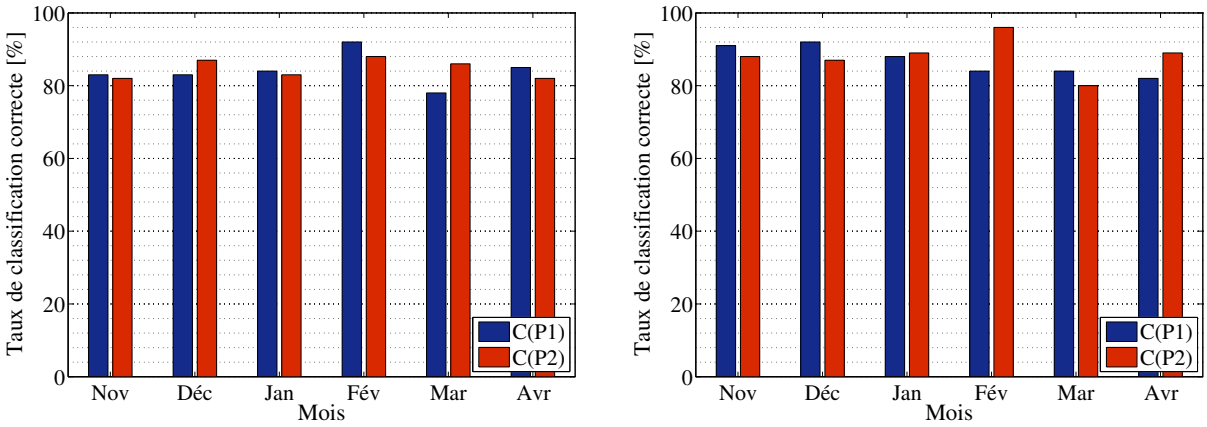
FIGURE 6.34 – Méthode des divisions hiérarchiques appliquées aux fréquences propres « corrigées » (filtres linéaire et non linéaire)

est la plus adaptée, avec des pourcentages plus élevés que ceux obtenus en considérant des histogrammes. Par ailleurs, la méthode basée sur des nuées dynamiques produit les meilleurs résultats.



(a) Filtre linéaire (b) Filtre non linéaire

FIGURE 6.35 – Méthode des agglomérations hiérarchiques appliquées aux fréquences propres corrigées (filtres linéaire et non linéaire)



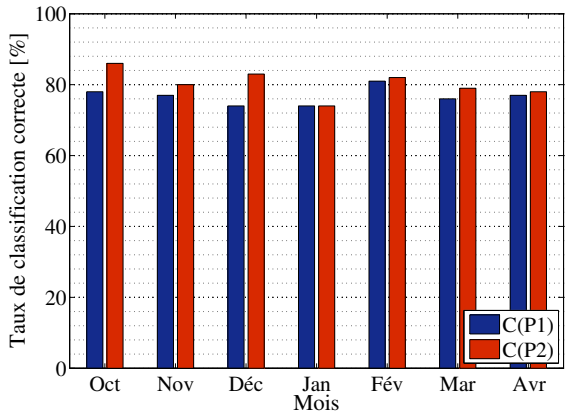
(a) Filtre linéaire (b) Filtre non linéaire

FIGURE 6.36 – Méthode des nuées dynamiques appliquées aux fréquences propres corrigées (filtres linéaire et non linéaire)

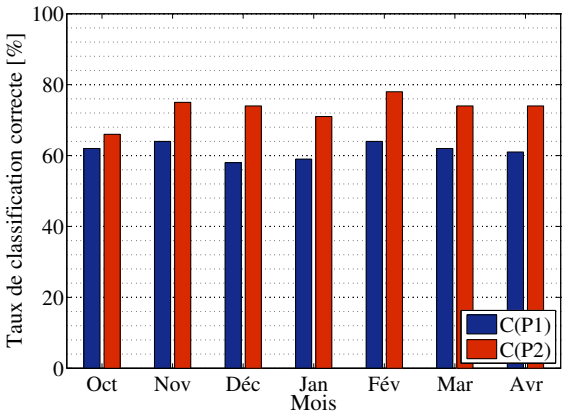
Bien que les taux de classification correcte soient pratiquement du même ordre que ceux avec des divisions hiérarchiques, il est observé que l'utilisation des nuées dynamiques permet d'obtenir des taux plus élevés pour les mois de janvier et février. Le Tableau 6.20 regroupe les variables discriminantes déterminées dans chaque analyse en considérant les divisions hiérarchiques. Il est noté que le deuxième mode propre est celui qui est le plus souvent utilisé pour la discrimination des clusters.

La procédure de correction thermique est maintenant considérée. Les pourcentages de classification correcte obtenus avec la méthode des divisions hiérarchiques sont présentés dans la Figure 6.38, en considérant les deux types de filtre. Même que les taux de classification correcte soient moins importants que de ceux calculés avec les fréquences propres, ils sont plus élevés que ceux obtenus sans prise en compte de la variation thermique. Il est également noté que le filtre non linéaire a produit des meilleurs résultats pour la majorité des simulations réalisées.

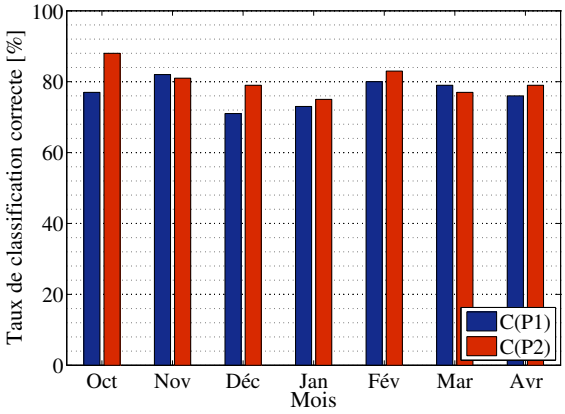
La Figure 6.39 présente les classifications obtenues en utilisant la méthode des agglomérations hiérarchiques. Dans ce cas, les taux de classification correcte obtenus par le filtre linéaire sont du



(a) Divisions hiérarchiques

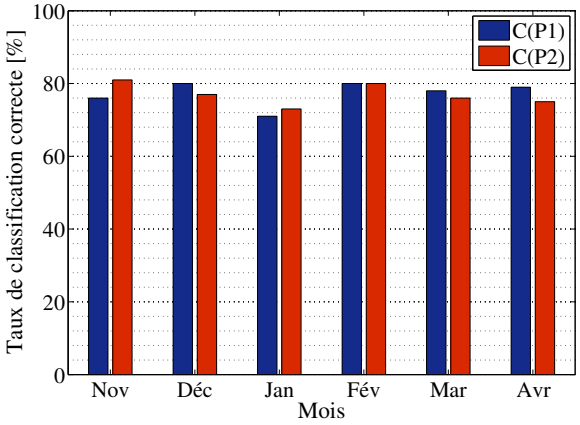


(b) Agglomérations hiérarchiques

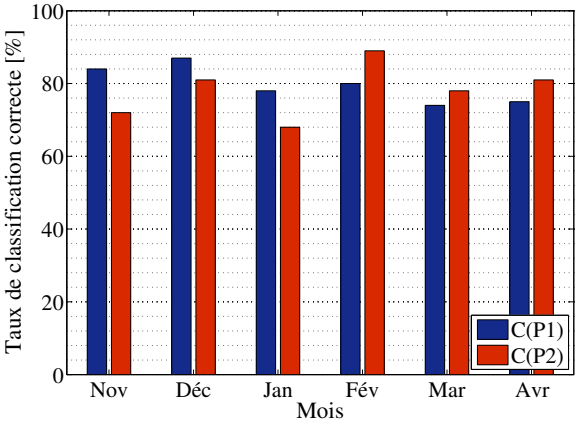


(c) Nuées dynamiques

FIGURE 6.37 – Méthodes de partitionnement appliquées aux modes propres transformés en histogrammes.



(a) Filtre linéaire

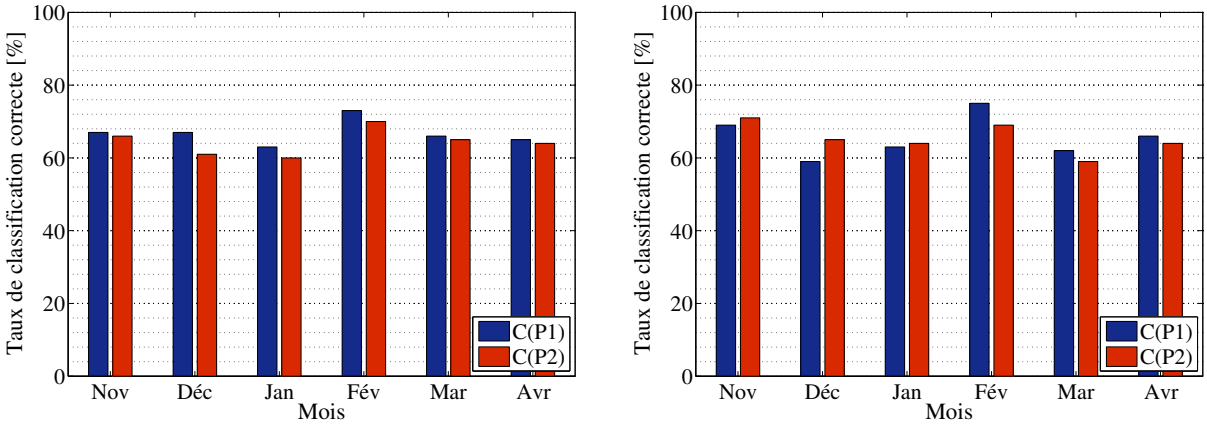


(b) Filtre non linéaire

FIGURE 6.38 – Méthode des divisions hiérarchiques appliquées aux modes propres « corrigés » (filtres linéaire et non linéaire)

même ordre que ceux évalués sans filtre. Ce résultat montre que la méthode de partitionnement n'est pas sensible aux corrections faites sur les modes propres dans ce cas. Toutefois, l'utili-

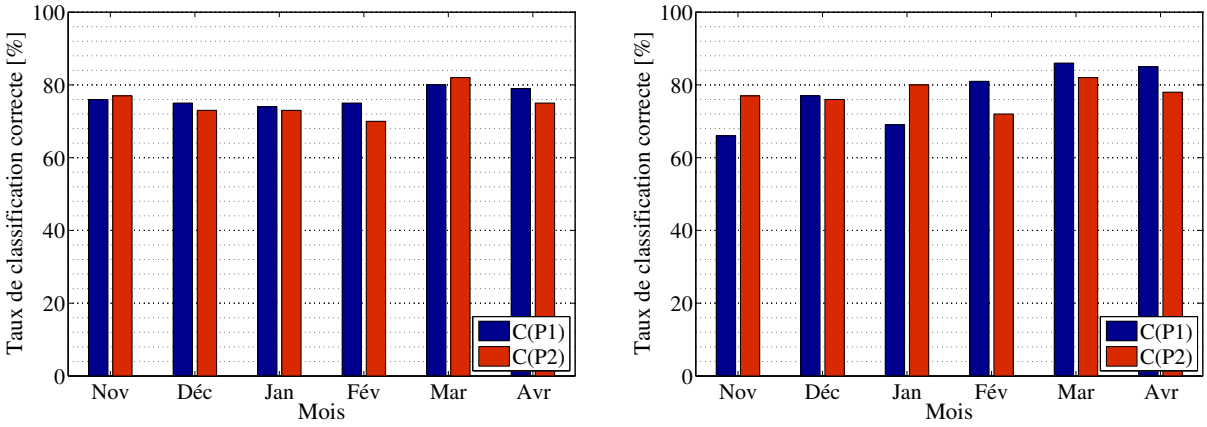
sation d'un filtre non linéaire a permis d'atteindre des pourcentages de classification correcte relativement plus élevés, et se révèle une technique plus pertinente dans le cas étudié.



(a) Filtre linéaire (b) Filtre non linéaire

FIGURE 6.39 – Méthode des agglomérations hiérarchiques appliquées aux modes propres corrigés (filtres linéaire et non linéaire)

La Figure 6.40 illustre les résultats obtenus avec la méthode des nuées dynamiques. Comme il a déjà été remarqué dans les paragraphes précédents, cette méthode produit généralement le meilleurs taux de classification correcte parmi toutes les méthodes considérées. De plus, les taux évalués en utilisant le filtre non linéaire sont généralement plus élevés que ceux obtenus en utilisant le filtre linéaire.



(a) Filtre linéaire (b) Filtre non linéaire

FIGURE 6.40 – Méthode des nuées dynamiques appliquées aux modes propres corrigés (filtres linéaire et non linéaire)

6.5.3.4 Variables discriminantes pour la méthode des divisions hiérarchiques

Le Tableau 6.20 regroupe l'ensemble des variables discriminantes déterminées par la méthode des divisions hiérarchiques appliquée aux signaux et aux paramètres modaux (sans correction thermique). Dans le cas des signaux, les capteurs situés à proximité des appuis sont ceux généralement choisis pour la séparation des clusters. Dans le cas des paramètres modaux, il est noté que

la première fréquence propre (et mode propre associé) est souvent utilisée pour la discrimination des groupes.

	Variable discriminante		
	Capteur	Fréquence propre	Mode propre
Oct.	A3	2	3
Nov.	A2	2	2
Déc.	A2	1	2
Jan.	A10	1	1
Fév.	A12	3	1
Mar.	A16	1	1
Avr.	A4	1	1

TABLEAU 6.20 – Variables discriminantes déterminées pour chaque mois analysé (histogrammes).

Le Tableau 6.20 regroupe l'ensemble des variables discriminantes déterminées par la méthode des divisions hiérarchiques appliquée paramètres modaux corrigés par le filtre linéaire et non linéaire. En général, la détermination des variables discriminantes n'est pas identique pour ces deux cas. Étant donné que les corrections mènent à des valeurs différentes pour les paramètres modaux, il est parfaitement envisageable que les divisions des arbres soient aussi différentes.

	Variables discriminantes			
	Régression linéaire		Régression non linéaire	
	Fréquence propre	Mode propre	Fréquence propre	Mode propre
Oct.	1	1	1	2
Nov.	1	2	3	2
Déc.	2	2	2	2
Jan.	2	1	2	1
Fév.	1	3	3	1
Mar.	2	1	2	1
Avr.	1	1	1	2

TABLEAU 6.21 – Variables discriminantes déterminées pour chaque mois analysé (histogrammes).

6.5.4 Indices de partitionnement optimal

Dans ce paragraphe, les indices introduits dans le paragraphe 3.1.5 sont employés afin de déterminer le nombre de clusters caractérisant le partitionnement optimal. L'objectif est de vérifier si ces indices peuvent indiquer le nombre d'états structuraux existants (dans ce cas, deux). Pour cela, les partitionnements obtenus en utilisant les méthodes des divisions hiérarchiques et des nuées dynamiques sont considérés dans cette étude. Le nombre maximal de clusters est fixé à 5.

6.5.4.1 Analyse sur l'ensemble des essais

Le but est d'évaluer les indices en tenant compte des partitionnements obtenus à partir de l'ensemble total des essais (i.e. les 1174 essais réalisés dans la phase 1 et les 1316 essais enregistrés dans la phase 2). En d'autres termes, l'objectif est de vérifier si les indices de partitionnement optimal indiquent correctement l'existence de deux comportements structuraux : avant et après les travaux de renforcement.

Signaux

Les indices sont d'abord calculés en utilisant les deux méthodes de clustering appliquées aux signaux transformés en histogrammes. Le Tableau 6.22 regroupe les valeurs calculées. Il est noté que l'indice CH est le seul à indiquer le nombre de clusters attendu (i.e., deux). Il est noté que l'indice Γ signale 3 clusters, pour la méthode des divisions hiérarchiques. Enfin, l'indice C^* n'indique jamais le nombre d'états structuraux théoriquement existants. Comme déjà remarqué, cet indice signale systématiquement un nombre de clusters toujours plus élevé que celui attendu. Vraisemblablement, il est possible que cet indice indique des comportements structuraux liés à différents effets extérieurs, tels que le type de trafic, la température, etc.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	CH	C^*	Γ
5	521	**0,076**	0,67	205	**0,098**	0,72
4	442	0,091	0,69	240	0,124	0,7
3	750	0,091	**0,72**	282	0,137	0,71
2	**1170**	0,111	0,68	**462**	0,175	**0,74**

TABLEAU 6.22 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en histogrammes (ensemble total des essais).

Fréquences propres

L'analyse suivante considère le calcul des indices en utilisant les fréquences propres. La lecture du Tableau 6.23 permet de vérifier que les indices CH et Γ indiquent toujours deux clusters comme étant le partitionnement optimal, pour les deux méthodes de clustering considérées. Enfin, l'indice C^* oscille entre 4 et 5 clusters.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	CH	C^*	Γ
5	30,8	**0,036**	0,63	44,3	0,022	0,37
4	32	0,079	0,65	48	**0,019**	0,51
3	35,3	0,117	0,71	52,7	0,075	0,59
2	**37,1**	0,114	**0,73**	**56,7**	0,166	**0,63**

TABLEAU 6.23 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en histogrammes (ensemble total des essais).

Les indices sont également évalués en utilisant les méthodes des divisions hiérarchiques et des nuées dynamiques appliquées aux fréquences propres corrigées par le filtre linéaire (cf. Tableau 6.24). Il est noté que seuls les indices CH et Γ indiquent toujours deux clusters, quelle que soit la méthode. Il est aussi observé que l'indice C^* signale toujours un nombre de groupes distincts plus important.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	CH	C^*	Γ
5	43,6	**0,014**	0,49	33,5	0,073	0,54
4	42,1	0,039	0,55	35,2	**0,072**	0,59
3	49,8	0,044	0,54	39,3	0,109	0,59
2	**51,5**	0,067	**0,61**	**43,6**	0,124	**0,63**

TABLEAU 6.24 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en histogrammes (filtre linéaire).

Les indices sont également calculés pour les partitionnement obtenus en utilisant le filtre non linéaire. Bien que les taux de classification correcte aient été plus élevés en utilisant ce type de filtre, cela ne se reflète pas sur le calcul des indices. En effet, les résultats sont pratiquement les mêmes que ceux obtenus dans le Tableau 6.24. La seule différence est que l'indice C^*, au lieu d'osciller entre quatre et cinq clusters, indique toujours quatre groupes (cf. Tableau 6.25).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	38,6	0,167	0,38	17,5	0,052	0,42
4	39,4	**0,013**	0,37	13,2	**0,044**	0,47
3	47,1	0,079	0,37	16,8	0,076	0,51
2	**50,5**	0,098	**0,43**	**15,9**	0,107	**0,54**

TABLEAU 6.25 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en histogrammes (filtre non linéaire).

Modes propres

Finalement, les indices sont calculés pour chaque partitionnement obtenu en utilisant les modes propres (cf. Tableau 6.26). Dans ce dernier cas, il est noté que les indices indiquent plus souvent un nombre différent de clusters. En réalité, à l'exception de la méthode des divisions hiérarchiques, les indices CH et Γ signalent toujours deux clusters comme étant le partitionnement optimal. Ceci peut être justifié par les meilleurs taux de classification correcte obtenus par cette méthode vis-à-vis de ceux obtenus avec la méthode des divisions hiérarchiques. Par ailleurs, l'indice C^* signale deux clusters uniquement pour la méthode des nuées dynamiques appliquée à des intervalles interquartiles (cf. Tableau C.21). Pour les autres simulations, cet indice varie entre 3 et 5 groupes.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C*†*	Γ	*CH*	*C*†*	Γ
5	104,4	**4,7**	0,67	167,8	3,5	0,68
4	111,4	10,6	0,69	176,7	7	0,7
3	**129,1**	15,4	**0,74**	190,7	**2,3**	0,73
2	128,6	21,6	0,66	**193,3**	4,9	**0,74**

TABLEAU 6.26 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en histogrammes (ensemble total des essais). † Les valeurs de l'indice C^* sont multipliées par $(\times 10^{-2})$.

Le Tableau 6.27 regroupe les indices calculés pour les méthodes des divisions hiérarchiques et des nuées dynamiques en considérant le filtre linéaire, lorsque les modes propres corrigés sont pris en compte. Dans ce cas, aucune différence n'est observée pour les indices CH et Γ, étant donné qu'ils indiquent le nombre attendu de groupes pour toutes les méthodes. Toutefois, contrairement à ce qui a été présenté dans le Tableau 6.26, l'indice C^* indique deux clusters comme étant le partitionnement optimal pour la méthode des divisions hiérarchiques.

Lorsque le filtre non linéaire est utilisé, les résultats sont identiques à ceux obtenus avec le filtre linéaire. En effet, les indices CH et Γ concordent dans toutes les simulations et l'indice C^* oscille entre trois et quatre clusters (cf. Tableau 6.28).

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	45,8	0,124	0,63	42,7	**0,076**	0,65
4	44,4	0,067	0,67	47,1	0,078	0,69
3	51,1	0,058	0,68	51,8	0,183	0,71
2	**52,8**	**0,038**	**0,74**	**56,4**	0,104	**0,76**

TABLEAU 6.27 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en histogrammes (filtre linéaire).

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	48,8	0,073	0,69	43,9	0,083	0,72
4	49,4	0,056	0,74	47,9	**0,082**	0,77
3	56,5	**0,052**	0,75	53,2	0,099	0,81
2	**59,1**	0,075	**0,81**	**56,1**	0,108	**0,86**

TABLEAU 6.28 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en histogrammes (filtre non linéaire).

6.5.4.2 Comparaison des phases 1 et 2 mois par mois

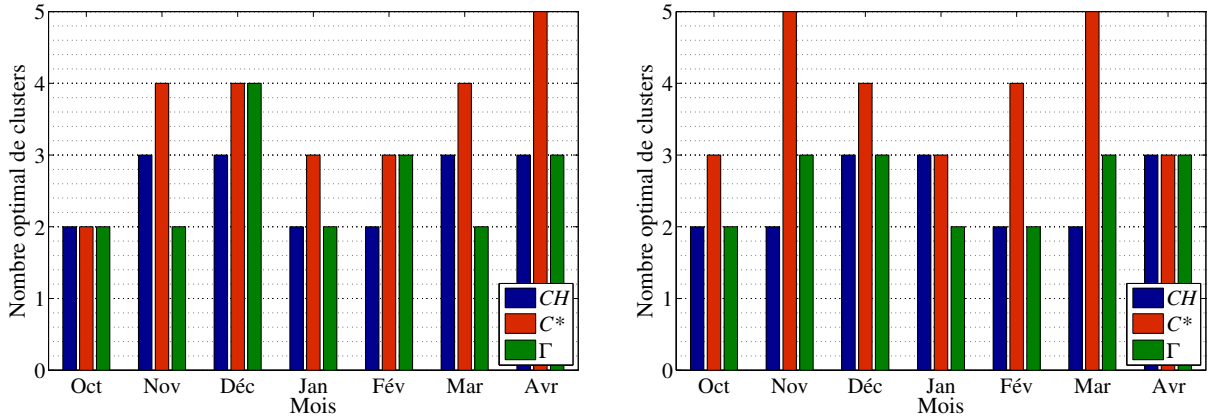
Dans le paragraphe 6.4.2, une étude complémentaire avait été menée en considérant l'analyse des essais mois par mois. Cette analyse a montré être plus pertinente que celle où tous les essais ont été étudiés en même temps. Les résultats obtenus ont mis en évidence que le fait de considérer une analyse des données mois par mois peut diminuer l'occurrence d'effets « parasites » liés aux perturbations extérieures. Ainsi, les indices de partitionnement optimal sont calculés à partir des signaux et des paramètres modaux, triés mois par mois. L'objectif est de vérifier s'il est possible d'améliorer les résultats obtenus au paragraphe 6.4.3.1. Dans la suite, les résultats sont présentés à l'aide de figures. Étant donné le nombre conséquent de simulations réalisées, seul le nombre optimal de clusters est montré.

Signaux

Les premières analyses prennent en compte les partitionnements obtenus en utilisant les divisions hiérarchiques et les nuées dynamiques appliquées aux signaux. La Figure 6.41 illustre les résultats obtenus mois par mois, en considérant la transformation en histogrammes. Dans ce cas, il est noté que les indices CH et Γ indiquent soit le nombre attendu (deux) soit trois clusters. Il est également remarqué que dans le mois de décembre, aucun indice n'indique le nombre attendu de groupes. En utilisant des histogrammes, les trois indices mettent en évidence deux clusters, comme étant le partitionnement optimal. Ceci reflète directement les résultats obtenus lors des procédures de clustering, où 100% des essais ont été correctement classés. L'indice C^*, toutefois, indique un nombre optimal entre trois et cinq clusters dans les deux cas analysés. Par ailleurs, la méthode des nuées dynamiques est considérée. Bien que pour le mois d'avril, tous les indices signalent trois clusters comme étant le partitionnement optimal, l'utilisation des histogrammes permet, en général, d'observer un meilleur accord entre les indices CH et Γ.

Fréquences propres

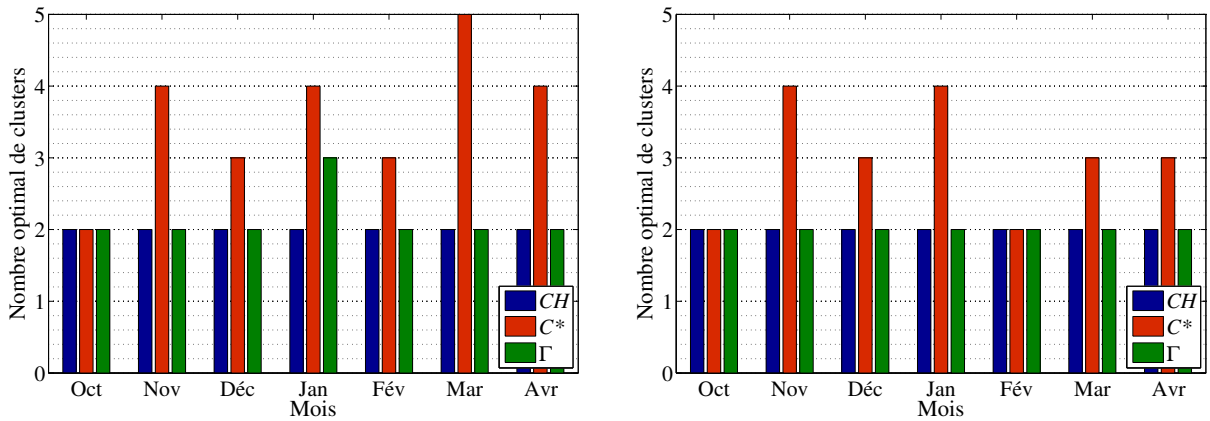
La même démarche est proposée en tenant compte des fréquences propres. La Figure 6.42(a) montre les résultats obtenus en utilisant les divisions hiérarchiques. Si la transformation en histogrammes est considérée, les indices CH et Γ indiquent deux clusters quelle que soit la paire de mois analysée, sauf pour le mois de janvier, où l'indice Γ indique trois groupes. L'indice C^*



(a) Divisions hiérarchiques (b) Nuées dynamiques

FIGURE 6.41 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés histogrammes (analyse mois par mois).

est toujours le plus erratique. Mis à part le mois d'octobre, pour lequel cet indice signale un partitionnement optimal de deux clusters, il oscille entre des valeurs plus importantes. Il a été déjà montré dans les études précédentes que cet indice peut être plus sensible à des variations extérieures, telles que le trafic, le vent, la température, l'humidité, etc. En considérant la méthode des nuées dynamiques, les indiquent CH et Γ indiquent correctement trois clusters pour tous les mois.



(a) Divisions hiérarchiques (b) Nuées dynamiques

FIGURE 6.42 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées histogrammes (analyse mois par mois).

L'objectif maintenant est de vérifier si la démarche utilisée pour la correction thermique des paramètres modaux peut influencer les résultats présentés dans la Figure 6.42. Dans la suite, les résultats sont présentés de façon synthétique au travers de tableaux. Étant donné le nombre conséquent de simulations réalisées, seulement le nombre optimal de clusters est fourni.

Dans les simulations suivantes la méthode des divisions hiérarchiques est appliquée aux fréquences propres corrigées. Le Tableau 6.29 synthétise les nombres optimaux de clusters déterminés pour chacun des indices, évalués mois après mois et en considérant les deux types de filtres. Il est noté qu'en utilisant le filtre linéaire, les indices CH et Γ signalent l'existence de deux clusters quel que soit le type de transformation symbolique utilisée. Toutefois, l'indice C^* ne concorde

jamais avec les autres. Les résultats sont similaires lorsque le filtre non linéaire est considéré. En effet, la seule différence est observée pour le mois de mars, où l'indice Γ indique trois groupes. Par ailleurs, l'indice C^* signale correctement le nombre de clusters pour les mois d'avril. Les indices calculés sont regroupés dans le Tableau 6.29, lorsque la méthode des nuées dynamiques est considérée. Dans ce cas, les valeurs évaluées en considérant le filtre linéaire montrent que les indices CH et Γ signalent l'existence de trois groupes distincts. En utilisant le filtre non linéaire, toutefois, ces indices concordent dans toutes les simulations, quel que ce soit le type de données symboliques utilisé. L'indice C^* oscille légèrement moins lorsque des filtres thermiques sont appliqués. Cela peut être expliqué par le fait que cet indice est plus sensible à des variations extérieures (comme la température, par exemple). Ainsi, pour les analyses sans correction thermique, il est possible que l'indice C^* indique non seulement le nombre de groupes avec un comportement structural différent, mais aussi le nombre de groupes avec un profil de températures distinctes.

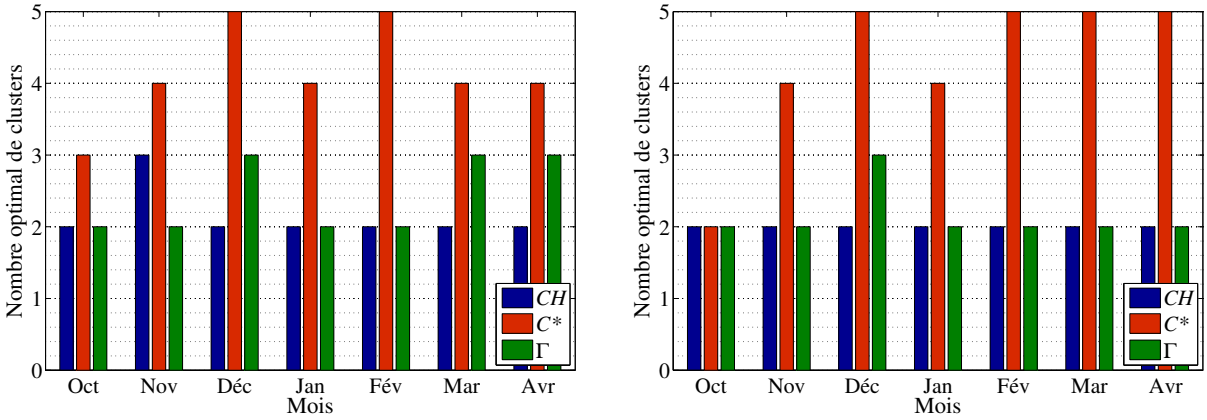
	Divisions hiérarchiques						**Nuées dynamiques**					
Mois	Filtre linéaire			Filtre non linéaire			Filtre linéaire			Filtre non linéaire		
	CH	C^*	Γ	CH	C^*	Γ	CH	C^*	Γ	CH	C^*	Γ
Nov	**2**	3	**2**	**2**	4	**2**	**2**	3	**2**	**2**	4	**2**
Déc	**2**	4	**2**	**2**	4	**2**	**2**	3	**2**	**2**	4	**2**
Jan	**2**	4	**2**	**2**	4	**2**	**2**	3	**2**	**2**	5	**2**
Fév	3	3	**2**	**2**	4	**2**	**2**	4	**2**	**2**	3	**2**
Mar	**2**	3	**2**	**2**	3	3	3	5	**2**	**2**	3	**2**
Avr	**2**	4	**2**	**2**	**2**	**2**	**2**	3	3	**2**	4	**2**

TABLEAU 6.29 – Nombres optimaux de clusters basé sur les méthodes de clustering appliquées aux fréquences propres (histogrammes).

Modes propres

Finalement, les indices sont évalués en considérant les clusters déterminés à partir des modes propres. Dans un premier temps, les divisions hiérarchiques sont utilisées (Figure 6.43(a)). Si la transformation en histogrammes est utilisée, il est observé que les indices CH et Γ concordent dans la plupart des simulations réalisées en indiquant deux clusters. Cependant, quel que soit le type de données symboliques utilisé, l'indice C^* ne donne jamais un partitionnement optimal contenant deux clusters. La Figure 6.43(b) montre également les résultats obtenus en utilisant la méthodes des nuées dynamiques. Les résultats sont particulièrement pertinents en utilisant la représentation en histogrammes. Dans ce cas, tous les indices signalent le même nombre de clusters dans le mois d'octobre. Cependant, l'indice C^* varie significativement, notamment pour les derniers mois analysés.

Par ailleurs, le Tableau 6.30 regroupe les résultats obtenus en utilisant les divisions hiérarchiques appliquées aux déformées modales corrigées. Il est remarqué que, les indices indiquent généralement le nombre attendu de groupes après l'utilisation des filtres thermiques. Bien que ce ne soit pas observé pour certains mois, les indices CH et Γ signalent systématiquement l'existence de deux clusters. Ainsi, comme il a déjà été observé précédemment, l'indice C^* varie entre 2 et 5. Les résultats obtenus par la méthodes des nuées dynamiques sont essentiellement les mêmes que ceux obtenus par les divisions hiérarchiques (cf. Tableau 6.30). Les résultats sont relativement meilleurs que ceux obtenus sans l'utilisation des filtres thermiques. De plus, l'indice C^* oscille relativement moins dans ces dernières simulations.



(a) Divisions hiérarchiques

(b) Nuées dynamiques

FIGURE 6.43 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés histogrammes (analyse mois par mois).

Mois	Divisions hiérarchiques						Nuées dynamiques					
	Filtre linéaire			Filtre non linéaire			Filtre linéaire			Filtre non linéaire		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
Nov	3	4	**2**	**2**	3	**2**	**2**	5	**2**	**2**	3	**2**
Déc	**2**	4	**2**	**2**	**2**	**2**	**2**	5	**2**	**2**	4	**2**
Jan	**2**	3	3	**2**	5	**2**	**2**	3	3	**2**	5	**2**
Fév	**2**	3	**2**	**2**	4	**2**	3	4	**2**	**2**	4	**2**
Mar	**2**	5	**2**	**2**	5	**2**	**2**	3	**2**	**2**	**2**	**2**
Avr	**2**	4	3	**2**	3	**2**	**2**	3	**2**	**2**	3	**2**

TABLEAU 6.30 – Nombres optimaux de clusters basé sur les méthodes de clustering appliquées aux modes propres (histogrammes).

6.5.5 Bilan

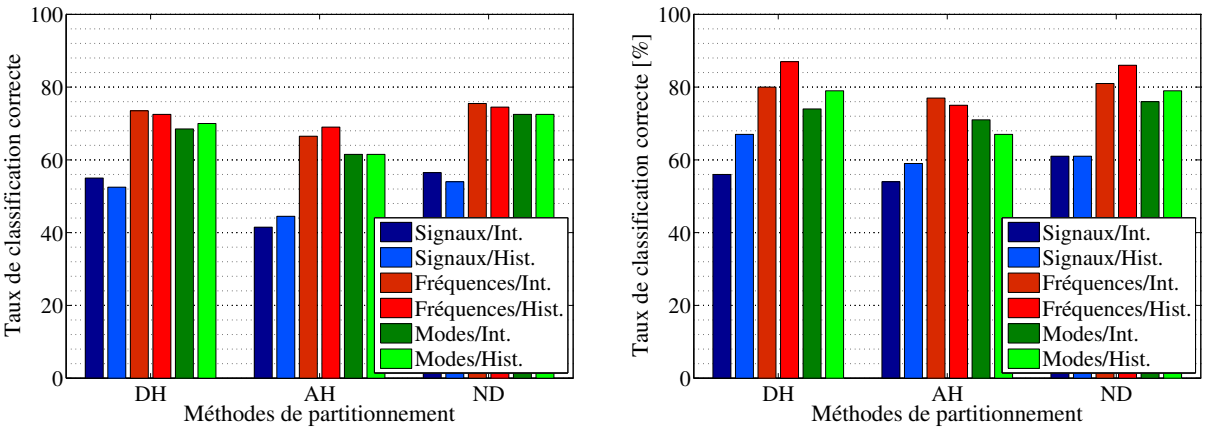
Dans les paragraphes précédents, les méthodes de partitionnement de données ont été utilisées pour la discrimination des deux états structuraux du pont routier PI-57 sur l'Oise. Ce pont, ayant subi des travaux de renforcement pendant l'été 2009, a été soumis à deux campagnes d'essais dynamiques. La première phase, réalisée entre les mois d'octobre 2008 et avril 2009, a permis d'enregistrer environ 1200 fichiers. Au cours de cette même période, entre octobre et avril 2010, une deuxième phase d'acquisition de mesures a été mise en œuvre, avec environ 1300 fichiers enregistrés.

La démarche proposée a consisté, dans un premier temps, à appliquer les méthodes de clustering à l'ensemble des 2490 essais acquis, le but étant de les séparer en deux groupes où chacun représente un des deux états structuraux : avant et après renforcement. Les résultats ont montré que les taux de classification sont relativement faibles, notamment dans le cas de l'utilisation des signaux. Les méthodes basées sur les divisions hiérarchiques et nuées dynamiques ont permis d'obtenir des résultats relativement pertinents en considérant les fréquences propres. Toutefois, les meilleurs taux de classification correcte n'ont pas dépassé la valeur de 75%.

Compte tenu de ces résultats, une analyse plus fine a été menée. Cette analyse a consisté à classer les essais mois par mois. En d'autres termes, au lieu de mener une étude où l'ensemble des essais des phases 1 et 2 sont considérés simultanément, les essais enregistrés dans un mois de l'année 2008 ont été comparés à ceux du même mois de l'année 2009. Cette démarche a

permis d'obtenir une meilleure classification des essais, en particulier pour le mois d'octobre. En effet, 100% des essais ont été correctement classés pour le mois d'octobre en considérant les signaux transformés en histogrammes. Ce résultat est intéressant, car il montre qu'il est possible d'extraire des informations sur le comportement de l'ouvrage directement à partir des mesures effectuées *in situ*.

La Figure 6.44(a) présente une synthèse de la moyenne globale (pour les deux états structuraux) des taux de classification pour chaque méthode et chaque type de transformation utilisée, en considérant tous les essais pris ensemble. Comme il a été déjà remarqué, les méthodes des divisions hiérarchiques et des nuées dynamiques sont les plus performantes pour la discrimination des états structuraux, bien que les pourcentages de réussite soient relativement faibles. Par ailleurs, l'application des méthodes de clustering aux fréquences propres a produit des meilleurs résultats que ceux avec les signaux ou encore les modes propres. La Figure 6.44(b) illustre les moyennes globales (en considérant tous les mois) des pourcentages de classification correcte évalués à partir de l'analyse des essais mois par mois. Dans ce dernier cas, il est remarqué que toutes les méthodes permettent d'atteindre des taux plus élevés, ce qui valide la procédure proposée. Les effets de la variation du trafic routier et de l'oscillation des conditions environnementales sont peut-être atténués avec cette approche.



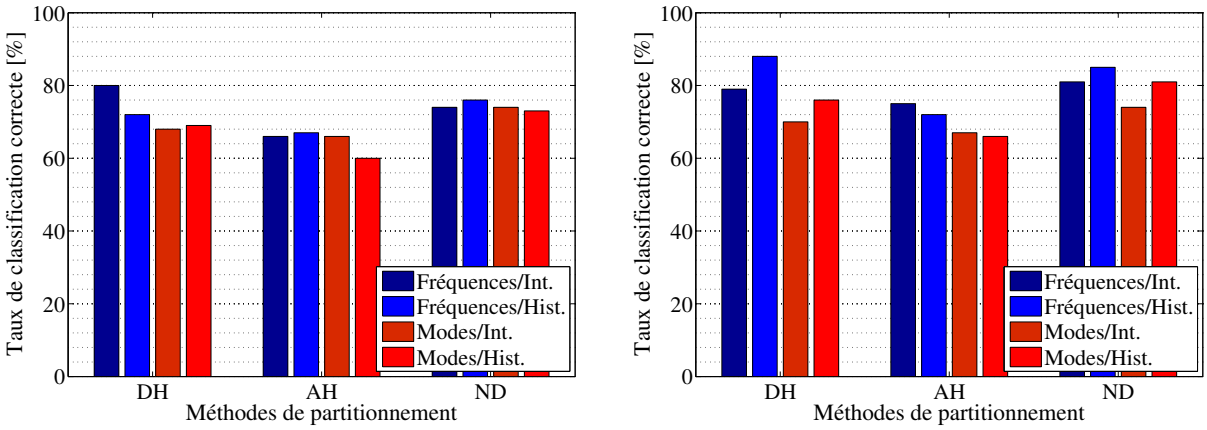
(a) Analyse simultanée

(b) Analyse « par mois »

FIGURE 6.44 – Synthèse des résultats obtenus par les méthodes des divisions hiérarchiques (DH), agglomérations hiérarchiques (AH) et nuées dynamiques (ND).

Par ailleurs, il a été noté que les indices de partitionnement optimal n'ont pas particulièrement changé en utilisant l'analyse mois par mois des essais. Bien entendu, l'analyse menée en considérant tous les essais simultanément tend à écraser les petites variations existantes dans les données, ce qui finalement peut ne pas influencer significativement le calcul des indices. En utilisant la deuxième approche, toutefois, il est possible d'observer la variation de chaque indice mois après mois. Il a été remarqué que pour certains mois (comme ceux de décembre 2008-2009 et avril 2009-2010) les indices n'arrivent pas à détecter de manière adéquate le nombre attendu de clusters. De manière générale, l'application des indices aux fréquences propres a produit des résultats plus pertinents que ceux obtenus en utilisant les signaux ou modes propres. Dans le premier cas, seuls les indices CH et Γ ont pu détecter correctement le nombre de groupes. Dans le deuxième cas, ces indices ont signalé plus fréquemment deux clusters comme étant le partitionnement optimal. Comme il a été déjà constaté plusieurs fois dans ce mémoire, l'indice C^* indique systématiquement un nombre plus important de comportements structuraux qu'en réalité. Il est possible que cet indice soit plus sensible aux variations liées à des changements structuraux plus subtils ou à d'autres effets extérieurs (thermiques, humidité, trafic, etc.).

Deux techniques de régression ont été proposées afin de « corriger » les effets thermiques sur les paramètres modaux identifiés. Contrairement à ce qui a été observé dans l'application expérimentale sur le pont-rail au chapitre 5, le couplage des lois de correction avec les méthodes de partitionnement est apparu pertinent. En considérant soit l'ensemble des essais simultanément, soit l'analyse mois par mois, les taux de classification correcte ont, en général, augmenté après la correction des paramètres modaux, surtout lorsque les filtres non linéaires ont été appliqués. Ces résultats sont pertinents, étant donné que les essais ont été acquis au cours de plusieurs jours et sous des conditions thermiques significativement différentes. Les résultats montrent que les méthodes des divisions hiérarchiques et des nuées dynamiques sont celles qui permettent d'atteindre les meilleurs taux de classification correcte. Les Figures 6.45 et 6.46 illustrent la synthèse des résultats obtenus par le filtre linéaire et non linéaire, respectivement, en considérant l'analyse avec l'ensemble des essais simultanément et celle où les essais sont considérés mois par mois. Bien que les pourcentages de classification correcte soient plus importants avec le filtre non linéaire, ces deux techniques présentent des résultats plus satisfaisants que ceux obtenus sans correction thermique (Figure 6.44).



(a) Analyse simultanée

(b) Analyse par mois

FIGURE 6.45 – Synthèse des résultats obtenus par les méthodes des divisions hiérarchiques (DH), agglomérations hiérarchiques (AH) et nuées dynamiques (ND) appliquées aux paramètres modaux corrigés (filtre linéaire).

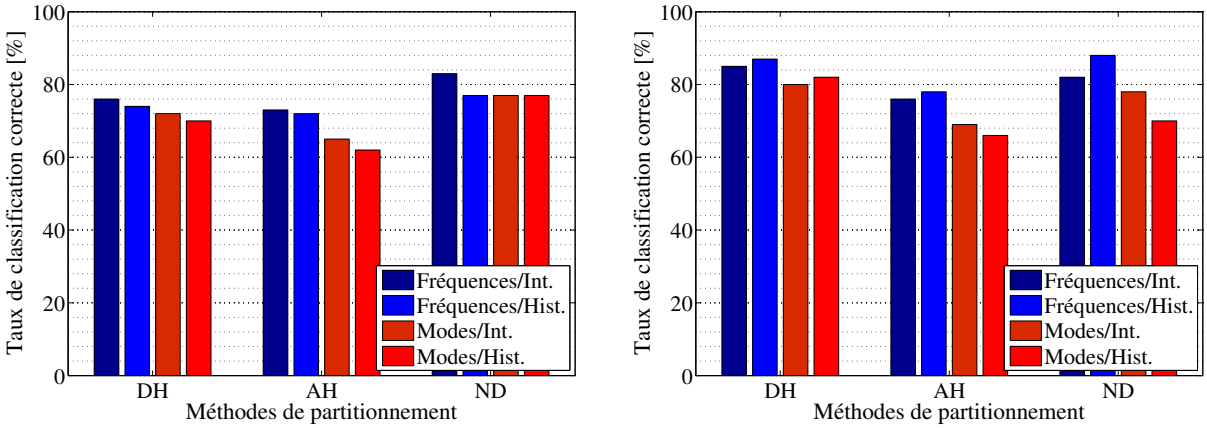
Les indices de partitionnement optimal ont également été calculés après correction thermique. Toutefois, peu de différence a été observé entre les résultats sans ou avec correction thermique. En général, les indices CH et Γ indiquent correctement le nombre de clusters dans la plupart des simulations. Par ailleurs, l'indice C^* oscille relativement moins lorsque la correction thermique est appliquée, en signalant rarement toutefois le nombre attendu de groupes. Ce résultat est intéressant, car avant corriger les effets thermiques des paramètres modaux, l'indice C^* indiquait aussi des comportements thermiques différents (en plus des comportements structuraux).

6.6 ADS couplée aux méthodes de classification supervisée

Comme il a été proposé dans les chapitre 4 et 5, des analyses utilisant le couplage entre l'ADS et les méthodes de classification supervisée sont maintenant présentées. Cette approche est appliquée aux signaux et aux trois premières paires des paramètres modaux identifiés.

Paramètres utilisés

Les simulations utilisant les ADB (Arbres de Décision Bayésiens) sont réalisées sous



(a) Analyse simultanée

(b) Analyse « par mois »

FIGURE 6.46 – Synthèse des résultats obtenus par les méthodes des divisions hiérarchiques (DH), agglomérations hiérarchiques (AH) et nuées dynamiques (ND) appliquées aux paramètres modaux corrigés (filtre non linéaire).

MatLab®. Pour la méthode basée sur les RN (réseaux de neurones), l'architecture du réseau utilisée est composée d'une couche cachée avec 20 neurones. Par ailleurs, les fonctions d'activation utilisées sont la fonction sigmoïde (cf. Équation (3.41)) dans la couche cachée et la fonction Heaviside (cf. Équation 3.40) dans la couche de sortie. Le taux d'apprentissage η est fixé à 0,001 et la constante de moment μ est fixée à 0,0001. Ces paramètres ont été choisis après quelques simulations réalisées au préalable. Dans le cas des SVM (Machines à Vecteurs Supports), la fonction noyau RBF 3.2.3.2 est utilisée et le paramètre σ est déterminé de manière itérative lors de la phase de validation.

Les simulations réalisées dans la suite de ce paragraphe utilisent des groupes d'entraînement, de validation et de test à partir de l'ensemble des essais enregistrés. De façon similaire à celle présentée pour les méthodes de clustering, deux études différentes sont proposées. Dans un premier temps, une analyse statistique est menée, en utilisant l'ensemble des 2490 essais et en simulant différents groupes d'entraînement, validation et test. Dans ce cas, deux classes sont créées, chacune représentant un état structural (i.e., avant et après les travaux de renforcement). Ainsi, 10000 simulations sont réalisées en considérant trois configurations :

- 30% (747) d'essais pour l'entraînement, 10% (249) pour la validation et 60% (1494) pour le test,
- 40% (996) d'essais pour l'entraînement, 10% (249) pour la validation et 50% (1245) pour le test,
- 50% (1245) d'essais pour l'entraînement, 10% (294) pour la validation et 40% (996) pour le test.

Dans un second temps, la même procédure est menée, mais en considérant une analyse mois par mois, comme déjà présentée. Les trois configurations proposés ci-dessus sont alors appliqués aux ensembles des essais enregistrés dans chaque mois. Pour le mois d'octobre 2008 et 2009, 182 essais ont été acquis ; 444 essais en novembre 2008/2009 ; 458 en décembre 2008/2009 ; 455 en janvier 2009/2010 ; 386 en février 2009/2010 ; 439 en mars 2009/2010 et 126 en avril 2009/2010.

6.6.1 Analyse sur l'ensemble des essais

Cette première étude considère la totalité des essais enregistrés. Une étude statistique est menée en faisant varier les dimensions des groupes d'entraînement et test. Au total, 10000 simulations sont réalisées pour chaque configuration proposée et les meilleurs taux de classification correcte, ainsi que les taux maximum et minimum, sont évalués.

6.6.1.1 Signaux

Les premières simulations correspondent à l'application des méthodes de classification aux signaux transformés en histogrammes (cf. Tableau 6.31). Les résultats montrent que les méthodes des RN et des SVM sont plus robustes pour la classification des essais que celles utilisant les ADB. De façon générale, la méthode des RN permet d'atteindre des taux de classification correcte les plus élevés. Dans cette étude, il est noté qu'en général, l'augmentation du nombre d'essais dans le groupe d'entraînement n'altère pas significativement les taux de classification correcte, notamment pour les méthodes des RN et des SVM. Cela peut indiquer que l'application de ces méthodes aux transformations en données symboliques est suffisamment discriminante, même si peu d'essais sont utilisés dans la phase d'apprentissage. Les pourcentages de réussite sont relativement plus faibles que ceux utilisant les intervalles, notamment pour la méthode des ADB (cf. Tableau C.26). Il est possible que ce type de transformation (dans cette application) ne soit pas le plus pertinent pour mettre en évidence les différences entre les deux états structuraux. De plus, bien que les RN et SVM réussissent à classer 100% des essais pendant une (ou plusieurs) simulation(s), les taux moyens sont légèrement plus faibles que ceux utilisant la transformation en intervalles interquartiles.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	94	99	96	93	100	97	90	100	100
Moyen	65	92	92	69	95	92	74	95	91
Minimal	31	86	89	45	88	86	50	90	85

TABLEAU 6.31 – Taux de classification correcte utilisant les signaux transformés en histogrammes pour l'ensemble total des essais (valeurs en %).

6.6.1.2 Fréquences propres

Par ailleurs, le Tableau 6.32 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en histogrammes est appliquée aux fréquences propres. Il est intéressant de noter que dans cette application, les taux de classification correcte sont pratiquement du même ordre que ceux obtenus en utilisant les signaux. Les méthodes basées sur les RN et SVM arrivent toujours à mieux classer les essais, quel que soit le nombre d'observations utilisées dans le groupe d'entraînement. En général, les résultats sont légèrement meilleurs pour la méthode des SVM. Toutefois, pour les ADB et RN, les taux de classification correcte sont soit du même ordre, soit plus faibles.

Les simulations suivantes correspondent à l'application des méthodes de classification aux fréquences propres corrigées par le filtre linéaire. Le Tableau 6.33 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en histogrammes est considérée. En général, les méthodes de classification supervisée permettent d'atteindre des résultats aussi satisfaisantes que ceux obtenus sans filtre thermique (cf. Tableau 6.32). La différence est notamment

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	92	100	100	90	100	100	92	100	100
Moyen	80	95	95	81	95	95	82	96	97
Minimal	71	88	76	74	86	84	79	88	83

TABLEAU 6.32 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour l'ensemble total des essais (valeurs en %).

notée sur les taux maximal et minimal qui ont une légère augmentation après la correction thermique. En effet, ces résultats montrent que les méthodes de classification supervisée sont, d'une certaine façon, plus robustes vis-à-vis des variations thermiques. Étant donné que ces algorithmes sont basés sur des algorithmes d'apprentissage, il est possible qu'ils arrivent à bien définir des seuils de séparation entre un état structural et un autre, même si les mesures ne sont pas filtrées. Ceci est notamment observé pour les méthodes des RN et des SVM.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	83	93	97	86	92	97	90	91	94
Moyen	81	94	91	80	92	95	81	94	93
Minimal	72	92	74	74	89	83	77	86	84

TABLEAU 6.33 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour l'ensemble total des essais (filtre linéaire ; valeurs en %).

L'utilisation du filtre non linéaire est maintenant considérée. Les résultats obtenus en considérant la transformation des fréquences propres en histogrammes sont détaillés dans le Tableau 6.34. Il est noté que les taux de classification pour les ADB sont significativement meilleurs que ceux obtenus sans l'utilisation des lois de correction thermique. Ces résultats montrent une meilleure adéquation du filtre non linéaire pour la correction des fréquences propres. Les classifications sont plutôt les mêmes que celles obtenues avec le filtre linéaire, avec une légère amélioration des taux. Toutefois, les ADB sont encore associées à un taux de classification correcte inférieur à 90%.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	89	100	104	86	96	97	96	98	98
Moyen	90	97	90	86	96	96	82	97	95
Minimal	68	94	76	73	91	79	79	85	81

TABLEAU 6.34 – Taux de classification correcte utilisant les fréquences propres transformées en histogrammes pour l'ensemble total des essais (filtre non linéaire ; valeurs en %).

6.6.1.3 Modes propres

Une démarche similaire est considérée en appliquant les méthodes de classification supervisée aux modes propres. Les résultats obtenus en utilisant la transformation en histogrammes sont détaillés dans le Tableau 6.35. Dans cette application, les taux moyens de classification correcte sont pratiquement du même ordre que ceux évalués pour les fréquences propres. Ceci montre que l'approche basée sur l'ADS appliquée aux modes propres peut être utilisée pour la discrimination des modifications structurales. Comme déjà remarqué, les méthodes basées sur les RN et

SVM sont celles avec les taux de classification correcte les plus importants. Toutefois, les taux minimal et maximal sont relativement plus importants comparés à ceux obtenus en utilisant la représentation par des intervalles (cf. Tableau C.30).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	88	95	95	88	96	96	89	96	96
Moyen	75	91	90	76	92	91	78	93	91
Minimal	62	86	79	66	88	82	69	89	83

TABLEAU 6.35 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour l'ensemble total des essais (valeurs en %).

Des simulations sont également réalisées en utilisant les modes propres corrigés par le filtre linéaire. Les taux de classification correcte sont pratiquement du même ordre que ceux trouvés avec les fréquences propres et relativement plus élevés que ceux obtenus sans le filtre. De plus, quel que soit le type de transformation en données symboliques utilisé, les résultats obtenus par les ADB et les RN ont évolué par rapport à ceux obtenus sans prise en compte des effets thermiques (cf. Tableau 6.35).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	86	95	98	87	96	97	90	94	96
Moyen	82	94	92	81	93	93	81	94	92
Minimal	68	90	77	70	90	80	74	85	83

TABLEAU 6.36 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour l'ensemble total des essais (filtre linéaire ; valeurs en %).

Le Tableau 6.37 regroupe les taux de réussite calculés en considérant les déformées modales par le filtre non linéaire. Dans ce cas, les améliorations apportées sont moins évidentes. Dans le cas de la transformation en histogrammes, seule l'utilisation des RN et SVM a permis de mieux classer les essais dans les deux groupes. Dans ce cas, les RN permettent d'atteindre des taux plus importants que ceux obtenus par le filtre linéaire ou encore sans utilisation des filtres.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	87	95	99	86	97	94	90	95	98
Moyen	84	92	91	83	97	94	80	96	94
Minimal	70	90	78	72	92	80	76	86	84

TABLEAU 6.37 – Taux de classification correcte utilisant les modes propres transformés en histogrammes pour l'ensemble total des essais (filtre non linéaire ; valeurs en %).

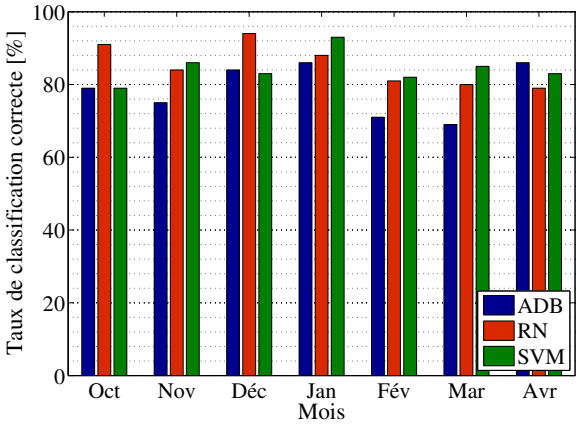
6.6.2 Comparaison des phases 1 et 2 mois par mois

La démarche basée sur l'analyse mois par mois des essais est considérée en utilisant les méthodes de classification supervisée. L'objectif est d'observer si les taux de classification correcte peuvent être améliorés. Dans le soucis de clarté, les résultats sont présentés de façon synthétique au travers de figures. Étant donné le nombre conséquent de simulations réalisées, seuls les taux moyens évalués pour chaque méthode sont montrés.

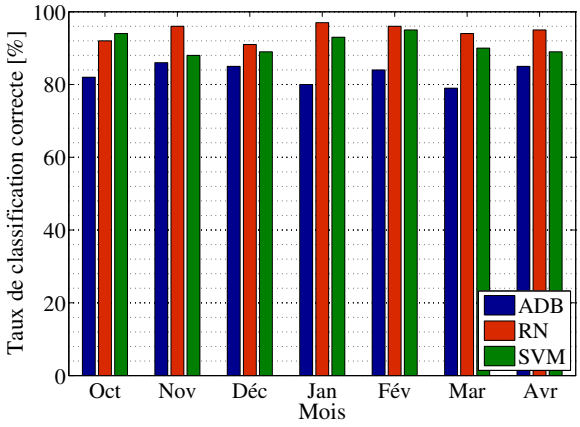
L'approche proposée est d'abord appliquée aux signaux transformés en histogrammes. La Figure 6.47(a) synthétise les résultats obtenus lorsque seulement 30% des essais (chaque mois séparément) sont retenus dans le groupe d'entraînement. Il est noté que les méthodes des RN et des SVM sont toujours les plus performantes. Par ailleurs, contrairement à ce qui a été observé dans le paragraphe 6.5.1 (avec la totalité des essais), la transformation en histogrammes semble être sensiblement plus adaptée que celle en intervalles (cf. Figure C.20(a)). En effet, il se peut que le couplage de ces méthodes à la représentation par des histogrammes soit plus sensible aux effets extérieurs (trafic, température, vent, etc.), ne permettant pas une discrimination adéquate des états structuraux. Toutefois, en utilisant l'analyse mois par mois, la diminution possible de ces effets a permis d'atteindre de meilleurs taux de classification. Les Figures C.18(a) et C.19(a) montrent les valeurs moyennes de classification correcte obtenues pour chaque méthode en considérant les groupes d'entraînement ayant 40% et 50% des essais, respectivement. Il est noté une sensible augmentation des taux moyens pour toutes les méthodes, à l'exception des ADB appliqués aux essais du mois d'octobre. En général, les taux moyens de classification sont de l'ordre de 70-80% en considérant les trois méthodes et les trois configurations de groupes simulées.

Une démarche similaire est appliquée aux fréquences propres en considérant les deux types de transformation en données symboliques. Les résultats obtenus en utilisant 30% des essais dans le groupe d'entraînement sont montrés à la Figure 6.47(b). Il est noté que les taux de classification correcte sont plus importants que ceux obtenus en utilisant les signaux. Par ailleurs, la transformation en histogrammes mène à des résultats relativement plus pertinents, notamment si les ADB sont considérés. En augmentant le nombre d'essais utilisés dans l'étape d'entraînement des méthodes, les taux de classification correcte augmentent significativement. La Figure C.18(b) regroupe les pourcentages moyens évalués en considérant 40% des essais pour l'entraînement. Il est noté que pour le mois d'octobre, la méthode des RN peut classer parfaitement tous les essais dans les simulations réalisées. Les résultats obtenus en considérant la moitié des essais dans le groupe d'entraînement sont présentés à la Figure C.19(b). Les taux de classification correcte sont relativement plus élevés et, en général, l'utilisation des histogrammes mène à de meilleurs résultats que ceux avec les intervalles interquartiles (cf. Figure C.22(b)).

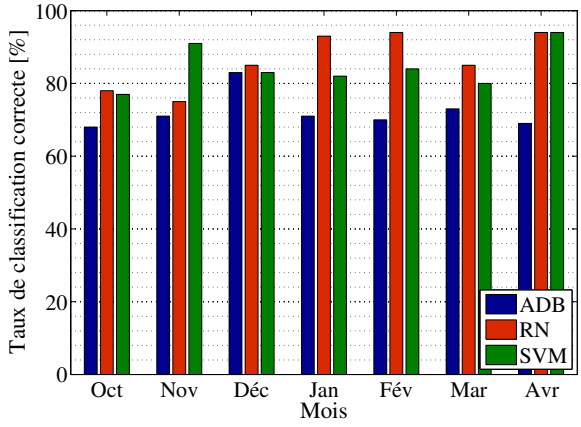
Finalement, les méthodes de classification sont appliquées aux modes propres. Comme il a été observé dans l'étude précédente et contrairement à ce qui a été remarqué dans les chapitres précédents, les taux de classification sont aussi importants que ceux obtenus avec des fréquences propres. Il est possible de conclure que l'ADS appliquée aux déformées modales a permis de mieux détecter les variations liées aux travaux de renforcement réalisés sur l'ouvrage (ce qui n'a pas été le cas de l'application sur le pont-rail au chapitre 5). La Figure 6.47(c) présente les taux moyens de classification correcte évalués en utilisant la première configuration des simulations. Il est noté que la transformation en histogrammes mène à des résultats légèrement plus adéquats, surtout si les SVM sont considérées. Par ailleurs, la Figure C.18(c) illustre les résultats obtenus en utilisant 40% des essais dans l'étape d'entraînement. Dans ce cas, il est observé que les RN et SVM permettent d'atteindre des pourcentages plus importants. Toutefois, les ADB ne classent, en moyenne, que 70% des essais environ quel que soit le type de transformation symbolique considéré. Les résultats sont plutôt pertinents pour le mois de février, où pratiquement tous les essais sont classés correctement. Les taux de classification correcte sont, en général, significativement meilleurs en considérant la moitié des essais dans le groupe d'entraînement. La Figure C.19(c) montre les pourcentages évalués pour chaque mois et chaque méthode. De manière générale, les taux de classification correcte en utilisant les modes propres sont de l'ordre de 80-85%.



(a) Signaux



(b) Fréquences propres



(c) Modes propres

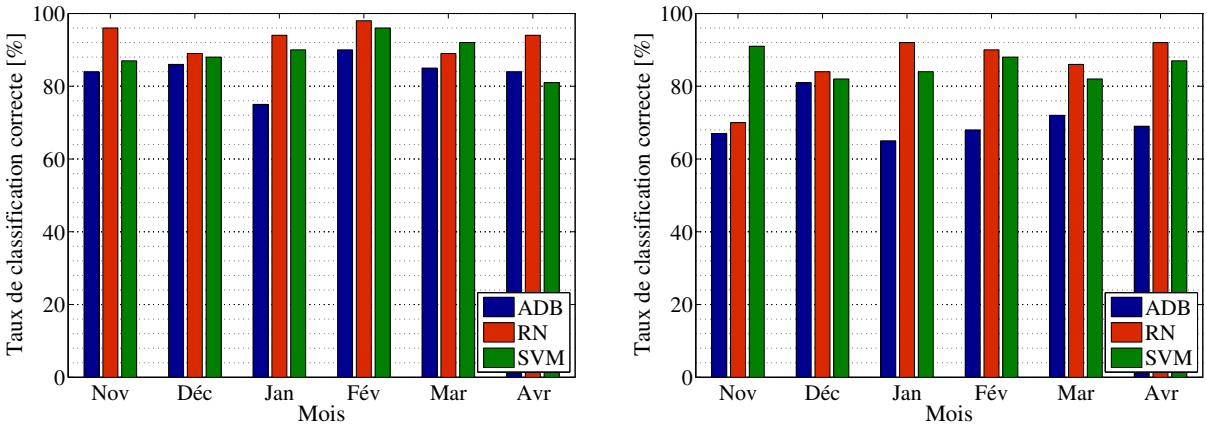
FIGURE 6.47 – Méthodes de classification supervisée appliquées aux signaux et paramètres modaux transformés en histogrammes (analyse mois par mois ; 30% entr., 10% val., 60% test ; valeurs en %).

6.6.2.1 Application aux paramètres modaux corrigés

L'objectif ici est d'observer si les taux de classification correcte peuvent être améliorés par rapport à ceux calculés au paragraphe précédent sans prise en compte des effets thermiques. Dans un premier temps, l'approche proposée est appliquée aux fréquences propres corrigées par le filtre linéaire. La Figure 6.48(a) synthétise les résultats obtenus lorsque seulement 30% des essais (pour chaque mois) sont retenus dans le groupe d'entraînement, en considérant la transformation en histogrammes. Il est noté que, dans la plupart des simulations, les méthodes des RN et des SVM sont les plus performantes. Il est remarqué que les SVM arrivent à classer tous les essais du mois de janvier correctement dans toutes les simulations réalisées. En général, les taux de classification correcte obtenus en utilisant des histogrammes sont relativement plus faibles que ceux obtenus avec les intervalles, notamment pour les ADB (cf Figure C.27(a)). Par ailleurs, comme il a été observé dans l'étude précédente, lorsque tous les essais ont été considérés simultanément, les résultats obtenus ne sont pas très différents de ceux obtenus sans utiliser de filtres thermiques. Cela peut être dû au fait que le filtre linéaire n'arrive pas à modifier significativement les valeurs des fréquences propres identifiées. En ce sens, la détermination des seuils de séparation entre classes reste pratiquement inchangée. Les Figures C.23(a) et C.24(a) montrent les valeurs moyennes de classification correcte obtenues pour chaque méthode, en considérant les groupes d'entraînement

avec 40% et 50% des essais, respectivement. Il est noté une légère augmentation des taux moyens pour toutes les méthodes, à l'exception des SVM appliquées aux essais du mois de janvier. Ceci pourrait être expliqué par le fait que les groupes d'essais définis pour l'entraînement n'ont pas pu déterminer de manière adéquate les marges de séparation entre les classes.

Les méthodes de classification sont maintenant appliquées aux modes propres corrigés par le filtre linéaire. Les taux de classification sont relativement moins pertinents que ceux obtenus avec des fréquences propres. La Figure 6.48(b) présente les taux moyens de classification correcte évalués en utilisant la première configuration des simulations. Il est noté que la transformation en histogrammes produit des résultats moins satisfaisants, surtout si les ADB sont considérés. De manière générale, les taux moyens de classification correcte varient entre 70% et 95%, sachant que les RN arrivent à mieux classer les essais. De plus, les pourcentages évalués en utilisant le filtre linéaire sont similaires à ceux évalués dans l'étude précédente. La Figure C.23(b) montre les résultats obtenus en utilisant 40% des essais dans l'étape d'entraînement. Il est observé que les toutes les méthodes permettent d'atteindre des pourcentages plus importants. Par ailleurs, les résultats sont, en général, significativement meilleurs en considérant la moitié des essais dans le groupe d'entraînement. La Figure C.24(b) montre les pourcentages évalués pour chaque mois et chaque méthode. De manière générale, les taux de classification correcte en utilisant les modes propres sont de l'ordre de 85% à 90%.



(a) Fréquences propres

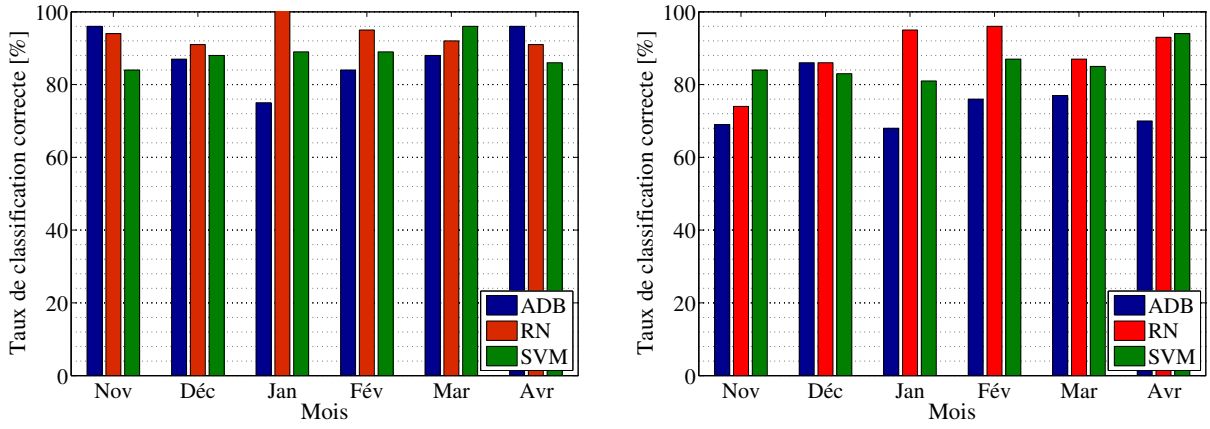
(b) Modes propres

FIGURE 6.48 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en histogrammes (filtre linéaire ; analyse mois par mois ; 30% entr., 10% val., 60% test ; valeurs en %).

L'étape suivante consiste à appliquer les méthodes de classification aux fréquences propres corrigées par le filtre non linéaire. Dans ce cas et en considérant 30% des essais dans le groupe d'entraînement, les résultats sont relativement meilleurs que ceux obtenus par le filtre linéaire (Figure 6.49(a)). En général, il est noté que ce type de filtre semble être le plus adéquat pour cette analyse. Bien que les RN et SVM arrivent à mieux classer les essais, il est noté que les taux de classification correcte pour les ADB sont relativement plus importants (notamment pour les mois de novembre et décembre). La transformation en histogrammes mène à des résultats légèrement meilleurs pour les RN, mais les pourcentages sont plus faibles pour les ADB. La Figure C.25(a) illustre les résultats obtenus en considérant 40% d'essais dans l'étape d'entraînement des méthodes. En général, les taux de classification correcte sont essentiellement les mêmes. Inversement, la transformation en histogrammes semble produire des taux plus faibles pour les ADB. En effet, tout dépend des essais choisis pour l'apprentissage des algorithmes. Si ce sont des essais avec un comportement « ambigu » (où la différence entre les fréquences de la phase 1 et 2

n'est pas suffisamment nette), il est tout à fait possible que les résultats dans l'étape de test ne soient pas satisfaisants. Par ailleurs, la Figure C.26(a) illustre les taux moyens calculés lorsque la moitié des essais est comprise dans le groupe d'entraînement. Il est noté que les résultats sont significativement meilleurs si comparés à ceux du filtre linéaire (ou encore sans utiliser aucun filtre).

Finalement, les déformées modales corrigées par le filtre non linéaire sont considérées. La Figure 6.49(b) illustre les résultats obtenus lors des premières simulations réalisées. Il est intéressant de noter que les taux de classification correcte ne changent pratiquement pas par rapport à ceux évalués en utilisant le filtre linéaire ou encore sans aucun filtre. Dans ce cas, il est possible de dire que les fréquences propres sont plus sensibles aux variations thermiques que les déformées propres. En général, les meilleurs taux sont obtenus par les RN (surtout au mois de janvier). La transformation en histogrammes rend les classifications relativement plus pertinentes que celles obtenus avec les intervalles (cf. Figure C.30(b)). De plus, en augmentant le nombre d'essais dans le groupe d'entraînement, les taux sont légèrement plus élevés (Figure C.25(b)). Cela est aussi observé lorsque 50% des essais sont utilisés pour la phase d'entraînement des méthodes, comme montré à la Figure C.26(b). Pour ces dernières simulations, la transformation en intervalles interquartiles permet, de manière générale, d'obtenir de meilleurs résultats. De plus, comme il a été remarqué dans toutes les simulations réalisées, les méthodes basées sur les RN et SVM sont celles capables de distinguer des états structuraux avec les meilleurs taux de réussite.



(a) Fréquences propres

(b) Modes propres

FIGURE 6.49 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en histogrammes (filtre non linéaire ; analyse mois par mois ; 30% entr., 10% val., 60% test ; valeurs en %).

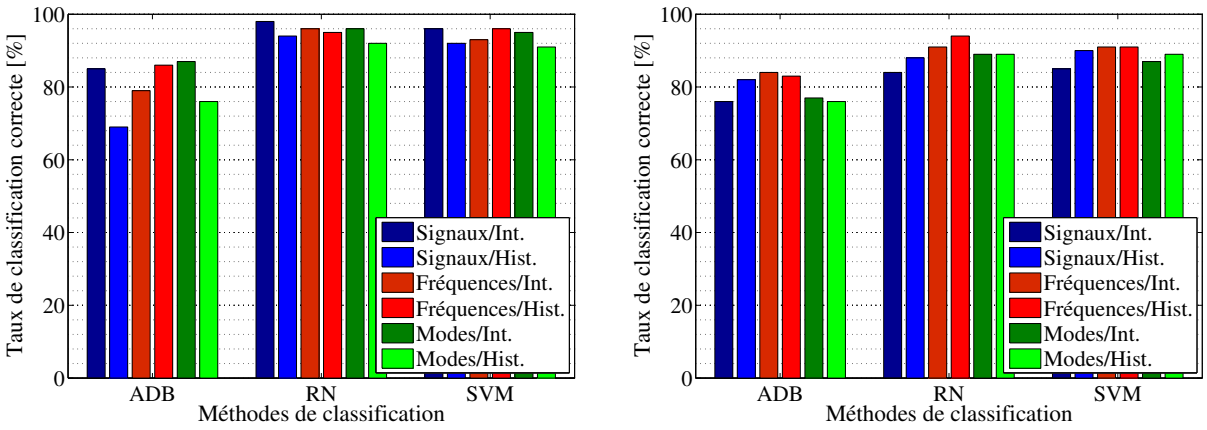
6.6.3 Bilan

Dans les paragraphes précédents, les méthodes de classification supervisée ont été appliquées aux signaux et paramètres modaux identifiés lors des campagnes de mesures réalisées sur pont sur l'Oise. Dans un premier temps, l'approche basée sur le couplage de l'ADS et les ADB, RN et SVM a été appliquée à l'ensemble des 2490 essais enregistrés. L'objectif a été de les discriminer en deux groupes : le premier, correspondant à l'état avant le renforcement et le deuxième, après les travaux. Les résultats ont montré que les taux de classification ont été significativement plus importants que ceux obtenus en utilisant les méthodes clustering. En particulier, les méthodes des RN et des SVM ont réussi à classer correctement tous les essais dans certaines simulations lorsque les signaux sont considérés. Ces résultats montrent que lorsque des outils adéquats sont

utilisés, les réponses dynamiques des ouvrages peuvent être utilisées pour la discrimination d'états structuraux. Cette approche a été également appliquée aux paramètres modaux. En considérant les fréquences propres, les trois méthodes ont pu indiquer, avec un taux de classification correcte supérieur à 90%, l'état structural de chaque essai. Par ailleurs, contrairement à ce qui a été observé lors des applications précédentes, l'utilisation des modes propres a aussi produit des résultats pertinents.

Dans un deuxième temps, la même démarche a été appliquée aux essais dynamiques triés mois par mois. Il a été observé qu'en utilisant les méthodes de clustering, les taux de classification correcte augmentent sensiblement. Toutefois, cela n'a pas été le cas pour les méthodes de classification supervisée. En général, les taux moyens globaux (en considérant tous les résultats pour tous les mois ensemble) ont été plus faibles que ceux obtenus en utilisant la totalité des essais en même temps. En effet, ces méthodes sont basées sur une procédure d'apprentissage à partir de laquelle les algorithmes sont « calibrés » après avoir vu un certain nombre d'observations. Il est tout à fait possible que, en considérant la totalité d'essais, les groupes d'entraînement soient plus représentatifs de chaque état structural et permettent donc un ajustement plus adéquat des paramètres relatifs à chaque méthode. En ce sens, la détermination des seuils pour la séparation des états structuraux peut être plus efficace que celle ne considérant qu'un sous-ensemble d'essais. Bien que les taux de réussite aient été relativement élevés dans certains mois, ceci n'a pas été observé de manière globale.

La Figure 6.50(a) illustre une synthèse des résultats évalués à partir du calcul de la moyenne globale (pour les deux états structuraux) des taux de classification obtenus pour chaque méthode et chaque type de transformation utilisée. En considérant tous les essais pris ensemble, les méthodes des RN et des SVM ont été les plus performantes en utilisant soit les signaux ou les paramètres modaux. Dans ces deux cas, des taux de classification correcte de l'ordre de 90% ont été obtenus. Par ailleurs, la Figure 6.50(b) illustre les moyennes globales (en considérant tous les mois) des pourcentages de classification correcte évalués à partir de l'analyse des essais mois par mois. Dans ce dernier cas, il est remarqué que toutes les méthodes permettent d'atteindre des taux relativement plus faibles.



(a) Analyse simultanée

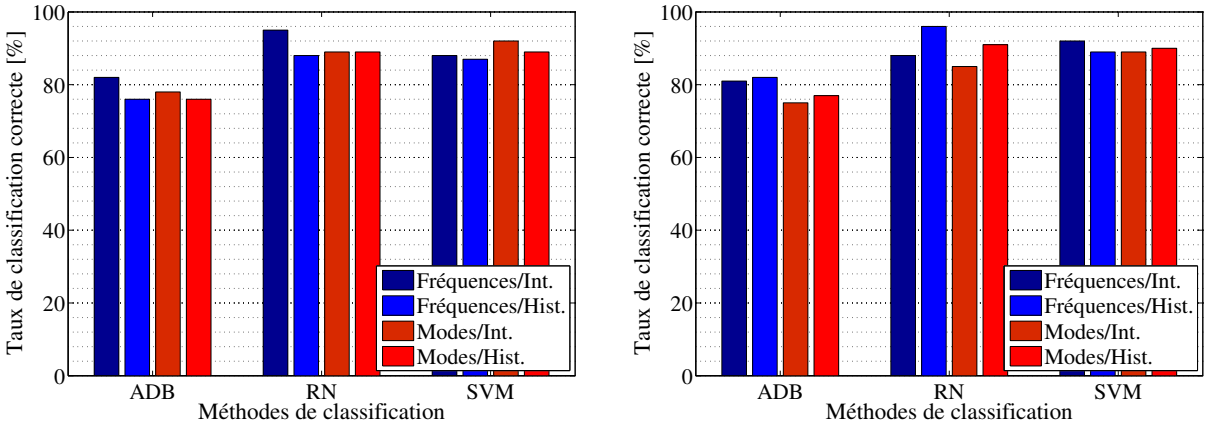
(b) Analyse « par mois »

FIGURE 6.50 – Synthèse des résultats obtenus par les méthodes de classification supervisée.

Les méthodes de classification supervisée ont été également appliquées aux paramètres modaux « compensés » des effets thermiques. Cela a été réalisé à partir de l'utilisation des techniques de régression linéaire et non linéaire proposées dans le paragraphe 3.4. Les résultats obtenus ont montré que la mise en œuvre de cette démarche est pertinente dans le cadre de l'étude considérée. En effet, il a été observé que les taux de classification correcte ont eu une significative améliora-

tion lorsque les filtres thermiques ont été appliqués aux fréquences et modes propres. Même que la différence sur les résultats ne soit pas évidente dans le cas du filtre linéaire (les pourcentages sont du même ordre sans ou avec correction thermique), les modifications obtenues en appliquant le filtre non linéaire sont plus nettes. Ces résultats montrent que la prise en compte des effets de la variation de la température n'est pas négligeable dans cet ouvrage. Dans le cas des méthodes de classification supervisée, il a été constaté que les fréquences propres sont plus influencées par les changements thermiques que les déformées modales, car les taux de classification correcte pour les fréquences propres ont relativement augmenté après la correction thermique (ce qui a été moins évident pour les modes propres).

Par ailleurs, en considérant soit l'analyse de l'ensemble d'essais, soit l'analyse mois par mois, les taux de classification correcte ont généralement amélioré. Toutefois, contrairement à ce qui a été observé avec les méthodes de clustering, l'analyse de tous les essais simultanément a mené à des meilleurs résultats. Ceci peut être dû à la capacité des méthodes de classification supervisée de mieux déterminer des seuils de séparation entre classes, même en considérant de grandes bases de données. Les résultats montrent que les méthodes basées sur les RN et SVM sont celles qui permettent d'atteindre les taux de classification correcte les plus élevés. Les Figures 6.51 et 6.52 illustrent la synthèse des taux moyens globaux obtenus à partir de l'utilisation du filtre linéaire et non linéaire, respectivement. Par ailleurs, un comparatif des résultats en considérant l'analyse de l'ensemble des essais et la démarche mois par mois est aussi fourni. Bien que les pourcentages de classification correcte soient plus importants pour le filtre non linéaire, il est noté que ces deux techniques présentent des résultats relativement plus pertinents que ceux obtenus sans la correction thermique Figure 6.50).



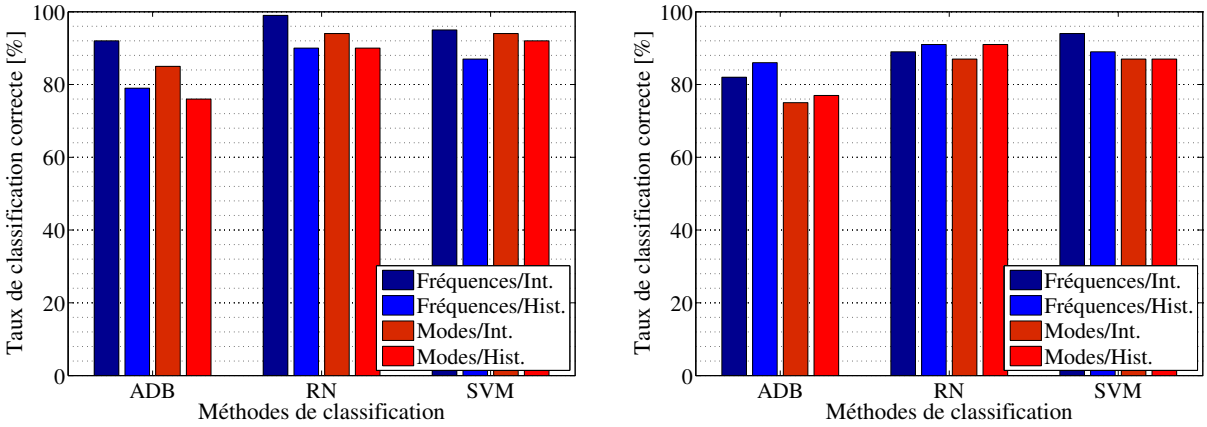
(a) Analyse simultanée

(b) Analyse mois par mois

FIGURE 6.51 – Synthèse des résultats obtenus par les méthodes de classification supervisée appliquées aux paramètres modaux corrigés (filtre linéaire).

6.6.4 Conclusions

Ce présent chapitre a porté sur l'étude d'une application expérimentale basée sur les mesures dynamiques acquises sur le pont PI-57. Cet ouvrage en béton précontraint et de longue portée permet à l'autoroute A1, reliant Paris à Lille, de franchir l'Oise. Pendant l'été 2009, ce pont a subi des travaux de renforcement qui ont consisté à ajouter des câbles de précontraint permettant de rigidifier l'ouvrage. Ainsi, comme présenté dans le chapitre précédent, cette étude a été divisée en deux axes principaux. Le premier objectif a été d'apprécier les effets des travaux réalisés à partir du suivi des caractéristiques dynamiques du pont, tels que la variation des fréquences



(a) Analyse simultanée

(b) Analyse « par mois »

FIGURE 6.52 – Synthèse des résultats obtenus par les méthodes de classification supervisée appliquées aux paramètres modaux corrigés (filtre non linéaire).

propres, des taux d'amortissement et des déformées modales. Le deuxième objectif a été d'évaluer l'impact des effets environnementaux (principalement de la variation des températures) sur les fréquences et les modes propres. De nombreux essais dynamiques ont été enregistrés durant les deux campagnes. Dans la première, réalisée entre octobre 2008 et avril 2009, 1174 ont été acquis et dans la deuxième, ayant eu lieu entre octobre 2009 et avril 2010, 1316 essais ont été enregistrés.

Pour répondre au premier objectif, des études portant sur l'analyse classique des paramètres modaux ont été menées. En premier lieu, un comparatif entre les histogrammes des fréquences propres identifiées avant et après renforcement a été réalisé. Par ailleurs, des intervalles de confiance pour les valeurs moyennes ont été calculés. Les résultats obtenus ont montré que cette approche n'est pas suffisamment sensible pour détecter les modifications structurales imposées, étant donné que les intervalles de confiance sont superposés. Des tests d'hypothèses ont été aussi réalisées, le but étant de déterminer si les distributions des fréquences propres identifiées pendant la phase 1 (avant les travaux) et la phase 2 (après les travaux) peuvent être considérées comme identiques. Les résultats ont montré que quelques fréquences propres ont échoué à ce test, en indiquant que les deux états peuvent être distingués. De plus, les coefficients MAC ont été évalués pour les déformées modales, en considérant là aussi les deux phases d'instrumentation. Il a été montré que ces coefficients ont varié entre 95% et 100% pour les deux premiers modes et 85% environ pour le troisième. Ces valeurs indiquent une faible sensibilité de ce type d'analyse pour la discrimination de modifications structurales. De manière générale, les résultats obtenus à partir des approches dites classiques ne permettent pas de tirer des conclusions définitives sur le comportement de l'ouvrage.

L'approche basée sur le couplage de l'ADS aux méthodes de clustering (divisions hiérarchiques, agglomérations hiérarchiques et nuées dynamiques) a d'abord été considérée. Afin de vérifier l'adéquation et la robustesse de la démarche proposée, deux études ont été menées. Dans un premier temps, la totalité des essais (2490) enregistrés pendant les deux campagnes de mesures ont été considérés de manière simultanée. L'objectif était de discriminer les essais de la première phase de ceux de la deuxième phase. Des analyses ont été réalisées en considérant les signaux et les paramètres modaux représentés par des intervalles interquartiles et des histogrammes. Il a été observé que les taux de classification correcte ont été relativement faibles, même lorsque les fréquences propres ont été considérées. En général, les méthodes des divisions hiérarchiques et des nuées dynamiques ont été plus pertinentes pour la distinction des états structuraux en utilisant soit les signaux, soit les paramètres modaux. Afin de compléter cette étude, une ana-

lyse plus fine a été menée. Cette analyse a consisté à classer les essais séparément, triés mois par mois. En d'autres termes, au lieu de mener une étude où tous les essais des phases 1 et 2 sont considérés au même temps, le but a été de comparer les essais enregistrés dans un mois de l'année 2008 avec ceux du même mois de l'année 2009. Les résultats obtenus ont montré que l'utilisation de cette démarche a permis une meilleure classification des essais, notamment pour le mois d'octobre. Dans ce cas, en utilisant les signaux transformés en histogrammes, 100% des essais ont été correctement classés. Ce résultat est intéressant, car il montre qu'il est possible d'extraire des informations sur le comportement de l'ouvrage à partir des mesures effectuées *in situ*. Par ailleurs, l'analyse des indices de partitionnement optimal appliqués aux méthodes de clustering a été considérée. Comme il a été observé lors des simulations numériques et expérimentales, les indices CH et Γ ont été, en général, être les plus fiables. Ces indices ont pratiquement toujours indiqué le nombre attendu de groupes distincts, notamment en utilisant les paramètres modaux. L'indice C^* a été le plus sensible aux variations structurales, en signalant souvent un nombre de groupes de comportement plus élevé. Les indices ont été également calculés pour les partitionnements obtenus par l'analyse mois par mois. Les résultats ont été essentiellement les mêmes, étant donné que les indices CH et Γ ont souvent indiqué l'existence de deux clusters.

De plus, la démarche basée sur le couplage de l'ADS aux méthodes de classification supervisée (arbres de décision Bayésiens, réseaux de neurones et machines à vecteur support) a été considérée. De façon similaire à celle présentée pour les méthodes de clustering, deux études ont été menées afin d'observer la performance des ces méthodes pour la discrimination de comportements structuraux. Contrairement à ce qui a été observé auparavant, les résultats obtenus en utilisant l'ensemble des essais ont été significativement meilleurs que ceux obtenus en considérant l'analyse mois par mois des essais. En effet, ces méthodes sont basées sur une procédure d'apprentissage à partir de laquelle les algorithmes sont « calibrés » après avoir vu un certain nombre d'observations. Il est donc possible, en considérant la totalité d'essais, que les groupes d'entraînement soient plus « représentatifs », et permettent alors une meilleure définition des seuils de séparation entre les états structuraux. Les RN et les SVM sont les méthodes qui ont permis d'obtenir des résultats les plus pertinents. Il a été remarqué que même en utilisant les signaux, près de 90% ont été classés correctement. Les ADB, toutefois, ont produit des taux de classification relativement plus faibles. En ce qui concerne l'analyse des essais mois par mois, bien que pour certaines périodes les taux de réussite aient été relativement importants, cela n'a pas été observé de manière globale.

Afin d'apporter des nouveaux éléments de réponse sur les effets de la variation des températures sur les paramètres modaux, deux techniques de régression de données - une à caractère linéaire multiple et l'autre basée sur des réseaux de neurones - ont été proposées. Une étude préliminaire a montré que la variation de la première fréquence propre lors de la première campagne de mesures a été fortement influencée par un changement abrupt de température. Ceci a donc amené à une analyse plus approfondie des effets thermiques sur les paramètres modaux. La procédure a d'abord consisté à « corriger » les fréquences propres et les modes propres identifiés pendant les deux phases d'instrumentation. Cela a été réalisé à partir des modèles de régression linéaire et non linéaire, permettant de filtrer les effets de la variation des températures. En second lieu, les paramètres modaux, « compensés » des effets thermiques, ont été appliqués aux méthodes de clustering et de classification supervisée. Comme déjà mentionné, l'objectif a été d'observer comment les résultats obtenus précédemment pouvaient changer en tenant en compte des effets thermiques. Dans l'ensemble des simulations réalisées en tenant compte des méthodes de clustering, il a été remarqué que les taux de classification correcte ont été relativement plus importants. En réalité, l'utilisation du filtre non linéaire s'est avérée plus pertinente que celle du filtre linéaire. Dans les deux études - en analysant les essais ensemble ou mois par mois - les résultats ont été meilleurs que ceux sans correction thermique. Par ailleurs, les indices de partitionnement optimal ont aussi été calculés. Peu de modification a été remarqué en considérant

les indices CH et Γ. Ces derniers ont fréquemment indiqué correctement le nombre de clusters dans la plupart des simulations réalisées. Par ailleurs, il a été observé que l'indice C^* a oscillé relativement moins que dans l'étude sans correction thermique. En effet, cela montre que l'indice C^* peut indiquer aussi des comportements thermiques différents (en plus des comportements structuraux). Les résultats obtenus par les méthodes de classification supervisée n'ont pas montré une amélioration évidente. Bien que la différence dans les résultats n'ait pas été nette dans le cas du filtre linéaire (où les pourcentages ont été pratiquement du même ordre), les améliorations obtenues par le filtre non linéaire ont été relativement plus marquées. Ces résultats montrent que la prise en compte des effets de la variation des températures, dans le cadre de cette étude, n'est pas négligeable.

Conclusions

Les ouvrages et structures du génie civil sont généralement soumis à des processus de détérioration pouvant se produire de manière graduelle ou abrupte et avec une gravité plus ou moins importante. Ce processus mène en général à des modifications de la réponse dynamique de la structure et, bien souvent, à des changements des paramètres modaux tels que les fréquences propres, les coefficients d'amortissement et les déformées propres. L'un des apports majeurs de cette thèse est lié non seulement à l'étude des nouvelles approches pour la détection et la discrimination de comportements structuraux anormaux, mais aussi à l'étude des effets des variations thermiques sur les paramètres modaux des structures. La contribution à ces sujets a été développée selon six axes principaux.

En premier lieu, une étude bibliographique sur les méthodes et indicateurs pour l'identification et la détection d'endommagements structuraux a été menée. Cette étude a permis de montrer que les techniques existantes sont essentiellement basées soit sur la variation des fréquences propres, soit sur celle des modes propres soit encore sur celle d'indicateurs dits « évolués ». Par ailleurs, la plupart des indicateurs existants sont issus de l'utilisation des paramètres modaux ou bien des propriétés physiques structurelles, telles que les matrices de masse et de rigidité. De manière générale, il est souvent considéré que les fréquences propres ne peuvent pas constituer des indicateurs fiables pour la détection d'endommagement, du fait des incertitudes liées à leur identification ou du manque d'outils adéquats pour en extraire des informations pertinentes sur le comportement de l'ouvrage. À l'inverse, l'utilisation des modes propres pour la détection d'endommagements est plus répandue dans la communauté scientifique. En réalité, c'est l'utilisation des indicateurs constitués à partir des déformées modales ou de leurs variations qui a connu un large développement ces dernières années. Toutefois, l'un des principaux inconvénients de ces approches est lié à leur incapacité à extraire des connaissances sur le comportement de la structure à partir de la prise en compte des mesures dynamiques « brutes ». Les mesures *in situ* sont peu (voire jamais) exploitées faute d'outils efficaces ou des difficultés à gérer les quantités importantes de données brutes.

Le premier enjeu majeur de cette thèse est d'apporter des nouveaux éléments de réponse à cette problématique. Pour cela, une approche originale basée sur l'Analyse de Données Symboliques (ADS) a été proposée. Cette méthodologie permet de manipuler des grandes bases de données alliant un stockage compact sans perdre la généralité de l'information originale. Actuellement, l'ADS n'est encore généralement employée que dans le cadre d'applications économiques, comme l'analyse de comportements de consommateurs, des ventes croisées, etc. Dans le cadre de cette thèse, l'ADS a été couplée à des méthodes de partitionnement de données - divisions hiérarchiques, agglomérations hiérarchiques et nuées dynamiques - et à des méthodes basées sur la classification supervisée - arbres de décision Bayésiens, réseaux neuronaux et machines à vecteur support. Il est noté que ces méthodes sont déjà connues dans le domaine de la classification de données. Toutefois, leur couplage avec l'ADS pour distinguer différents états structuraux représente une alternative innovante pour la détection d'endommagements. Les chapitres 2 et 3 de ce mémoire ont porté sur la description de cette approche et, plus spécifiquement, sur leur utilisa-

tion avec des données issues de campagnes de mesures dynamiques. La démarche utilisée pour transformer des signaux et des paramètres modaux dits classiques en représentations symboliques a été présentée. Par la suite, les méthodes citées précédemment ont été décrites et des exemples succincts ont été présentés afin de mieux appréhender la procédure utilisée. L'enjeu majeur de l'approche proposée est de pouvoir discriminer des comportements structuraux et permettre, par conséquent, de distinguer un état considéré « normal » d'un état « anormal ».

Le chapitre 4 est consacré à une étude de sensibilité et de validation de la démarche proposée. Pour cela, un modèle numérique d'une poutre bi-appuyée discretisée par éléments finis a été utilisé. Cette application permet de simuler quatre scénarios d'endommagement, chacun représentant un état structural de la poutre. Par ailleurs, trois niveaux de bruit ont été ajoutés aux mesures dynamiques simulées. L'objectif principal a été de vérifier la robustesse et l'adéquation des méthodes utilisées pour la discrimination des états structuraux, en considérant des mesures bruitées. L'utilisation des méthodologies dites classiques, telles que l'analyse d'intervalles de confiance, les tests d'hypothèses, ou encore le calcul des coefficients MAC, a montré leur insuffisance pour discriminer les scénarios d'endommagements simulés. Les méthodes de clustering ont alors été mises en œuvre en considérant plusieurs simulations avec différents sous-ensembles d'essais. Des analyses ont été menées en utilisant les signaux et les paramètres modaux transformés en intervalles interquartiles et en histogrammes. Les résultats ont montré que, de manière générale, la démarche proposée est capable de discriminer les états structuraux simulés. Toutefois, pour des niveaux importants de bruit, les taux de classification correcte sont relativement plus faibles. Il est remarqué par ailleurs que l'utilisation des fréquences propres produit de meilleurs résultats que l'utilisation des signaux ou des modes propres. Contrairement aux approches classiques, l'utilisation de l'ADS permet de mieux représenter les variations fréquentielles. Les méthodes de clustering permettent alors de constituer un indicateur pertinent pour la détection d'endommagements. En effet, en utilisant une représentation en histogrammes par exemple, il est possible d'analyser la variabilité des fréquences propres identifiées pour un essai donné. En cela, au lieu d'utiliser la variation des valeurs moyennes comme un indicateur d'endommagement (cas classique) et par conséquent de condenser de manière importante les données originales, l'ADS permet d'apporter une description plus riche. De plus, les méthodes des nuées dynamiques et des divisions hiérarchiques sont plus efficaces pour la discrimination des différents scénarios. La deuxième partie de cette étude porte sur l'utilisation d'indices de partitionnement optimal appliqués aux méthodes de clustering. Ces indices sont importants dans le cas où le nombre de groupes de comportement n'est pas connu au préalable. De cette façon, ils permettent, dans un certain sens, de valider les résultats obtenus par les méthodes de partitionnement de données. Trois indices existants dans la littérature ont été considérés. Il est noté que, de façon générale, deux de ces indices (CH et Γ) produisent des résultats similaires (et pertinents) dans les simulations réalisées. Le troisième indice, C^* a montré être relativement plus erratique et moins stable. Il est cependant constaté que cet indice est plus sensible à des modifications structurales, pouvant indiquer des comportements « cachés » qui ne sont pas détectés par les autres indices. Par ailleurs, les méthodes dites supervisées ont également été étudiées. Il a été observé que les réseaux neuronaux et les machines à vecteurs support produisent généralement les résultats les plus pertinents. En effet, si comparées aux méthodes de clustering, les méthodes de classification supervisée permettent d'atteindre des taux de classification correcte plus élevés. Cela est dû au fait que ces dernières sont basées sur une procédure d'apprentissage. L'utilisation de ces méthodes a montré être robuste, surtout dans le cas où les fréquences propres sont utilisées comme données d'entrée.

Le chapitre 5 porte sur une première application expérimentale. Des essais dynamiques acquis à partir de plusieurs campagnes de mesures réalisées sur le pont-rail PK 075+317 ont été utilisées. Ce pont, en béton armé, fait partie de la ligne à grande vitesse qui relie Paris à Lyon. Dans le cadre d'une campagne de surveillance dynamique menée au début de l'année 2003, il a été

constaté que la fréquence de passage des TGV était très proche de la première fréquence propre de vibration de l'ouvrage. Afin de restreindre les effets de résonance, la SNCF a décidé de mettre en œuvre un système d'encastrement mécanique des abouts du tablier sur les culées, afin de rigidifier la structure et, par conséquent, d'augmenter les fréquences propres du pont. Cet ouvrage a été choisi pour atteindre un double objectif. Le premier à consisté à apprécier les améliorations apportées par la procédure de resserrage menée. En d'autres termes, le but a été de discriminer chacun des états structuraux existants : avant, pendant et après le resserrage des bielles. Comme pour l'étude numérique proposée au chapitre 4, une analyse classique a été menée dans un premier temps. Cette étude a montré être inadaptée pour tirer des conclusions sur l'efficacité des travaux. L'approche symbolique couplée avec des méthodes de classification non supervisée et supervisée a alors été proposée. En général, les résultats ont montré que ces approches sont capables de discriminer des modifications structurales avec un taux de réussite adéquat. Par ailleurs, les méthodes supervisées sont apparues relativement plus performantes que les méthodes de clustering. De plus, il a été noté que les méthodes de clustering basées sur les nuées dynamiques et les divisions hiérarchiques ont été plus efficaces pour la discrimination des états structuraux, notamment lorsque des fréquences propres sont utilisées (le même constat avait été effectué dans le cas des simulations numériques). De même, les réseaux neuronaux et les machines à vecteur support ont été plus performants en séparant (dans certains cas, de manière parfaite) les trois états structuraux. Par ailleurs, une étude complémentaire a été menée en ne considérant que les deux états « extrêmes » (i.e., avant et après le resserrage). Il a été noté que les méthodes ont généralement pu mieux discriminer ces deux états avec des taux de classification correcte relativement importants.

Quatre approches pour l'affectation de nouveaux essais ont été proposées : la première, basée sur la méthode des divisions hiérarchiques ; la deuxième, sur les nuées dynamiques ; la troisième, sur le calcul des distances entre essais, et la quatrième, basée sur les méthodes de classification supervisée. Ces approches ont été appliquées dans un contexte des nouvelles campagnes de mesures réalisées par la SNCF, lors des années 2004, 2005 et 2006. Ces séries d'essais ont été menées afin d'apprécier les effets des travaux de resserrage au cours du temps. L'objectif a alors été d'affecter les nouveaux essais à un des clusters déjà existants (i.e., les groupes représentant l'état avant, pendant et après resserrage). La possibilité de les affecter à un nouveau groupe (représentant un état différent) a également été considérée. Une analyse classique simple a d'abord été réalisée, en évaluant les intervalles de confiance pour la moyenne des fréquences propres identifiées. Les résultats ne permettant pas de tirer de conclusions définitives, des nouvelles démarches ont été proposées. Les résultats obtenus ont montré que la plupart des nouveaux essais sont rarement affectés dans le cluster associé à l'état après resserrage. Une conclusion peut alors être tirée (ce qui n'était pas été possible en utilisant l'analyse classique) : l'efficacité des travaux de resserrage n'est plus remarquée au cours des années suivantes les travaux. Ceci est surtout observé pour les campagnes de 2005 et 2006, où pratiquement toutes les approches d'affectation indiquent que le comportement de l'ouvrage est revenu à l'état « pendant » (voire même « avant ») les travaux menés.

Le second enjeu majeur de la thèse a été d'évaluer l'impact des variations environnementales (en particulier celles liées à la température) sur les caractéristiques modales (i.e., les fréquences et modes propres). Pour cela, deux techniques de régression de données ont été introduites. La première, basée sur la régression linéaire multiple et la seconde (non linéaire) basée sur des réseaux de neurones. La procédure a été menée en deux temps avec (i) le filtrage des effets de la variation de températures sur les fréquences et modes propres, à partir de l'utilisation des techniques proposées, (ii) l'utilisation des paramètres modaux « corrigés » des effets thermiques pour appliquer les méthodes de clustering, de classification supervisée et les approches d'affectation de nouveaux essais. L'objectif a été d'observer comment les résultats obtenus précédemment peuvent changer en tenant en compte des effets thermiques. Dans le cas des méthodes de cluste-

ring et de classification supervisée, il a été remarqué que les taux de réussite sont, dans certaines études, plus élevés et dans d'autres, plus faibles. Ceci permet de conclure que, dans cette analyse, la variation des températures n'influence pas significativement les paramètres modaux. Pour cet ouvrage, pendant la période analysée, il n'est donc pas nécessaire de prendre en compte les effets de la variation de température. Il est à noter que pratiquement tous les essais ont été enregistrés sous les mêmes horaires et sous une plage de températures similaires.

Le chapitre 6 considère une seconde application expérimentale, basée sur des campagnes d'essais réalisées sur le pont PI-57. Ce pont, en béton précontraint, permet à l'autoroute A1 (qui relie Paris à Lille) de franchir l'Oise. Pendant l'été de 2009, ce pont a subi des travaux de renforcement qui ont consisté à ajouter des câbles de précontrainte permettant de rigidifier l'ouvrage. Comme il a été proposé dans le chapitre 5, une étude des effets thermiques sur les paramètres modaux a été menée. Cette étude a été divisée en deux étapes. Le couplage des techniques de régression avec l'ADS (plus spécifiquement avec les méthodes d'affectation des essais) a d'abord été considéré. En utilisant des intervalles de confiance pour les fréquences propres, calculés par les modèles de régression, il a été possible de détecter des observations dites « anormales » (celles qui se situent en dehors des intervalles calculés). Toutefois, la distinction de ces essais n'implique pas l'existence d'une anomalie structurale. En réalité, ces observations représentent les effets causés par les travaux de renforcement réalisés. En d'autres termes, toutes les observations qui se situent en dehors de la bande de confiance sont celles pour lesquelles les effets du renforcement sont remarqués. Théoriquement, si le modèle de régression est bien ajusté, toutes les observations prédites doivent être caractérisées comme « anormales ». Il faut cependant considérer la possibilité d'erreurs de détection. Il a été cependant montré que les modèles de régression n'ont pas signalé un nombre important d'essais comme étant des « anomalies » et donc, n'ont pas été capables de mettre en évidence les effets des travaux de renforcement réalisés. Afin de compléter l'analyse, une deuxième étape a été mise en œuvre. L'idée a consisté à appliquer les méthodes d'affectation aux essais considérés comme « normaux » (situés à l'intérieur des bandes de confiance) par les modèles de régression en les affectant soit au groupe d'essais de la phase 1 (avant travaux de renforcement), soit dans un nouveau groupe. Il est alors possible de déterminer si la détection par les modèles de régression est pertinente ou non. En utilisant cette deuxième étape d'affectation, la plupart des essais n'appartenaient généralement pas au cluster d'essais de la phase 1. Ces résultats montrent qu'il ne suffit pas d'utiliser uniquement les modèles de régression pour discriminer les effets d'une modification structurale de ceux liés à la variation de la température. Par ailleurs, une approche basée sur l'ACP a été proposée. L'avantage est qu'elle ne nécessite pas des valeurs mesurées par les sondes de température, une fois que les effets thermiques sont considérés comme étant inhérents aux fréquences propres identifiées. Les résultats obtenus par l'ACP ne donnent généralement pas suffisamment d'éléments de conclusion pour distinguer une modification structurale d'un changement lié aux effets thermiques. Il a été montré qu'un nombre élevé de fausses alarmes a été détecté par l'approche d'affectation des essais.

Par ailleurs, une procédure similaire à celle menée dans le chapitre 5 pour la correction des effets de la température sur les paramètres modaux a été considérée. Le but de cette étude a été d'appliquer l'approche basée sur le couplage de l'ADS aux méthodes de classification de données, pour la discrimination des états structuraux existants : avant et après renforcement. La phase 1 d'instrumentation a eu lieu entre octobre 2008 et avril 2009 et la phase 2 pendant la même période des années suivantes. Dans un premier temps, la totalité des essais enregistrés (2500 environ) a été considérée. Après avoir réalisé quelques analyses classiques préliminaires, les mêmes limitations que celles mentionnées dans le chapitre 5 ont été mises en évidence. Par conséquent, des simulations tenant compte des méthodes de clustering ont été proposées. Il a été observé que les taux de classification correcte ont été relativement faibles, même quand les fréquences propres ont été considérées. Afin d'analyser ces résultats plus en détail, une étude plus

fine a été menée. La démarche mise au point a consisté à classer les essais en considérant à chaque fois un mois particulier en 2008 avec le même mois en 2009 (ou un mois de 2009 avec le même mois en 2010). En d'autres termes, au lieu de mener une étude globale où tous les essais des phases 1 et 2 sont considérés tous ensemble, une comparaison avant et après travaux en considérant les essais mois par mois, a été effectuée. Les résultats obtenus ont montré que cette approche permet une meilleure classification des essais. Par ailleurs, en tenant compte des essais enregistrés au mois d'octobre, il a été remarqué qu'en utilisant les signaux transformés en histogrammes, tous les essais ont été correctement classés. Ce résultat est important, car il montre qu'il est possible d'extraire des informations sur le comportement de l'ouvrage à partir des mesures effectuées directement *in situ*. De façon générale, les méthodes basées sur les divisions hiérarchiques et nuées dynamiques ont été plus performantes que les agglomérations hiérarchiques. En ce qui concerne les méthodes supervisées, il a été observé que l'utilisation de l'analyse mois par mois n'apporte pas une amélioration significative sur les résultats. Par ailleurs, les réseaux de neurones et les machines à vecteurs support ont permis d'atteindre des taux de classification correcte plus importants que les arbres de décision Bayésiens. En considérant les méthodes de clustering, il a été remarqué que les taux de classification correcte sont significativement plus élevés. Toutefois, les résultats obtenus par les méthodes de classification supervisée n'ont pas montré une amélioration évidente. Bien que la différence sur les résultats n'ait pas été nette dans le cas du filtre linéaire (où les pourcentages sont pratiquement du même ordre sans et avec correction des effets thermiques), les modifications obtenues par le filtre non linéaire sont plus évidentes. Ces résultats montrent que la prise en compte des effets de la variation des températures ne sont pas négligeables dans le cadre de cette étude.

L'un des apports majeurs de cette thèse a été de proposer des nouveaux outils permettant la discrimination de comportements structuraux, pouvant être liés à l'existence d'endommagements. En utilisant l'approche basée sur le couplage de l'ADS aux méthodes de classification, il a été possible d'utiliser les mesures dynamiques acquises directement *in situ* comme des indicateurs de l'état de santé de l'ouvrage. De plus, la prise en compte des fréquences propres, bien souvent considérées comme des indicateurs inefficaces, a apporté des résultats pertinents montrant que ces paramètres peuvent être utilisés pour la discrimination de comportements structuraux. En effet, la plupart des analyses classiques trouvées dans la littérature ne font que comparer des valeurs ponctuelles (valeurs moyennes, par exemple) des fréquences propres d'un état non endommagé et d'un autre, endommagé. Toutefois, cette démarche peut éventuellement masquer des variations, parfois petites, mais qui doivent être prises en compte pour la détection d'un endommagement. En utilisant la représentation en données symboliques, notamment celle basée sur des histogrammes, il est possible de mieux caractériser les variations fréquentielles inhérentes à un processus endommageant. En revanche, l'utilisation des déformées modales semble être moins robuste. Même que la plupart des indicateurs d'endommagements recensés dans la littérature soient basée sur la variation des modes propres, les approches proposées dans le cadre de cette thèse n'ont pas indiqué de façon claire une meilleure capacité des indicateurs basés sur les modes propres, à détecter des modifications structurales.

La mise en œuvre des démarches proposées dans cette thèse est parfaitement envisageable dans le cadre d'une surveillance structurelle continue. Les méthodes de clustering basées sur les divisions hiérarchiques et les nuées dynamiques sont recommandés aussi bien que les méthodes de classification supervisée basées sur les réseaux neuronaux et les machines à vecteur support. Vraisemblablement, l'utilisation de ces dernières dépend de la connaissance préalable des conditions structurales afin d'entraîner les algorithmes de classification. Dans la pratique, il n'est généralement pas aisé de connaître *a priori* les différents états d'une structure. Toutefois, si cela est possible, l'utilisation de méthodes supervisées est fortement conseillée pour la détection de modifications structurales. Ces méthodes ont davantage montré être plus adaptées et relativement robustes par rapport à la présence de bruit dans les mesures. Un éventuel inconvénient de

la démarche basée sur l'ADS repose cependant sur le type de transformation symbolique à utiliser. Comme expliqué précédemment, le choix de la représentation (par intervalles, histogrammes, etc.) peut dépendre de l'application étudiée et il revient à l'utilisateur de choisir la représentation symbolique la plus adéquate.

Finalement, la prise en compte des effets thermiques sur les paramètres modaux pour des longues campagnes de mesures doit être considérée. Les techniques de régression linéaire et non linéaire couplées avec les méthodes d'affectation d'essais permettent au gestionnaire de l'ouvrage de suivre, en temps réel, les variations fréquentielles inhérentes, soit aux variations environnementales, soit aux modifications structurales contrôlées, soit encore à des effets endommageants. Un point important à considérer est l'étude des modèles de prédiction et des filtres de correction thermique appliqués aux mesures dynamiques. Bien que la manipulation directe des signaux soit plus délicate (notamment à cause du bruit inhérent aux mesures), il est essentiel d'être capable de définir des seuils d'alarme à partir des mesures dynamiques. Cela permettrait au gestionnaire de suivre en temps réel le comportement dynamique de l'ouvrage et de prendre une décision pertinente si une modification structurale anormale est détectée. Pour cela, des modèles d'autorégression tels que l'ARX, l'ARMA ou encore l'ARMAX peuvent être considérés. Par ailleurs, le développement de démarches complémentaires à celles proposées pour la localisation et la quantification des endommagements est une perspective à apprécier, la thèse s'étant principalement consacrée au problème de leur détection.

Références bibliographiques

Abdel Wahab, M., De Roeck, G. & Peeters, B. (1999), 'Parametrisaion of damage in reinforced concrete structures using model updating', *Journal of Sound and Vibration* **228**(4), 717–730.

Abdelghani, M., Goursat, M., Biolchini, T., Hermans, L. & Van Der Auweraer, H. (1999), Performance of output-only identification algorithms for modal analysis of aircraft structures, *in* 'Proceedings of the 17th International Modal Analysis Conference', Orlando, États-Unis, pp. 19–21.

Aktan, A., Lee, K., Chuntavan, C. & Aksel, T. (1994), Modal testing for structural identification and condition assessment of constructed facilities, *in* 'Proceedings of the IMAC 12', Honolulu, Hawaï, pp. 462–468.

Alampalli, S. (1995), Influence of in-service environment on modal parameters, *in* 'Proceedings of the IMAC 16', Santa Barbara, États-Unis, pp. 111–116.

Allemang, R. & Brown, D. (1982), A correlation coefficient for modal vector analysis, *in* 'Proceedings of the IMAC 1', Orlando, États-Unis, pp. 145–151.

Alvandi, A. (2004), Contribution à l'utilisation pratique de l'évaluation dynamique pour la détection d'endommagements dans les ponts, Thèse de doctorat, École Nationale des Ponts et Chaussées, France.

Alvandi, A. & Crémona, C. (2006), 'Assessment of vibration-based damage identification techniques', *Journal of Sound and Vibration* **292**(2), 179–202.

Alvin, K. (1997), 'Finite element model update via bayesian estimation and minimization of dynamic residuals', *AIAA Journal* **35**(5), 879–886.

Asmussen, J. (1997), Modal analysis based on the random decrement technique : Application to civil engineering structures, Thèse de doctorat, University of Aalborg, Danemark.

Asmussen, J. & Brincker, R. (1996), Estimation of correlation functions by random decrement, *in* 'Proceedings of the 21st International Conference on Noise and Vibration Engineering', Louvain, Belgique, pp. 1215–1224.

Azeez, M. & Vakakis, A. (2001), 'Proper orthogonal decomposition of a class of vibroimpact oscillations', *Journal of Sound and Vibration* **240**(5), 859–889.

Baker, F. & Hubert, L. (1975), 'Measuring the power of hierarchical cluster analysis', *Journal of the American Statistical Association* **1**(70), 31–38.

Barbosa, F. & Crémona, C. (2001), Identification modale de structures sous sollicitation ambiante, Technical report, Laboratoire Central des Ponts et Chaussées, France.

Bardos, M. & Rasson, J. (2001), *Analyse discriminante , Application au risque et au scoring financier*, Dunod, pp. 145–181.

Basseville, M., Benveniste, A., Goursat, M. & Mevel, L. (2006), 'Subspace-based algorithms for structural identification, damage detection, and sensor data fusion', *EURASIP Journal on Advances in Signal Processing* **2007**(1), 1–13.

Bernard, P. (1988), 'Identification de grandes structures : une remarque sur la méthode du décrément aléatoire', *Journal de Mécanique Théorique et Appliquée* **7**(3), 269–280.

Bernard, P. (1990), 'Slepian model : a new approach', *Structural Safety* **8**, 301–310.

Bertrand, P. (1986), Etude de la Représentation Pyramidale., Thèse de doctorat, Université Paris-Dauphine, France.

Billard, L. & Diday, E. (2006), *Symbolic Data Analysis*, John Wiley and Sons.

Bishop, C. (1995), *Neural Networks for Pattern Recognition*, 1 edn, Oxford University Press.

Bock, H. & Diday, E. (2000), *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin.

Bodeux, J.-B. & Golinval, J.-C. (2003), 'Modal identification and damage detection using the data-driven stochastic subspace and armav methods', *Mechanical systems and signal processing* **17**(1), 83–89.

Brito, P. (1994), *New Approaches in Classification and Data Analysis*, Springer-Verlag, chapter Use of Pyramids in Symbolic Data Analysis, pp. 378–386.

Brito, P. & Diday, E. (1990), *Knowledge, Data and Computer-Assisted Decisions*, Springer-Verlag, chapter Use of Pyramids in Symbolic Data Analysis, pp. 3–16.

Burbridge, R. & Buxton, B. (2002), 'An introduction to support vector machines for data mining'.

Calinski, T. & Harabasz, J. (1974), 'A dendrite method for cluster analysis', *Communications in Statistics* **1**(3), 1–27.

Cawley, P. & Adams, R. (1979), 'The location of defects in structures from measurements of natural frequencies', *Journal of Strain Analysis* **14**(2), 49–57.

Celeux, G., Diday, E., Govaert, G., Lechevalier, Y. & Ralambondrainy, H. (1989), 'Classification automatique des données', *Environnement Statistique et Informatique* pp. 25–37.

Chance, J., Tomlinson, J. & Worden, K. (1994), A simplified approach to the numerical and experimental modeling of the dynamics of a cracked beam, *in* 'Proceedings of the IMAC 12', Honolulu, Hawaï, pp. 778–785.

Chang, C. (1975), Study of dynamic characteristics of aeroelastic systems utilizing randomdec signatures, Technical Report CR-13563, NASA, Langley Research Center, Hampton, États-Unis.

Chang, C. & Lin, C. (2001), *LIBSVM : a library for support vector machines.* `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Chavent, M. (1997), Analyse de Données Symboliques. Une Méthode Divisive de Classification, Thèse de doctorat, Université Paris-Dauphine, France.

Chun, X., Weilian, Q. & Dongmei, T. (2005), An application of data fusion technology in structural health monitoring and damage identification, *in* 'Proceedings of Smart Structures and Materials 2005 : Smart Sensor Technology and Measurement Systems', San Diego, États-Unis, pp. 451–461.

Cole, H. (1968), 'On-the-line analysis of random vibrations', *AIAA/ASME Ninth Structures, Structural Dynamics and Materials Conference* .(1), 68–288.

Cole, H. (1973), On-line failure detection and damping measurements of aerospace structures by random decrement signature, Technical Report CR-2205, NASA, Langley Research Center, Hampton, États-Unis.

Cristianini, N. & Shawe-Taylor, J. (2000), *An introduction to Support Vector Machines*, 2 edn, Cambridge University Press.

Crémona, C. (1990), Recursive time domain modal identification techniques, Thèse de doctorat, University of Wales College of Cardiff, Royaume-Uni.

Crémona, C. (2004), *Évaluation dynamique des ponts*, Presses LCPC, Paris, France.

Crémona, C. (2007), *Suivi du comportement mécanique d'une poutre de VIPP sous chargement - Le cas du VIPP de Merlebach*, Presses LCPC, Paris, France.

Crémona, C., Barbosa, F. & Alvandi, A. (2003), 'Identification modale sous excitation ambiante : application à la surveillance des ponts', *Revue Mécanique et Industries* **4**, 259–271.

Crémona, C. & Brandon, J. (1992), 'Modal order testing : an overview', *Journal of Mechanical Systems and Signal Processing* **6**(3), 229–336.

Crémona, C., Hallak, P., Barbosa, F., Lavigne, J. & Moretti, O. (2004), Field dynamic performance of an exceptional timber bridge, *in* 'Proceedings of IABMAS'04 Conference', Kyoto, Japon, pp. 19–21.

Cury, A., Borges, C. & Barbosa, F. (2010), 'A two-step technique for damage assessment using numerical and experimental vibration data', *in press* .

Cury, A. & Crémona, C. (2010*a*), Long term dynamic monitoring of a psc box girder bridge, *in* 'Proceedings of the IABSE (International Association for Bridge and Structural Engineering) Conference', Venise, Italie.

Cury, A. & Crémona, C. (2010*b*), Novelty detection based on symbolic data analysis applied to structural health monitoring, *in* 'IABMAS 2010', Philadelphie, États-Unis, pp. 267–276.

Cury, A. & Crémona, C. (2010*c*), 'Pattern recognition of structural behaviors based on learning algorithms and symbolic data concepts', *Journal of Structural Control and Health Monitoring* . accepté.

Cury, A. & Crémona, C. (2010*d*), Pi-57 bridge : monitoring the effects of temperature variation on modal parameters, *in* 'International Conference on Noise and Vibration Engineering', Louvain, Belgique.

Cury, A., Crémona, C. & Diday, E. (2009*a*), A methodology based on symbolic data analysis for structural damage assessment, *in* 'IOMAC 2009 - International Operational Modal Analysis Conference', Portonovo, Italie, pp. 267–276.

Cury, A., Crémona, C. & Diday, E. (2009*b*), Symbolic data analysis applied to structural damage evaluation, *in* 'EVACES 2009 - Experimental Vibration Analysis in Civil Engineering Structures', Wroclaw, Pologne, pp. 335–445.

Cury, A., Crémona, C. & Diday, E. (2010), 'Application of symbolic data analysis for structural modification assessment', *Engineering Structures* **32**(3), 762–775.

Da Silva, S., Dias Junior, M., Lopes Junior, V. & Brennan, M. J. (2008), 'Structural damage detection by fuzzy clustering', *Mechanical systems and signal processing* **22**(7), 1636–1649.

De Boe, P. & Golinval, J. (2003), 'Principal component analysis of piezo-sensor array for damage localization', *Structural Health Monitoring* **2**(2), 137–144.

Diday, E. (1971), 'La méthode des nuées dynamiques', *Révue de Statistique Appliquée* **2**(19), 19–34.

Diday, E. & Noirhomme-Fraiture, M. (2007), *Symbolic Data Analysis and the SODAS Software*, John Wiley and Sons.

Doebling, S., Farrar, C., Prime, M. & Shevitz, D. (1996), Damage identification and health monitoring on structural and mechanical systems form changes in their vibration characteristics : a literature review, Technical report, Los Alamos National Laboratory, États-Unis.

Duggan, D., Wallance, E. & Caldwell, S. (1980), Measured and predicted vibrational behavior of gulf of mexico platforms, *in* 'Proceedings of the the 12th Annual Offshore Techical Conference', Houston, États-Unis, pp. 92–100.

EMPA (1999), Results of tasks a1 and a2 : Long-term monitoring and bridge tests, Technical Report 168, Laboratoire Fédéral d'Essai des Matériaux et de Recherche, Suisse.

EN1990 (1990), *EN 1990 Eurocode : Bases de calcul des structure.*

Esposito, F., Malerba, D. & Tamma, V. (2000), *Analysis of Symbolic Data : Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, chapter Dissimilarity Measures for Symbolic Data, pp. 165–185.

Ewins, D. (2000), *Modal Testing : Theory and Practice*, 2 edn, John Wiley and Sons.

Fang, X., Luo, H. & Tang, J. (2005), 'Structural damage detection using neural network with learning rate improvement', *Computers and Structures* **83**, 2150–2161.

Farrar, C., Doebling, S., Cornwell, P. & Straser, E. (1995), Variability of modal parameters measured on the alamosa canyon bridge, *in* 'Proceedings of the IMAC 15', États-Unis, pp. 257–263.

Farrar, C. & Jauregui, D. (1996), Damage detection algorithms applied to experimental and numerical modal data from the i-40 bridge, Technical report, Los Alamos National Laboratory report, États-Unis.

Fisher, R. A. (1954), *Statistical Methods for Research Workers*, 12 edn, Oliver and Boyd.

Fox, C. (1992), The location of defects in structures : A comparison of the use of natural frequency and mode shape data, *in* 'Proceedings of the IMAC 10', San Diego, États-Unis, pp. 522–528.

Friswell, M. & Inman, D. (2000), Sensor validation for smart structures, *in* 'Proceedings of the IMAC 18', San Antonio, États-Unis, pp. 483–489.

Friswell, M., Penny, G. & Wilson, D. (1994), 'Using vibration data and statistical measures to locate damage in structures', *The International Journal of Analytical and Experimental Modal Analysis* **9**(4), 239–254.

Fujino, Y. & Yoshida, Y. (2002), 'Wind induced vibration and control of trans-tokyo bay crossing bridge', *ASCE Journal of Structural Engineering* **128**(8), 1012–1025.

Gudmundson, P. (1982), 'Eigenfrequency changes of structures due to cracks, notches, or other geometrical changes', *Journal of the Mechanics and Physics of Solids* **30**(5), 339–353.

Guo, H. (2006), 'Structural damage detection using information fusion technique', *Mechanical Systems and Signal Processing* **20**(1), 1173–1188.

Guo, T., Li, A. & Wang, H. (2007), 'Influence of ambient temperature on the fatigue damage of welded bridge decks', *International Journal of Fatigue* **30**(6), 1092–1102.

Hagiwara, H. & Mita, A. (2003), 'Quantitative damage diagnosis of shear structures using support vector machine', *KSCE Journal of Civil Engineering* **7**(6), 683–689.

Han, S. & Feeny, B. (2003), 'Enhanced proper orthogonal decomposition for the modal analysis of homogeneous structures', *Journal of Vibration and Control* **8**, 19–40.

Haritos, N. & Owen, J. (2004), 'The use of vibration data for damage detection in bridges : a comparison of system identification and pattern recognition approaches', *Structural Health Monitoring* **3**(2), 141–163.

Hastie, T., Tibshirani, R. & Firedman, J. (2009), *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*, 2 edn, Springer Series in Statistics.

He, H. & Yan, W. (2005), 'Structural damage detection with wavelet support vector machine : introduction and applications', *Structural Control and Health Monitoring* **14**(1), 162–176.

Hermans, L. & Van der Auwaraer, H. (1999), 'Modal testing and analysis of structures under operational conditions : industrial applications', *Mechanical Systems and Signal Processing* **13**(2), 193–216.

Hopfield, J. (1982), 'Neural networks and physical systems with emergent collective computational abilities', *Proceedings of the the National Academy of Sciences* **79**(1), 2554–2558.

Hsu, C. & Lin, C. (2002), 'A comparison of methods for multi-class support vector machines', *IEEE Transactions on Neural Networks* **13**, 415–425.

Hubert, L. & Levine, J. (1976), 'A general statistical framework for assessing categorical clustering in free recall', *Psychological Bulletin* **3**(83), 1073–1080.

Ibrahim, S. (1977), 'Random decrement technique for modal identification of structures', *Journal of Spacecraft and Rockets* **14**(11), 696–700.

Ibrahim, S. (1987), 'An upper hessenberg sparse matrix algorithm for modal identification on minicomputer', *Journal of sound and vibration* **113**(1), 47–57.

Ibrahim, S. & Milkucik, E. (1977), 'A method for the direct identification of vibration parameters from the free response', *Shock and Vibration Bulletin* **1**(4), 183–198.

Iwasaki, A., Todoroki, A., Shimamura, Y. & Kobayashi, H. (2004), 'An unsupervised statistical damage detection method for structural health monitoring', *Smart Material Structures* **13**, 80–85.

Juang, J. & Pappa, R. (1984), Galileo spacecraft modal identification using an eigensystem realization algorithm, Technical Report AIAA-1984-1070, Structures and Dynamics Division, NASA, Langley Research Center, Hampton, États-Unis.

Juang, J. & Pappa, R. (1985), 'An eigensystem realization algorithm for modal parameter identification and model reduction', *Journal of Guidance, Control and Dynamics* **8**(5), 620–627.

Juang, J., Pappa, R. & Wright, J. (1988), 'An eigensystem realization algorithm using data correlation for modal parameter identification', *Journal of Control Theory and Advanced Technology* **4**(1), 5–14.

Kato, M. & Shimada, S. (1986), 'Vibration of pc bridge during failure process', *ASCE Journal of Structural Engineering* **112**, 1692–1703.

Kaundal, R. (2005), *A SVM-based server for rice blast prediction dedicated to the farming community.* `http://www.imtech.res.in/raghava/rbpred/algorithm.html`.

Kawiecki, G. (2001), 'Modal damping measurement for damage detection', *Smart Materials and Structures* **10**(3), 466–480.

Kerschen, G., Feeny, B. & Golinval, J.-C. (2003), 'On the exploitation of chaos to build reduced-order models', *Computer Methods in Applied Mechanics and Engineering* **192**(13-14), 1785Ű1795.

Kim, J. & Stubbs, N. (1993), Assessment of the relative impact of model uncertainties on the accuracy of global nondestructive damage detection in structures, Technical report, New Mexico University, États-Unis.

Kim, J., Yun, C. & Yi, J. (2003), 'Temperature effects on frequency-based damage detection in plate-girder bridges', *Korean Society of Civil Engineers* **7**(6), 725–733.

Kim, S. & Lee, J. (2000), 'Use of modal testing to identify damage on steel members', *KSCE Journal of Civil Engineering* **4**(2), 75–82.

Klein, L. (1999), *Sensor and Data Fusion : Concepts and Application*, 1 edn, SPIE.

Krzanowski, W. (2000), *Principles of Multivariate Analysis - A User's Perspective*, Oxford University, Oxford.

Lechevallier, Y. (1974), Optimisation de quelques criteres en classification automatique et application à l'étude des modifications des protéines sériques en pathologie clinique, Thèse de doctorat, Université Paris-VI Pierre et Marie Curie, France.

Leconte, R., Goepfer, F. & Crémona, C. (2007), Dynamic and static behaviour of a beam of the merlebach bridge, *in* 'Proceedings of the EVACES 2005', Bordeaux, France, pp. 129–138.

Lenaerts, V., Kerschen, G. & Golinval, J. (2003), 'Identification of a continuous structures with a geometrical non-linearity, part ii : proper orthogonal decomposition', *Journal of Vibration and Control* **262**(4), 907–919.

Lieven, N. & Ewins, D. (1988), Spatial correlation of mode shapes, *in* 'Proceedings of the the IMAC 6', Kissimmee, États-Unis, pp. 690–695.

Lim, T. & Kashangaki, T. (1994), 'Structural damage detection of space truss structures using best achievable eigenvectors', *AIAA Journal* **32**(5), 1049–1057.

MacQueen, J. B. (1967), Some methods for classification and analysis of multivariate observations, *in* '5th Berkeley Symposium on Mathematical Statistics and Probability', California, États-Unis, pp. 281–297.

Malerba, D., Esposito, F., Gioviale, V. & Tamma, V. (2001), Comparing dissimilarity measures for symbolic data analysis, *in* 'ETK-NTTS', Hersonissos, Crète, pp. 473–481.

Massey, F. (1951), 'The kolmogorov-smirnov test for goodness of fit', *Journal of the American Statistical Association* **46**(253), 68–78.

McCulloch, W. & Pitts, W. (1943), 'A logical calculus of the ideas immanent in nervous activity', *Bulletin of Mathematical Biophysics* **5**(1), 115–133.

Messina, A., Williams, E. & Contursi, T. (1998), 'Structural damage detection by a sensitivity and statistical-based method', *Journal of Sound and Vibration* **216**(5), 791–808.

Mevel, L., Hermans, L. & Van der Auweraer, H. (1999), 'Application of a subspace-based fault detection method to industrial structures', *Mechanical systems and signal processing* **13**(6), 823–838.

Milligan, G. & Cooper, M. (1985), 'An examination of procedures for determining the number of clusters in a data set', *Psychometrika* **50**(2), 159–179.

Minor, C., Johnson, K., Rose-Pehrsson, S., Owrutsky, J., Wales, S., Steinhust, D. & Gottuk, D. (2007), Data fusion with a multisensor system for damage control and situational awareness, *in* 'Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance', Londres, Royaume-Uni, pp. 1055–1068.

Mita, A. & Taniguchi, R. (2004), 'Active damage detection method using support vector machine and amplitude modulation', *Smart Structures and Materials* **53**(1), 21–29.

Ndambi, J., Vantomme, J. & De Visscher, J. (2000), *Modal damping as a damage detection parameter in reinforced concrete structures*, Civil-Comp press, Edinburgh, Royaume-Uni, pp. 1–7.

Pandey, A. & Biswas, M. (1994), 'Damage detection in structures using changes in flexibility', *Journal of Sound and Vibration* **169**(1), 3–17.

Pandey, A., Biswas, M. & Samman, M. (1991), 'Damage detection from changes in curvature mode shapes', *Journal of Sound and Vibration* **145**(2), 321–332.

Peeters, B. & De Roeck, G. (2001), 'One-year monitoring of the z24-bridge : environmental effects versus damage events', *Earthquake Engineering and Structural Dynamics* **30**(2), 149–171.

Peeters, B., Maeck, J. & De Roeck, G. (2001), 'Vibration-based damage detection in civil engineering : excitation sources and temperature effects', *Smart Materials And Structures* **10**, 518–527.

Principe, J., Euliano, N. & Lefebvre, C. (2000), *Neural and Adaptive Systems*, 1 edn, John Wiley and Sons.

Reda Taha, M. & Lucero, J. (2005), 'Damage identification for structural health monitoring using fuzzy pattern recognition', *Engineering Structures* **27**(1), 1774–1783.

Roberts, G. (1995), 'Analysis and testing of bridges', One-Day Seminar, ISTRUCE/SECED, Londres, Royaume-Uni.

Roberts, G. & Pearson, A. (1998), Health monitoring of structures-towards a stethoscope for bridges, *in* 'Proceedings of the ISMA 23, International Conference On Noise and Vibration Engineering', Louvain, Belgique.

Rohrmann, R., Baessler, M., Said, S., Schmid, W. & Ruecker, W. (2000), Structural causes of temperature affected modal data of civil structures obtained by long time monitoring, *in* 'Proceedings of 18th International Modal Analytical Conference', San Antonio, Texas, pp. 1–7.

Rosenblatt, F. (1958), 'The perceptron : probabilistic model for information storage and organization in the brain', *Psychological Review* **65**(2), 386–408.

Rumelhart, D., Hinton, G. & Williams, R. (1986), 'Learning representations by back-propagating errors', *Nature* **323**(1), 533–536.

Rytter, A. (1993), Vibration based inspection of civil engineering structures, Thèse de doctorat, University of Aalborg, Danemark.

Salane, H. & Baldwin, J. (1990), 'Identification of modal properties of bridges', *ASCE Journal of Structural Engineering* **116**(7), 2008–2021.

Salawu, O. (1997), 'Detection of structural damage through changes in frequency : a review', *Engineering Structures* **19**(9), 718–723.

Seber, G. & Wild, C. (2003), *Nonlinear Regression*, John Wiley and Sons-Interscience, New Jersey.

Silva, A. (2005), 'Dissimilarity functions analysis based on dynamic clustering for symbolic data'.

Silverman, B. (1986), 'Using kernel density estimates to invetigate multimodality', *Journal of the Royal Statistical Society* **43**(2), 97–99.

Smith, S. & Beattie, C. (1991), 'Secant-method adjustement for structural models', *AIAA Journal* **29**(1), 119–126.

Sohn, H., Dzwonczyk, M., Straser, E., Kiremidjian, A., Law, K. & Meng, T. (1999), 'An experimental study of temperature effects on modal parameters of the alamosa canyon bridge', *Earthquake Engineering and Structural Dynamics* **28**(1), 879–897.

Sohn, H., Worden, K. & Farrar, C. (2002), Consideration of environmental and operational variability for damage diagnosis, *in* 'Proceedings of Smart Structures and Materials (SPIE)', San Jose, California, pp. 100–111.

Srinivasan, M. & Kot, C. (1992), Effects of damage on the modal paramaters of a cylindrical shell, *in* 'Proceedings of the IMAC 10', San Diego, États-Unis, pp. 529–535.

Su, Z., Wang, X., Cheng, L., Yu, L. & Chen, Z. (2009), 'On selection of data fusion schemes for structural damage evaluation', *Structural Health Monitoring* **9**(3), 223–241.

SysIdent-LCPC (2007), *Dynamic assessment of structures, User's guide.* `http://or.lcpc.fr/sys-ident`.

Teughels, A., Maeck, J. & De Roeck, G. (2002), 'Damage assessment by fe model updating using damage functions', *Computers and Structures* **80**(25), 1869–1879.

Trendafilova, I. & Heylen, W. (2003), 'Categorisation and pattern recognition methods for damage localisation from vibration measurements', *Mechanical Systems and Signal Processing* **17**(4), 825–836.

Vapnik, V. (1995), *The Nature of Statistica Learning Theory*, Springer-Verlag.

Wahebm, M. & De Roeck, G. (1997), 'Effect of temperature on dynamic system parameters of a highway bridge', *Structural Engineering International* **97**(4), 266–270.

Wen, C., Hung, S., Huang, C. & Jan, J. (2005), 'Unsupervised fuzzy neural networks for damage detection of structures', *Structural Control and Health Monitoring* **14**(1), 144–161.

West, W. (1984), Illustration of the use of modal assurance criterion to detect structural changes in an orbiter test specimen, *in* 'Proceedings of the the Air Force Conference on Aircraft Structural Integrity', États-Unis, pp. 1–6.

Williams, E., Contursi, T. & Messina, A. (1996), Damage detection and localisation using natural frequency sensivity, *in* 'Identification in Engineering Systems', Swansea, Royaume-Uni, pp. 368–376.

Woods, M. (1992), Damage Analysis of Bridge Structures Using Vibrational Techniques, Thèse de doctorat, University of Aston in Birmingham, England.

Worden, K., Manson, G. & Fieller, N. (2000), 'Damage detection using outlier analysis', *Journal of Sound and Vibration* **3**(229), 647–667.

Yan, A.-M., Kerschen, G., DeBoe, P. & Golinval, J. (2005), 'Structural damage diagnosis under varying environmental conditions - parti : A linear analysis', *Mechanical Systems and Signal Processing* **19**, 847–864.

Yan, Y., Cheng, L., Wu, Z. & Yam, L. (2007), 'Development in vibration-based structural damage detection technique', *Mechanical Systems and Signal Processing* **21**, 2198–2211.

Yeung, W. & Smith, J. (2005), 'Damage detection in bridges using neural networks for pattern recognition of vibration signatures', *Engineering Structures* **27**, 685–698.

Zhang, Q. (2007), 'Statistical damage identification for bridges using ambient vibration data', *Computers and Structures* **85**, 476–485.

Zhang, Z. & Atkan, A. (1995), The damage indices for constructed facilities, *in* 'Proceedings of the IMAC 13', Nashville, États-Unis, pp. 1520–1529.

Zienkiewicz, O. & Taylor, R. (2000), *The Finite Element Method*, 3 edn, Butterworth-Hienemann.

Annexe A

Résultats complémentaires - Application Numérique

Dans cette annexe, les résultats complémentaires des analyses menées sur le modèle de poutre bi-appuyée sont présentés. Ces résultats sont obtenus à partir de la transformation des signaux et des paramètres modaux en intervalles interquartiles. Chaque paragraphe correspond à une analyse exploitée dans le chapitre 4.

A.1 Analyse classique

Dans ce paragraphe, les résultats complémentaires concernant l'analyse classique de données sont montrés. Le Tableau A.1 présente les valeurs moyennes ($\overline{f}$) et les écarts-types (σ_f) des réalisations pour chacune des fréquences propres identifiées, par scénario d'endommagement, pour le niveau de bruit n0 (sans bruit). Par ailleurs, les différences en pourcentage entre les fréquences de la structure non endommagée (D0) et endommagée (D3) sont calculées.

Sans Bruit - n0										
Scénario	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4		Fréquence 5	
	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
D0	0,58	0,0103	2,31	0,0062	5,21	0,0087	9,26	0,0112	14,46	0,0068
D1	0,57	0,0800	2,31	0,0063	5,18	0,0099	9,25	0,0060	14,39	0,0067
D2	0,57	0,0091	2,31	0,0060	5,17	0,0112	9,24	0,0049	14,36	0,0073
D3	0,57	0,0076	2,30	0,0071	5,13	0,0083	9,24	0,0065	14,25	0,0094
Diff.(%)	**1,85**		**0,61**		**1,52**		**0,13**		**1,51**	

Tableau A.1 – Comparaison des fréquences propres identifiées pour chaque scénario d'endommagement, sans bruit (valeurs moyennes et écarts-types en [Hz]).

Le Tableau A.2 regroupe les valeurs moyennes et écarts-types des fréquences propres identifiées à partir des signaux bruités (niveau n1).

Bruit - n1										
Scénario	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4		Fréquence 5	
	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
D0	0,58	0,0056	2,31	0,0070	5,20	0,0095	9,25	0,0104	14,46	0,0137
D1	0,58	0,0060	2,31	0,0073	5,18	0,0075	9,25	0,0089	14,39	0,0120
D2	0,57	0,0053	2,31	0,0077	5,17	0,0090	9,24	0,0074	14,37	0,0125
D3	0,56	0,0058	2,30	0,0082	5,13	0,0078	9,24	0,0084	14,25	0,0152
Diff.(%)	**1,72**		**0,61**		**1,44**		**0,09**		**1,51**	

TABLEAU A.2 – Comparaison des fréquences propres identifiées pour chaque scénario d'endommagement, bruit niveau n1 (valeurs moyennes et écarts-types en [Hz]).

Les valeurs moyennes et écarts-types des fréquences identifiées à partir des signaux bruités (niveau n2) sont regroupés dans le Tableau A.3.

Bruit - n2										
Scénario	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4		Fréquence 5	
	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
D0	0,58	0,0046	2,31	0,0058	5,21	0,0065	9,25	0,0075	14,46	0,0113
D1	0,57	0,0052	2,31	0,0059	5,18	0,0067	9,25	0,0093	14,38	0,0088
D2	0,58	0,0049	2,31	0,0057	5,17	0,0073	9,25	0,0084	14,37	0,0098
D3	0,57	0,0049	2,30	0,0055	5,12	0,0057	9,24	0,0073	14,25	0,0101
Diff.(%)	**1,85**		**0,61**		**1,55**		**0,09**		**1,51**	

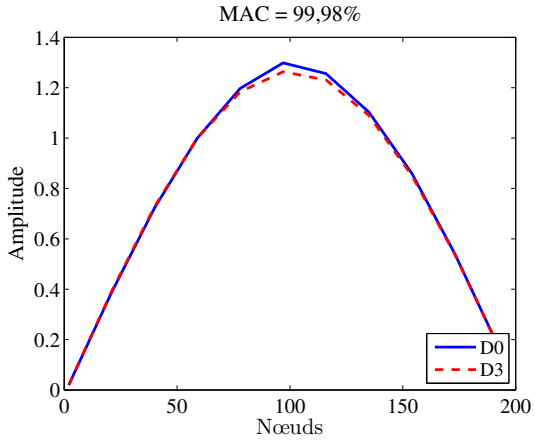
TABLEAU A.3 – Comparaison des fréquences propres identifiées pour chaque scénario d'endommagement, bruit niveau n2 (valeurs moyennes et écarts-types en [Hz]).

Par ailleurs, le Tableau A.4 regroupe les valeurs moyennes et écarts-types des fréquences identifiées à partir des signaux bruités (niveau n3).

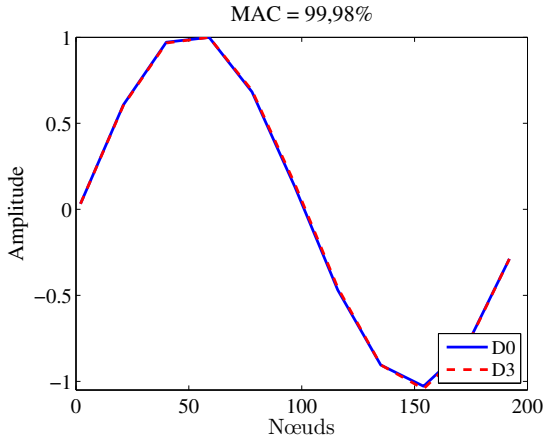
Bruit - n3										
Scénario	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4		Fréquence 5	
	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
D0	0,57	0,0039	2,32	0,0046	5,20	0,0060	9,25	0,0090	14,46	0,0112
D1	0,58	0,0039	2,31	0,0051	5,18	0,0067	9,25	0,0070	14,39	0,0091
D2	0,57	0,0036	2,31	0,0042	5,18	0,0066	9,24	0,0059	14,37	0,0113
D3	0,56	0,0034	2,30	0,0039	5,12	0,0057	9,24	0,0065	14,24	0,0075
Diff.(%)	**1,75**		**0,86**		**1,54**		**0,09**		**1,52**	

TABLEAU A.4 – Comparaison des fréquences propres identifiées pour chaque scénario d'endommagement, bruit niveau n3 (valeurs moyennes et écarts-types en [Hz]).

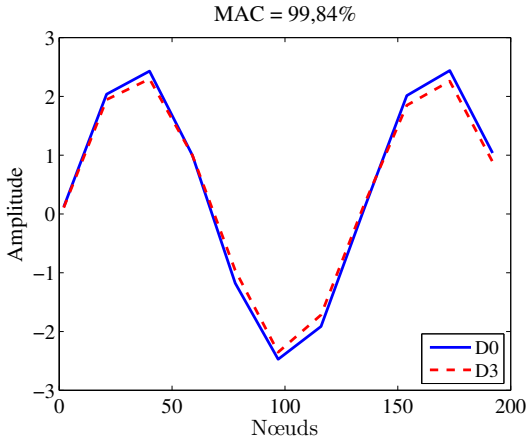
La Figure 4.5 montre une comparaison entre les cinq premiers modes propres identifiés, en ne considérant que les cas D0 et D3 pour le niveau de bruit n3.



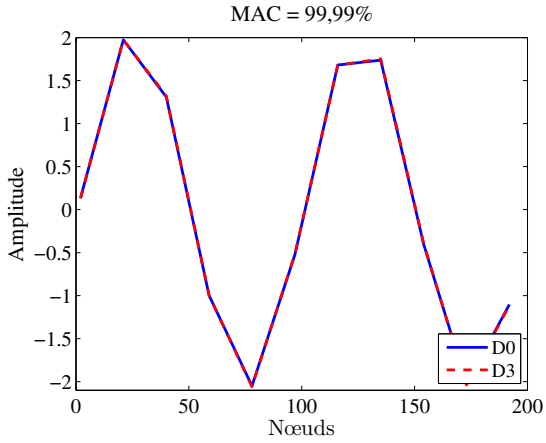
(a) Premier mode propre



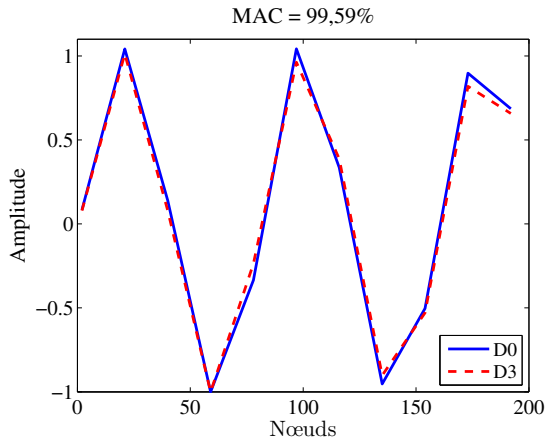
(b) Deuxième mode propre



(c) Troisième mode propre



(d) Quatrième mode propre



(e) Cinquième mode propre

FIGURE A.1 – Comparaison des modes propres non endommagés et endommagés, avec 5% de bruit (n3).

A.2 ADS couplée aux méthodes de clustering

A.2.1 Analyse de la sensibilité au bruit

A.2.1.1 Signaux

Dans ce paragraphe, les résultats complémentaires, concernant l'ADS couplée aux méthodes de clustering, sont montrés. Les premiers résultats correspondent à l'application de la méthode des divisions hiérarchiques aux signaux, transformés en intervalles interquartiles. La Figure A.2 illustre la classification obtenue. De manière générale, les taux de classification correcte sont assez satisfaisants. Dans cette analyse, 8 essais sur 10 de type D0 sont correctement classés dans un même groupe. Pour le scénario D1, le taux est de 70%. Pour les deux derniers scénarios, D2 et D3, les classifications sont légèrement meilleures : 80% et 100%, respectivement. Il est important d'observer que les essais de type D0 et D3 (qui représentent des conditions structurales les plus éloignées) ne sont jamais mélangés. Cependant, les essais qui correspondent aux états D1 et D2 (états relativement « proches ») tendent à être classés incorrectement. Pour ce partitionnement, les capteurs 5 et 9 ont été sélectionnés comme les variables les plus discriminantes.

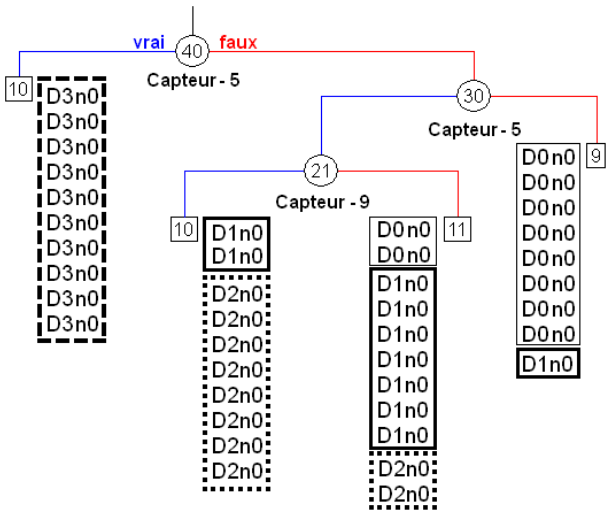


FIGURE A.2 – Classification des signaux sans bruit (n0) utilisant les divisions hiérarchiques (intervalles).

La Figure A.3 montre les résultats en considérant le niveau de bruit n1. Dans ce cas, 70% des essais D0 et D1 sont classés correctement, alors que seulement 60% des essais le sont pour le scénario D2. Pour l'état D3, ce taux s'élève à 80%.

La Figure A.4 montre l'arbre construit à partir des intervalles interquartiles. Il est noté que la classification n'est pas pertinente pour les états D1 et D2. En effet, pour ces deux scénarios, les taux sont de 60% et 70% seulement. Pour les états D0 et D3, ces taux sont légèrement plus élevés : 80% et 70%, respectivement. Par ailleurs, l'un des essais D2 est classé dans le groupe D0. Cela montre que la présence de bruit peut mener, dans certains cas, à des classifications incorrectes.

La Figure A.5 montre la classification obtenue en utilisant des intervalles interquartiles, quand 5% (n3) de bruit est ajouté aux signaux. Il est intéressant de noter que le bruit n'a pas trop influencé la classification des essais des états D0 et D3. En effet, les taux de classification correcte correspondant à ces conditions structurales n'ont pratiquement pas varié avec l'ajout de bruit. Cela montre que cette méthode de clustering est capable de détecter des niveaux d'en-

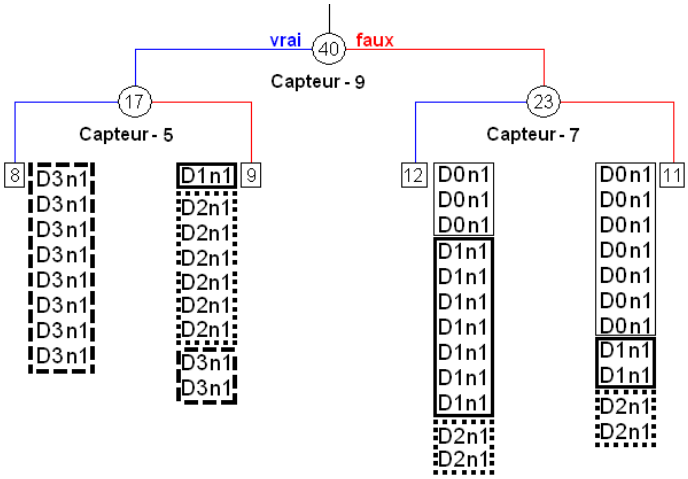


FIGURE A.3 – Classification des signaux avec bruit (n1) utilisant les divisions hiérarchiques (intervalles).

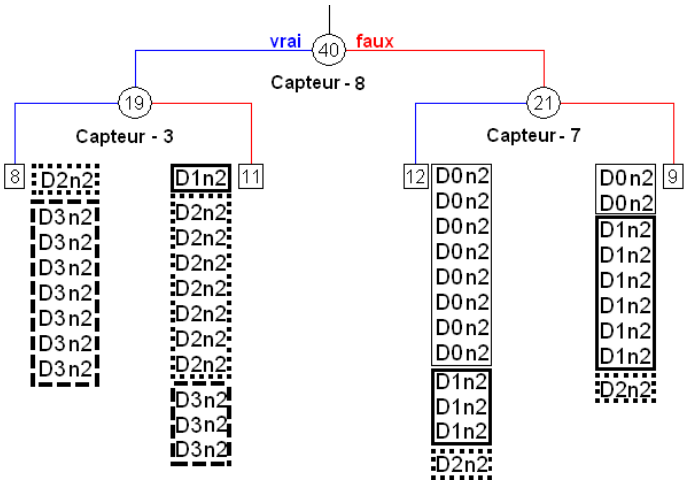


FIGURE A.4 – Classification des signaux avec bruit (n2) utilisant les divisions hiérarchiques (intervalles).

dommagement relativement élevés en utilisant directement les signaux, avec une probabilité de classification correcte satisfaisante : 80% pour les états D0 et D3, 50% pour D1 et 60% pour D2.

La deuxième méthode de clustering utilisée dans cette application est celle basée sur les agglomérations hiérarchiques. Dans le premier cas, sans bruit (n0) et en considérant les signaux transformés en intervalles, les taux de classification correcte sont satisfaisants pour les scénarios D0 et D3 (90%), mais relativement faibles pour les états D1 et D2 : 50% et 60%, respectivement (Figure A.6). Les essais sont donc plus mélangés que dans le cas où le partitionnement est effectué avec la méthode des divisions hiérarchiques.

La Figure A.7 illustre l'agglomération obtenue. Seuls les scénarios D0 et D3 présentent une diminution des taux de classification correcte. Inversement, les taux ont légèrement augmenté pour les scénarios intermédiaires D1 et D2. Il est possible que la présence de bruit ait altéré les valeurs interquartiles de telle sorte que les essais D1 et D2 apparaissent plus différents qu'ils ne le sont réellement. Par conséquent, l'augmentation des distances entre eux permet de les classer dans des clusters différents. Il est important d'apprécier ces premiers résultats avec précaution. En effet, il faut tenir compte de l'ensemble des résultats obtenus pour chaque méthode et observer

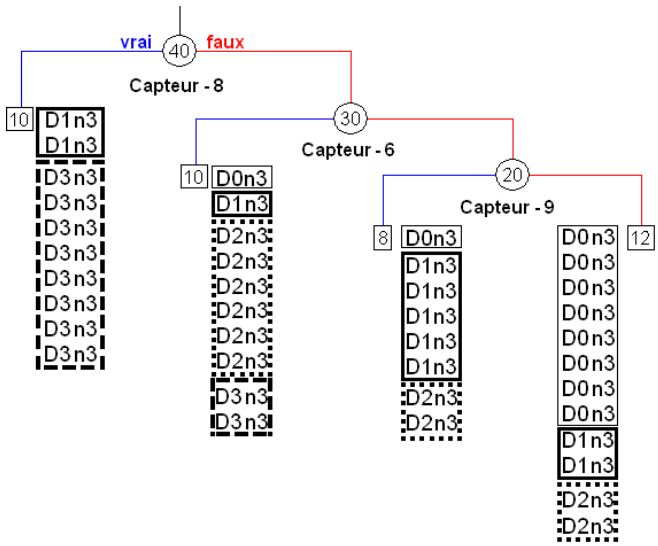


FIGURE A.5 – Classification des signaux avec bruit (n3) utilisant les divisions hiérarchiques (intervalles).

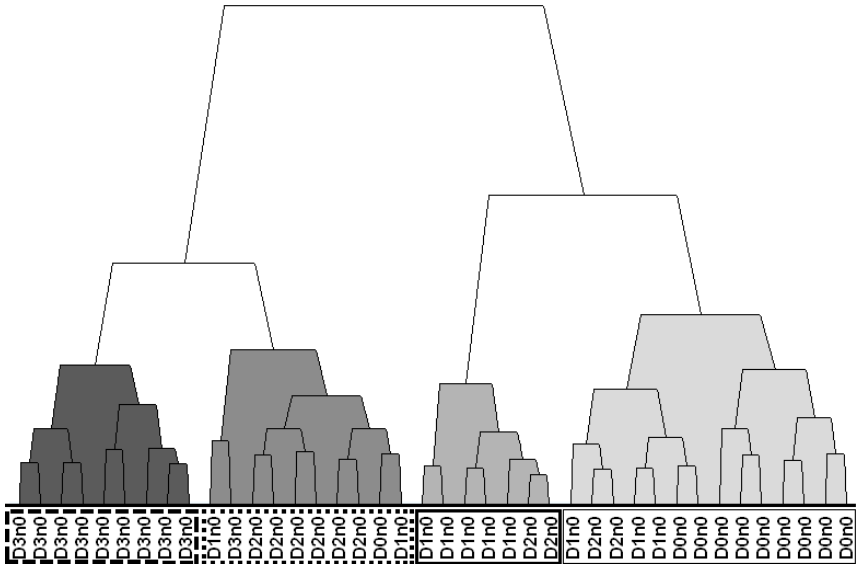


FIGURE A.6 – Classification des signaux sans bruit (n0) utilisant les agglomérations hiérarchiques (intervalles).

une tendance des classifications réalisées.

La simulation suivante consiste à appliquer 2% de bruit (n2) aux signaux. Dans ce cas, les taux de classification correcte ont diminué pour tous les scénarios. Lorsque des intervalles interquartiles sont utilisés, 70% des essais sont bien classés dans les groupes D0, D2 et D3, tandis que 60% sont classés dans D1 (Figure A.8).

Le dernier niveau de bruit est maintenant considéré (n3). En général, la méthode basée sur les agglomérations hiérarchiques est plus sensible au bruit que celle basée sur des divisions hiérarchiques. La Figure A.9 montre les résultats si des intervalles sont utilisés. Dans ce cas, les taux de classification correcte sont de 60% pour D0 et D3, 50% pour D1 et 40% pour D2. Par ailleurs, plusieurs essais se retrouvent mélangés, surtout dans le groupe d'essais D2. Un essai D3 apparaît même dans le cluster D0. Cela montre que la présence de bruit peut mener à une

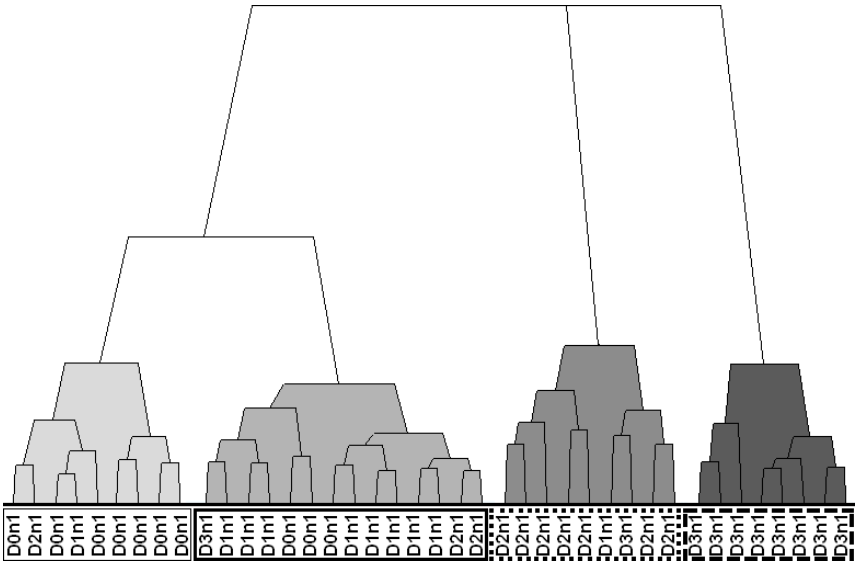


FIGURE A.7 – Classification des signaux avec bruit (n1) utilisant les agglomérations hiérarchiques (intervalles).

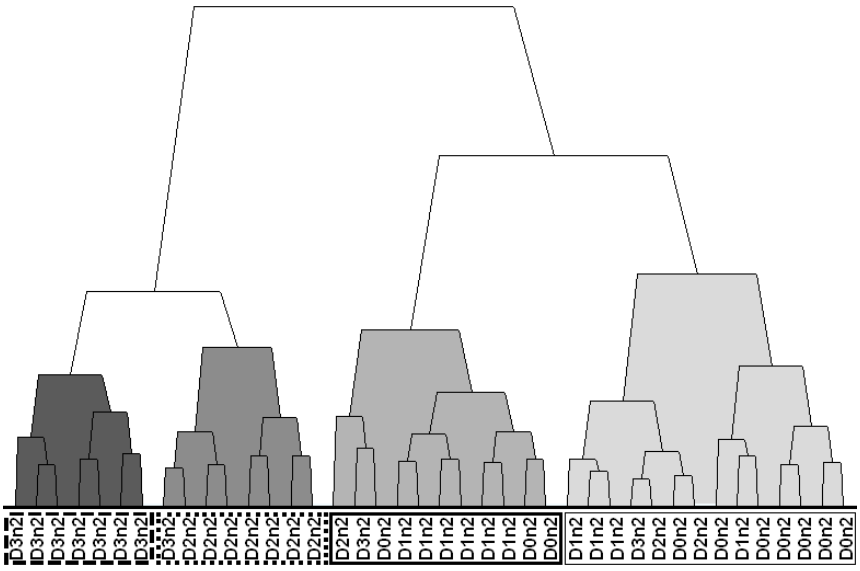


FIGURE A.8 – Classification des signaux avec bruit (n2) utilisant les agglomérations hiérarchiques (intervalles).

classification incorrecte.

Afin de compléter l'analyse des méthodes de clustering appliquées à l'exemple considéré, la méthode des nuées dynamiques est maintenant considérée. Comme décrit précédemment, pour chaque niveau de bruit, l'ensemble des 40 essais est séparé en 4 groupes où chacun représente un scénario d'endommagement. Le Tableau A.5 regroupe les taux de classification obtenus pour toutes les simulations de bruit réalisées, en utilisant des intervalles interquartiles. Dans ce tableau, chaque ligne correspond à un niveau de bruit, lorsque les colonnes sont les scénarios d'endommagement.

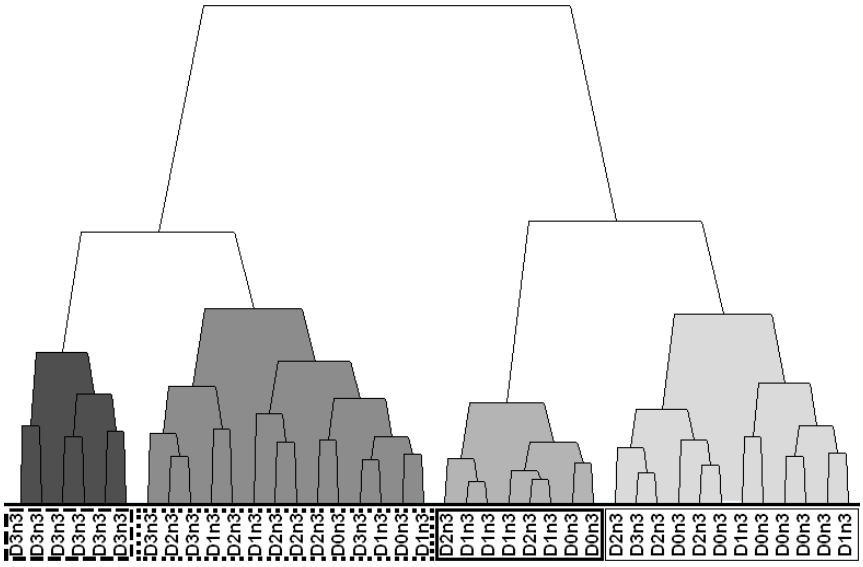


FIGURE A.9 – Classification des signaux avec bruit (n3) utilisant les agglomérations hiérarchiques (intervalles).

	Intervalles interquartiles			
	D0	D1	D2	D3
n0	90	80	70	80
n1	80	70	70	80
n2	80	70	60	80
n3	80	70	60	80

TABLEAU A.5 – Taux de classification correcte obtenus en utilisant les signaux appliqués aux nuées dynamiques (valeurs en %).

A.2.1.2 Fréquences propres

Le couplage entre l'ADS et les méthodes de clustering appliquées aux fréquences propres, préalablement identifiées, est maintenant considéré. La méthode des divisions hiérarchiques est considérée en premier. Pour toutes les simulations de bruit, 100% de classification correcte est obtenue pour les quatre scénarios d'endommagement. Ce résultat est important car il montre que les fréquences propres transformées en données symboliques peuvent être considérées comme des indicateurs pertinents pour détecter des modifications structurales. La Figure A.10 montre les arbres construits pour le dernier niveau de bruit, n3 (les arbres obtenus pour les autres niveaux de bruit ne sont pas montrés, car identiques à ceux obtenus pour le niveau de bruit, n3). Bien que la seule différence entre les arbres soit les variables discriminantes, le partitionnement obtenu est toujours le même. Il est également important de noter que les états D0 et D3 sont séparés dans des clusters différents dès le premier critère de décision (les clusters D1 et D2 sont les derniers à être séparés).

La méthode des agglomérations hiérarchiques est maintenant considérée. Comme dans le cas des divisions hiérarchiques, 100% des fréquences propres sont classées correctement, quel que soit le niveau de bruit ou le type de transformation en données symboliques utilisé. Les sorties graphiques de cette méthode ne sont pas montrées, par soucis de clarté du texte.

Le Tableau A.6 détaille les taux de classification correcte obtenus en utilisant la méthode des nuées dynamiques. Lorsque les fréquences propres sont transformées en intervalles interquartiles,

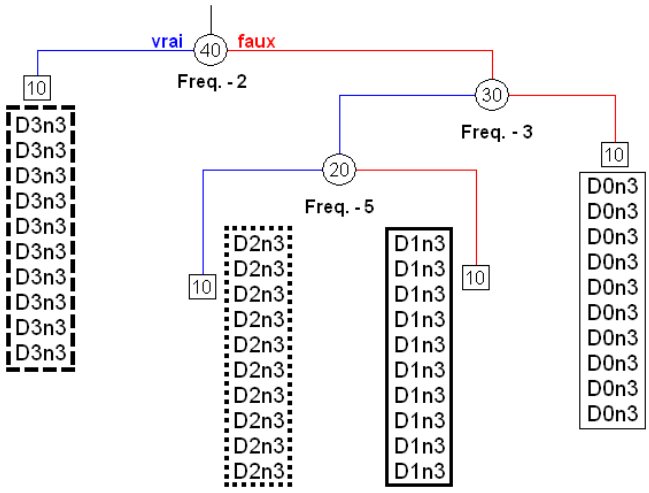


FIGURE A.10 – Classification des fréquences propres avec bruit (n3) utilisant les divisions hiérarchiques (intervalles).

la classification est parfaite uniquement dans le cas où il n'y a aucun bruit (cas n0). Dans les autres cas, le nombre d'essais D1 classés correctement diminue à mesure que le bruit augmente.

	Intervalles interquartiles			
	D0	D1	D2	D3
n0	100	100	100	100
n1	100	90	100	100
n2	100	70	100	100
n3	100	60	100	100

TABLEAU A.6 – Taux de classification correcte obtenus en utilisant les fréquences propres appliquées aux nuées dynamiques (valeurs en %).

A.2.1.3 Modes propres

La troisième partie de cette étude considère les cinq premiers modes propres transformés en données symboliques. Dans un premier temps, ces quantités transformées en intervalles sont appliquées à la méthode des divisions hiérarchiques. La Figure A.11 illustre l'arbre de décision obtenu dans le cas sans bruit. Une classification satisfaisante est obtenue : 100% pour D0, et 90% pour les autres scénarios.

Les simulations concernant le premier niveau de bruit (n1) sont maintenant considérées. Comme il a été observé avec l'utilisation directe des signaux, le nombre de classifications correctes diminue avec la présence de bruit. Quand des intervalles interquartiles sont utilisés, 90% des essais appartenant au cluster D0 sont correctement classés, tandis que 80% le sont pour les clusters D1, D2 et D3.

Des simulations avec 2% de bruit sont également réalisées. La Figure A.12 montre l'arbre construit. Les taux d'essais correctement classés sont légèrement plus faibles. Manifestement, l'ajout de bruit cause cette diminution.

Pour le troisième niveau de bruit, les pourcentages de classification correcte sont de 70% pour D0 et D1, et 80% pour D2 et D3 (cf. Figure A.14).

La méthode des agglomérations hiérarchiques est maintenant appliquée aux modes propres

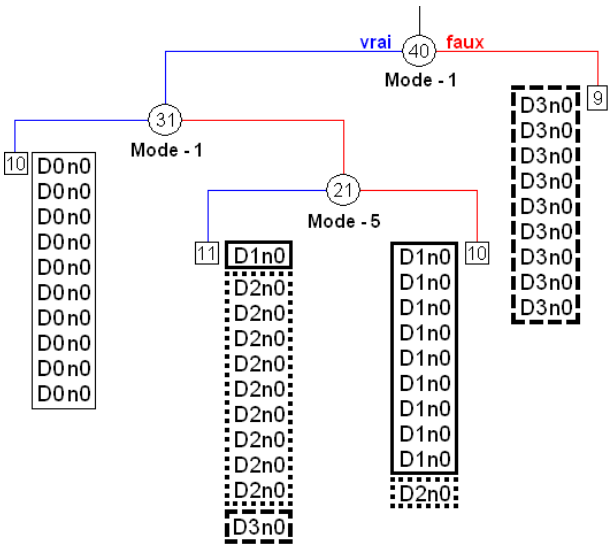


FIGURE A.11 – Classification des modes propres sans bruit (n0) utilisant les divisions hiérarchiques (intervalles).

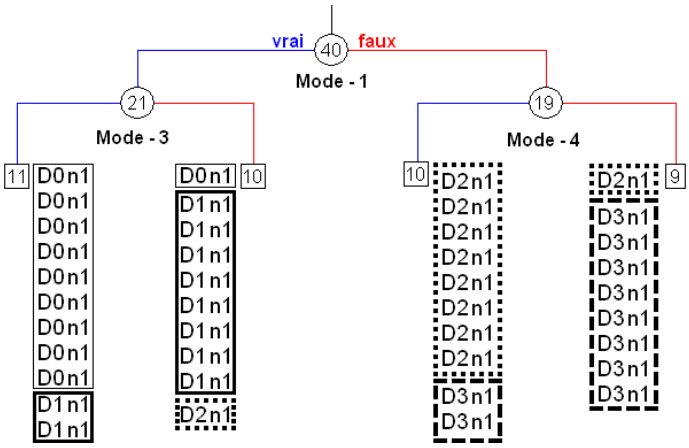


FIGURE A.12 – Classification des modes propres avec bruit (n1) utilisant les divisions hiérarchiques (intervalles).

identifiés à partir des signaux non bruités. Dans ce cas, une bonne classification est permet d'atteindre : 90% pour les états D0 et D2 et 100% pour les états D1 et D3 (cf. Figure A.15).

La Figure A.16 illustre les résultats obtenus pour le niveau de bruit n1. Il est constaté que les modes propres sont plus sensibles au bruit que les fréquences propres. En considérant des intervalles interquartiles, 80% des essais sont correctement classés quel que soit le scénario d'endommagement considéré.

Les simulations avec le niveau de bruit n2 sont également considérées. Dans ce cas, plusieurs essais se retrouvent mélangés, notamment dans les groupes D1 et D2. La Figure A.17 illustre l'agglomération obtenue pour la transformation en intervalles interquartiles. Le pourcentage de classification correcte est relativement faible : 50% pour D0, 60% pour D1 et D2 et 90% pour D3.

Les agglomérations construites pour le dernier niveau de bruit sont présentées à la Figure A.18. Des taux relativement faibles sont obtenus : 60% pour D0, 70% pour D1, et 50% pour D2 et D3. Dans cette analyse, les essais des groupes D2 et D3 se retrouvent isolés des

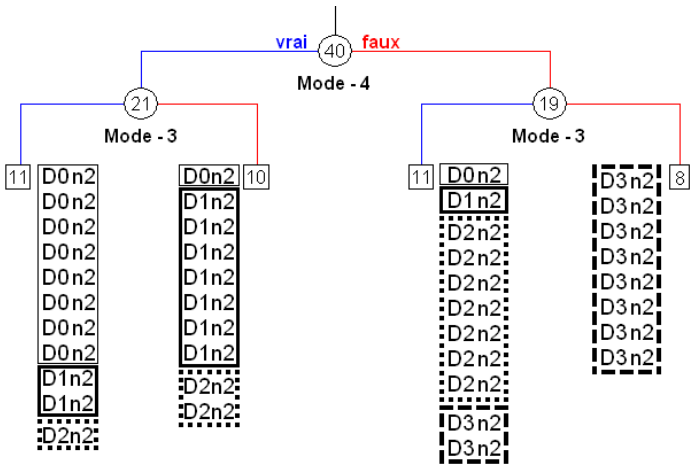


FIGURE A.13 – Classification des modes propres avec bruit (n2) utilisant les divisions hiérarchiques (intervalles).

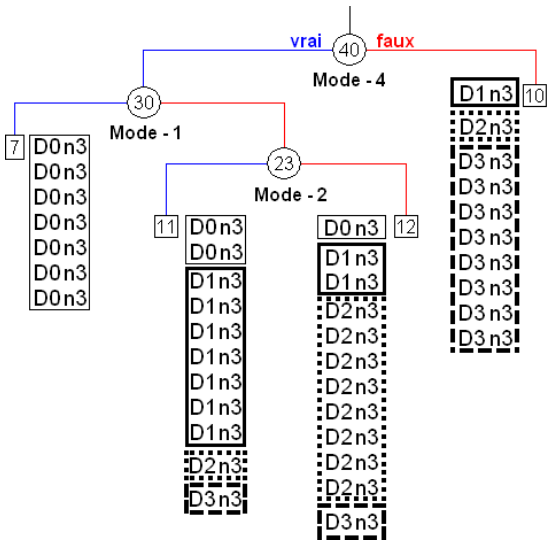


FIGURE A.14 – Classification des modes propres avec bruit (n3) utilisant les divisions hiérarchiques (intervalles).

autres, tandis que les groupes D0 et D1 contiennent des essais appartenant à des états structuraux distincts. Cela caractérise l'influence du bruit dans la classification. Cela peut également indiquer que cette méthode n'est pas suffisamment sensible pour discriminer ces états.

La méthode de clustering basée sur les nuées dynamiques est maintenant appliquée aux modes propres. Le Tableau A.7 détaille les taux de classification obtenus. Des pourcentages de classification correcte satisfaisants sont généralement obtenus.

A.2.2 Bilan des résultats

Dans ce paragraphe, un bilan des résultats obtenus par les méthodes de clustering est présenté. Les Figures A.19 et A.20 regroupent les pourcentages obtenus pour chaque méthode de partitionnement lorsque les signaux et les modes propres sont utilisés et convertis soit en intervalles interquartiles soit en histogrammes, respectivement. Pour chacune de ces figures, les taux de classification sont regroupés par niveau de bruit. De cette façon, chaque barre verticale corres-

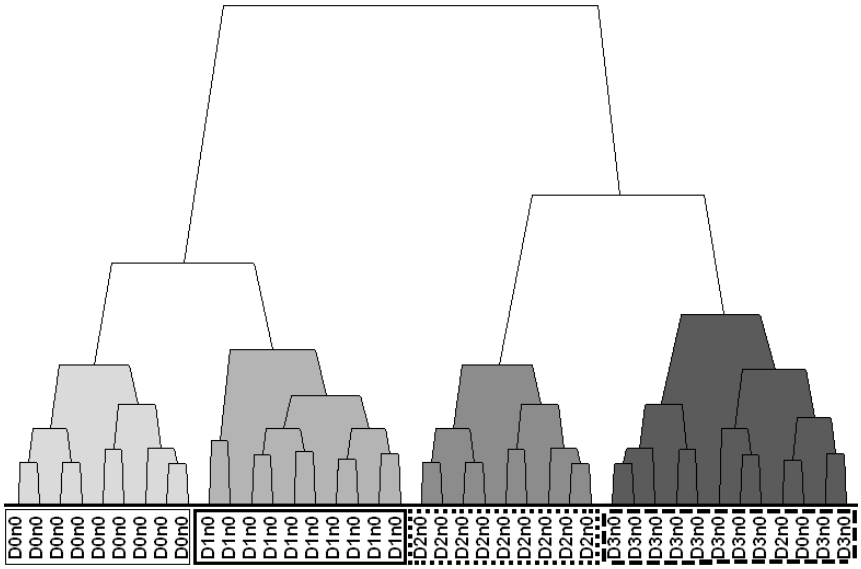


FIGURE A.15 – Classification des modes propres sans bruit (n0) utilisant les agglomérations hiérarchiques (intervalles).

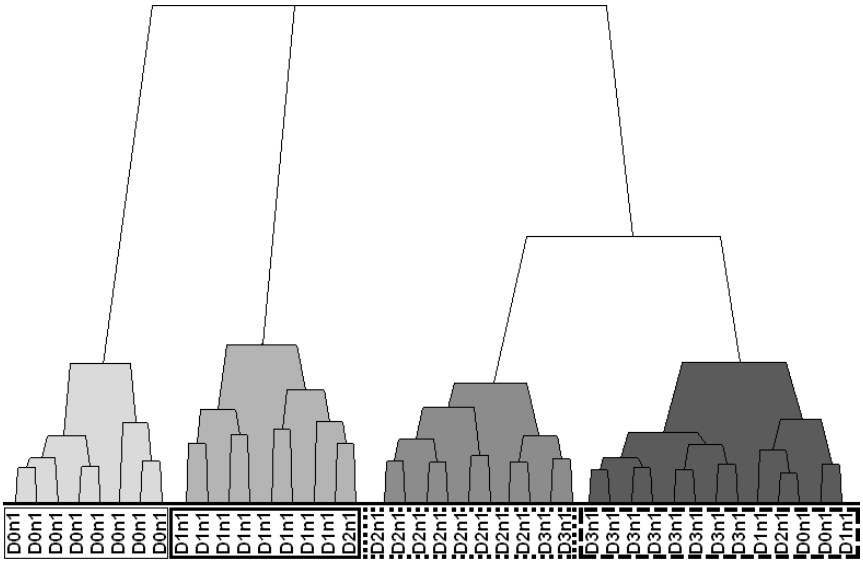


FIGURE A.16 – Classification des modes propres avec bruit (n1) utilisant les agglomérations hiérarchiques (intervalles).

	Intervalles interquartiles			
	D0	D1	D2	D3
n0	100	90	90	90
n1	90	80	70	70
n2	90	80	70	60
n3	90	80	80	80

TABLEAU A.7 – Taux de classification correcte obtenus en utilisant les modes propres appliqués aux nuées dynamiques (valeurs en %).

pond à un niveau d'endommagement différent. La discussion détaillée des résultats est présentée dans le chapitre 4.

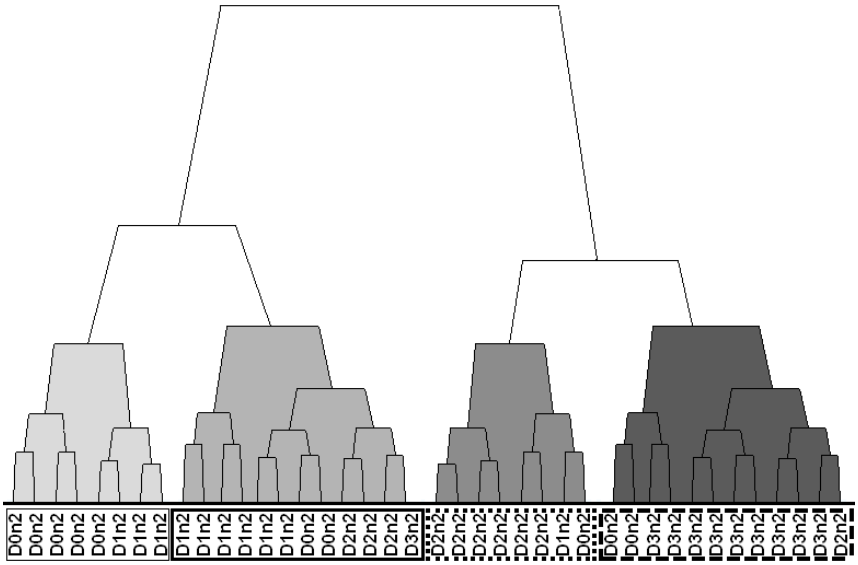


FIGURE A.17 – Classification des modes propres avec bruit (n2) utilisant les agglomérations hiérarchiques (intervalles).

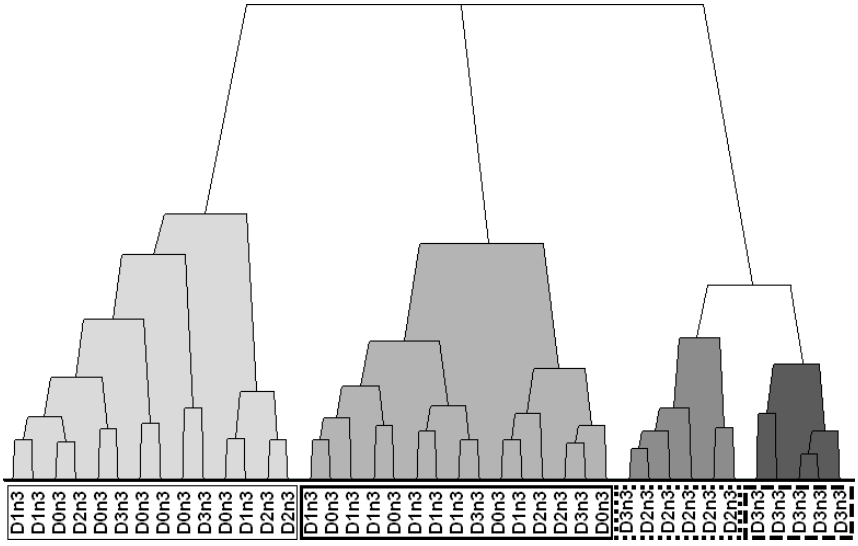
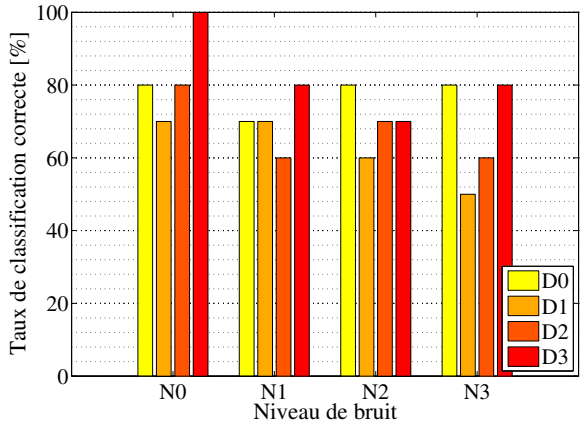


FIGURE A.18 – Classification des modes propres avec bruit (n3) utilisant les agglomérations hiérarchiques (intervalles).

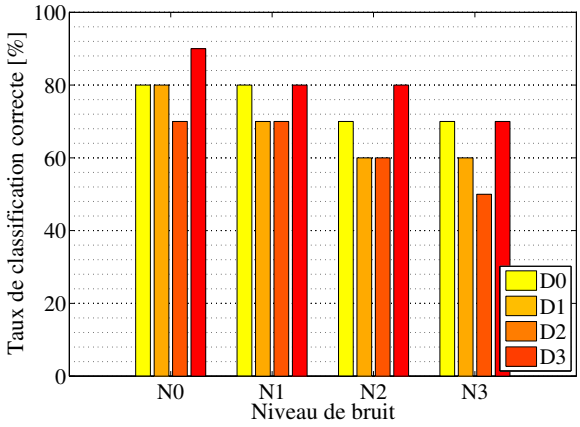
A.2.3 Études complémentaires

Dans ce paragraphe, trois analyses complémentaires sont considérées. La première analyse consiste à utiliser l'ensemble des essais de type D0 et D3 avec tous les niveaux de bruit confondus. Cela produit donc un ensemble de 80 essais (40 pour D0 et 40 pour D3). Une analyse similaire est ensuite réalisée en ne considérant que les états structuraux D1 et D2. Enfin, l'ensemble des 160 essais est utilisé. Dans ce cas, tous les niveaux d'endommagement et de bruit sont considérés. Le but est de discriminer les différents états structuraux indépendamment du niveau de bruit existant.

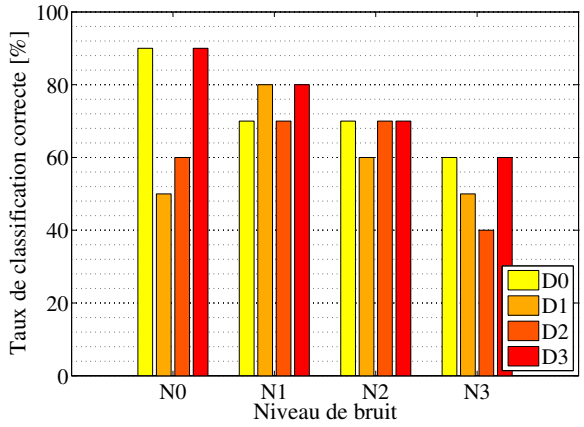
En premier lieu, les états D0 et D3 (représentant la structure non dégradée et celui le plus dégradée, respectivement) sont considérés. Le Tableau A.8 regroupe les taux de classification correcte obtenus par chacune des méthodes de partitionnement. La classification obtenue en utilisant



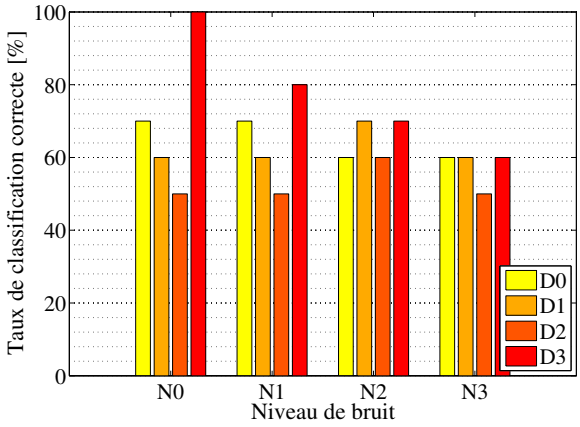
(a) Divisions hiérarchiques appliquées aux signaux transformés en intervalles interquartiles



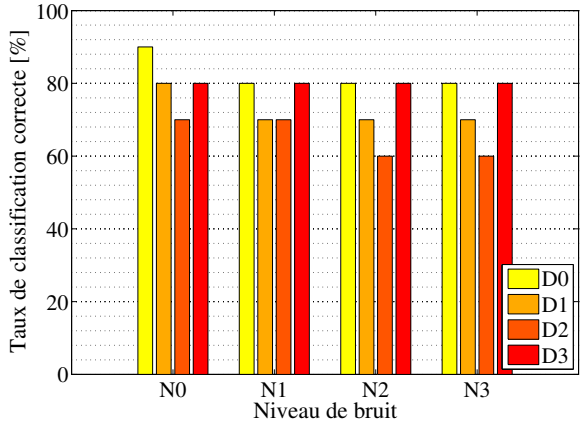
(b) Divisions hiérarchiques appliquées aux signaux transformés en histogrammes



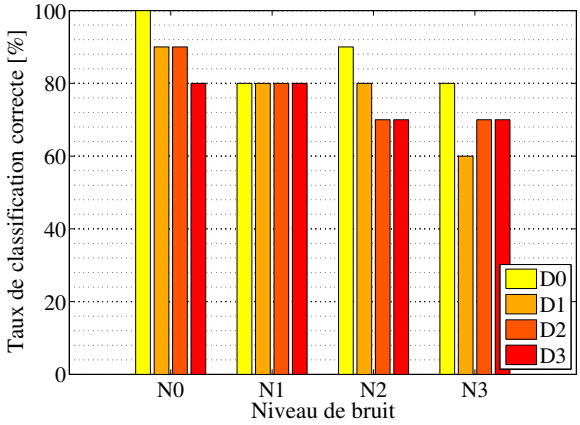
(c) Agglomérations hiérarchiques appliquées aux signaux transformés en intervalles interquartiles



(d) Agglomérations hiérarchiques appliquées aux signaux transformés en histogrammes



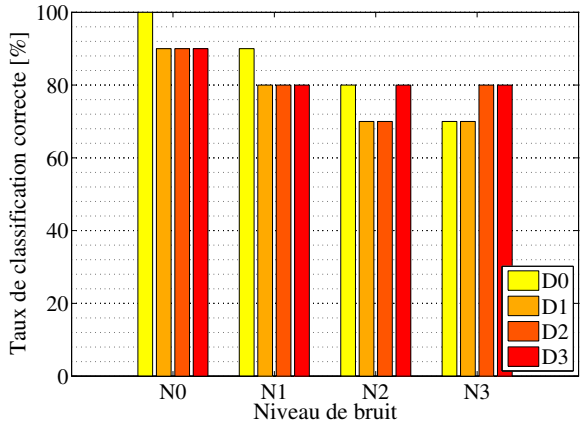
(e) Nuées dynamiques appliquées aux signaux transformés en intervalles interquartiles



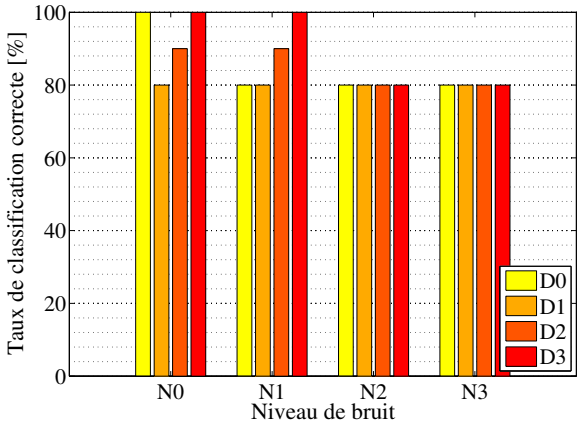
(f) Nuées dynamiques appliquées aux signaux transformés en histogrammes

FIGURE A.19 – Taux de classification correcte pour les trois méthodes de partitionnement utilisant des signaux.

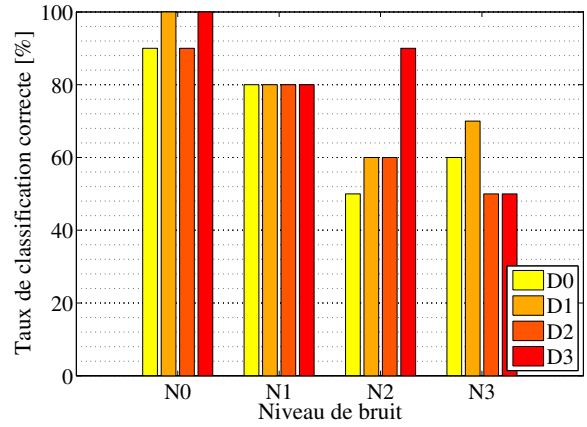
les signaux montre qu'il existe un nombre plus important d'essais mélangés, notamment dans le cas des agglomérations hiérarchiques. Il est également observé que si les fréquences propres sont considérées, la discrimination est parfaite en utilisant les méthodes des divisions hiérarchiques et



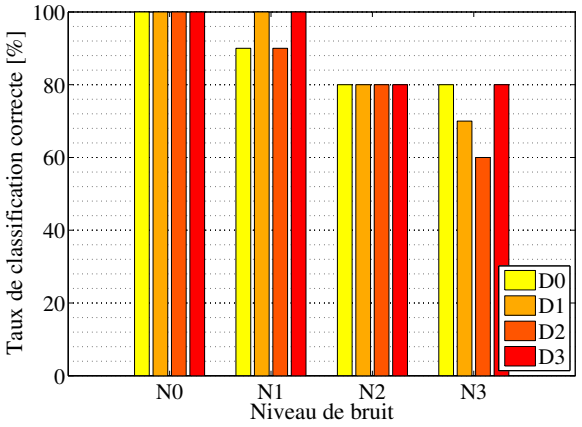
(a) Divisions hiérarchiques appliquées aux modes propres transformés en intervalles interquartiles



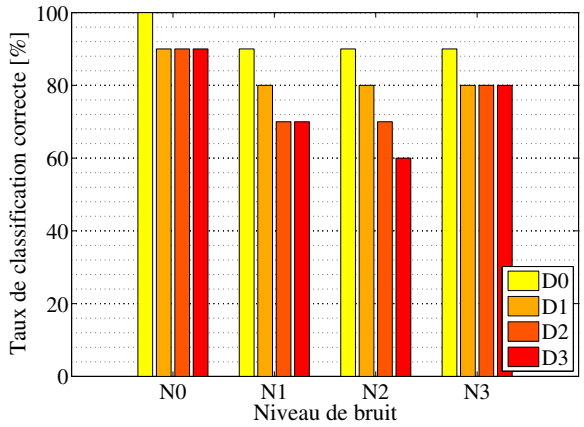
(b) Divisions hiérarchiques appliquées aux modes propres transformés en histogrammes



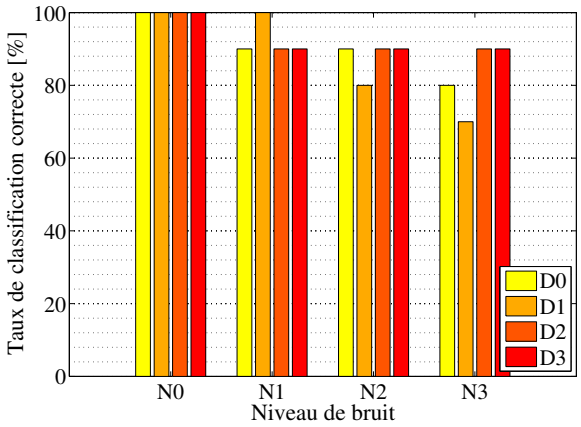
(c) Agglomérations hiérarchiques appliquées aux modes propres transformés en intervalles interquartiles



(d) Agglomérations hiérarchiques appliquées aux modes propres transformés en histogrammes



(e) Nuées dynamiques appliquées aux modes propres transformés en intervalles interquartiles



(f) Nuées dynamiques appliquées aux modes propres transformés en histogrammes

FIGURE A.20 – Taux de classification correcte pour les trois méthodes de partitionnement utilisant des modes propres.

des nuées dynamiques. La classification des modes propres est également pertinente pour ces deux méthodes. En général, les nuées dynamiques permettent d'atteindre des taux de classification les plus élevés parmi les trois méthodes.

	Signaux	Fréquences propres	Modes propres
Divisions hiérarchiques			
D0	58	100	93
D3	68	100	98
Agglomérations hiérarchiques			
D0	53	93	75
D3	58	100	83
Nuées dynamiques			
D0	73	100	98
D3	60	100	100

TABLEAU A.8 – Taux de classification correcte obtenus en utilisant les méthodes de partitionnement pour les états D0 et D3 (valeurs en %).

La deuxième analyse consiste à regrouper tous les essais de type D1 et D2. Dans ce cas, il est possible que les taux de classification correcte obtenus soient moins élevés. Les résultats obtenus en utilisant les signaux ne sont pas pertinents (cf Tableau A.9). L'utilisation des nuées dynamiques appliquées aux fréquences propres ne permet pas d'obtenir une classification parfaite. Au contraire, les taux les plus élevés sont de l'ordre de 80%. Ceci implique que pour des conditions structurales très proches, il est possible qu'un un endommagement ne soit pas détecté en utilisant les méthodes de clustering. En général, la méthode des divisions hiérarchiques permet d'obtenir des taux de classification correcte de l'ordre de 70%, ce qui est plus élevé que les pourcentages obtenus avec les agglomérations hiérarchiques (de l'ordre de 60%).

	Signaux	Fréquences propres	Modes propres
Divisions hiérarchiques			
D1	48	70	68
D2	45	70	73
Agglomérations hiérarchiques			
D1	35	63	65
D2	38	63	73
Nuées dynamiques			
D1	58	80	68
D2	58	83	70

TABLEAU A.9 – Taux de classification correcte obtenus en utilisant les méthodes de partitionnement pour les états D1 et D2 (valeurs en %).

La troisième analyse considère les 160 essais. Le Tableau A.10 regroupe les pourcentages de classification correcte calculés. Les taux obtenus en considérant les fréquences et modes propres sont à nouveau plus élevés. Pour les méthodes des divisions hiérarchiques, 100% de classification correcte est obtenue. Toutefois, aucune de ces méthodes n'arrive à classer plus de 50% des essais correctement quand les signaux sont considérés. Cela montre que la présence de bruit dans les mesures dynamiques peut conduire à des classifications incorrectes.

	Divisions hiérarchiques		
	Signaux	Fréquences	Modes
D0	43	100	83
D1	38	100	80
D2	35	100	75
D3	63	100	88
	Agglomérations hiérarchiques		
D0	38	100	75
D1	38	75	78
D2	40	80	73
D3	45	100	78
	Nuées dynamiques		
D0	55	100	88
D1	45	98	80
D2	55	98	80
D3	60	100	88

TABLEAU A.10 – Taux de classification correcte obtenus en utilisant les méthodes de partitionnement pour les états D0, D1, D2, D3 (valeurs en %).

A.2.4 Indices de partitionnement optimal

Ici, le nombre maximal de clusters est défini arbitrairement comme étant égal à 10. De cette façon, il est considéré que le nombre optimal de clusters se trouve dans l'intervalle $[2, 10]$. Les indices CH, C^* et Γ sont évalués pour chaque niveau de bruit, l'idée étant d'observer comment le bruit affecte la détermination du nombre correct de clusters.

A.2.4.1 Signaux

Le Tableau A.11 regroupe les indices calculés en utilisant la méthode des divisions hiérarchiques appliquées aux signaux, transformés en intervalles interquartiles. Dans ce tableau, il est noté que les indices CH et Γ indiquent correctement le nombre de clusters (i.e., 4 clusters) dans les cas ou il n'existe pas de bruit (n0) ou lorsque le bruit est faible (n1). Il est rappelé que les 4 clusters correspondent aux scénarios d'endommagements D0, D1, D2 et D3. Toutefois, pour les niveaux plus élevés de bruit tels que les niveaux n2 et n3, les indices CH et Γ indiquent chacun 5 clusters comme étant le partitionnement optimal. Cela montre clairement l'influence du bruit sur les résultats obtenus. L'indice C^* est également calculé et oscille entre 8 et 10 clusters, ne permettant pas de tirer des conclusions pertinentes.

Par ailleurs, la méthode des nuées dynamiques est utilisée. Le Tableau A.12 regroupe les valeurs de chaque indice dans le cas où les signaux sont transformés en intervalles interquartiles. Les indices CH et Γ indiquent correctement quatre clusters pour tous les niveaux de bruit (à l'exception du niveau n2, pour CH).

A.2.4.2 Fréquences propres

Le Tableau A.13 regroupe les valeurs calculées. Dans ce cas, il est noté un bon accord entre les trois indices. Bien que C^* présente une variation plus importante que les autres indices, il indique correctement le nombre de clusters pour les niveaux de bruit n0 et n2. Les indices CH et Γ signalent toujours le nombre correct (i.e., quatre groupes).

	n0			n1			n2			n3		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
10	22,5	**0,038**	0,68	11,60	0,019	0,26	27,7	0,22	0,80	17,6	0,09	0,58
9	21,1	0,041	0,68	11,14	**0,016**	0,27	30,0	0,19	0,74	30,1	**0,03**	0,71
8	20,6	0,069	0,70	9,10	0,040	0,28	30,7	**0,16**	0,69	21,0	0,08	0,71
7	23,7	0,081	0,67	11,83	0,047	0,26	25,4	0,30	0,74	23,6	0,11	0,70
6	27,0	0,074	0,61	13,49	0,046	0,22	41,8	0,22	0,82	25,3	0,15	0,68
5	40,7	0,082	0,78	23,15	0,035	0,34	40,9	0,30	**0,84**	**44,8**	0,08	**0,74**
4	**50,5**	0,108	**0,80**	**29,55**	0,062	**0,35**	**45,0**	0,26	0,67	34,3	0,16	0,73
3	33,2	0,131	0,71	17,24	0,060	0,31	38,0	0,50	0,71	33,8	0,19	0,67
2	30,9	0,220	0,68	17,56	0,075	0,34	34,8	0,40	0,76	39,5	0,27	0,68

TABLEAU A.11 – Indices basés sur les divisions hiérarchiques appliquées aux signaux transformés en intervalles interquartiles.

	n0			n1			n2			n3		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
10	46,2	**0,01**	0,73	25,4	0,09	0,80	58,6	0,08	0,66	63,5	0,08	0,70
9	41,8	0,04	0,57	12,3	0,09	0,73	67,6	0,05	0,64	60,5	0,06	0,74
8	35,6	0,07	0,61	24,2	0,07	0,71	54,3	0,10	0,69	66,5	**0,05**	0,74
7	33,4	0,11	0,61	29,1	0,06	0,75	39,3	0,23	0,69	71,7	0,08	0,75
6	37,9	0,02	0,62	30,8	0,08	0,74	53,8	**0,03**	0,59	67,4	0,09	0,75
5	40,4	0,14	0,56	27,1	0,06	0,82	60,9	0,33	0,55	69,0	0,09	0,75
4	**51,3**	0,11	**0,78**	**42,6**	0,06	**0,92**	60,1	0,23	**0,74**	**79,1**	0,10	**0,77**
3	34,7	0,25	0,54	13,1	0,03	0,87	**75,1**	0,48	0,49	73,9	0,14	0,70
2	33,7	0,39	0,36	6,0	**0,01**	0,86	70,8	0,55	0,30	75,1	0,22	0,67

TABLEAU A.12 – Indices basés sur les nuées dynamiques appliquées aux signaux transformés en intervalles interquartiles.

	n0			n1			n2			n3		
	CH	*C*†*	Γ	*CH*	*C*†*	Γ	*CH*	*C*†*	Γ	*CH*	*C*†*	Γ
10	3660	1,7	0,91	1112	1,0	0,52	2134	4,0	0,467	3861	**0,2**	0,52
9	2928	1,6	0,91	1243	1,1	0,72	2523	2,1	0,562	3128	4,5	0,51
8	2368	1,4	0,90	1207	1,3	0,71	2504	1,7	0,523	3327	6,6	0,76
7	1190	4,5	0,88	1286	1,1	0,74	2830	1,1	0,673	3105	11,6	0,65
6	1375	2,6	0,91	1570	**0,9**	0,77	3461	0,8	0,708	3243	4,7	0,65
5	3286	1,8	0,93	1705	1,2	0,80	3558	0,8	0,753	2697	4,2	0,70
4	**5640**	**0,3**	**0,98**	**1855**	2,4	**0,83**	**3630**	**0,4**	**0,903**	**4956**	4,8	**0,90**
3	3892	1,5	0,87	1008	3,1	0,62	1885	0,5	0,739	4781	13,1	0,84
2	738	1,2	0,79	896	4,8	0,46	1987	0,7	0,636	3753	15,4	0,83

TABLEAU A.13 – Indices basés sur les divisions hiérarchiques appliquées aux fréquences transformées en intervalles interquartiles. † Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Les résultats obtenus en utilisant la méthode des nuées dynamiques sont présentés dans le Tableau A.14. À l'exception de l'indice C^* pour les niveaux n1, n2 et n3, les autres indices indiquent toujours le nombre correct de groupes existants.

	n0			n1			n2			n3		
	CH	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ
10	2283	6,6	0,50	3557	**2,6**	0,45	2971	5,0	0,21	1677	3,6	0,60
9	2014	1,1	0,47	2802	8,5	0,58	2313	5,3	0,26	510	3,7	0,88
8	1734	0,6	0,68	2950	9,5	0,67	2695	4,5	0,62	1199	4,8	0,82
7	1245	1,0	0,56	2813	13,0	0,64	3244	3,6	0,56	2503	3,6	0,86
6	1473	1,5	0,55	2806	14,8	0,65	3809	4,0	0,56	2545	**2,7**	0,89
5	2423	1,3	0,67	2237	13,6	0,69	3831	**1,9**	0,71	3971	4,8	0,89
4	**3602**	**0,2**	**0,80**	**5166**	11,7	**0,77**	**4515**	2,2	**0,93**	**5881**	10,8	**0,91**
3	2355	1,1	0,71	3948	10,0	0,62	2383	4,5	0,83	5238	13,6	0,61
2	838	2,1	0,67	3565	4,1	0,52	2158	5,0	0,81	5073	21,5	0,38

TABLEAU A.14 – Indices basés sur les nuées dynamiques appliquées aux fréquences transformés en intervalles interquartiles.† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-3}$).

A.2.4.3 Modes propres

Le Tableau A.15 regroupe les valeurs calculées pour la méthode des divisions hiérarchiques. Il est noté que, contrairement à ce qui a été observé pour les fréquences, l'indice C^* n'indique jamais le nombre correct de groupes. Pour les autres indices, la valeur optimale est toujours égale à quatre.

	n0			n1			n2			n3		
	CH	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ
10	40,7	3,6	0,77	34,3	**0,4**	0,55	48,8	6,8	0,66	22,2	0,094	0,57
9	38,5	3,3	0,76	41,9	0,8	0,48	41,8	7,6	0,64	21,5	5,0	0,54
8	46,0	4,0	0,77	43,8	1,1	0,50	47,1	6,5	0,70	18,6	4,3	0,51
7	39,8	5,7	0,77	40,9	1,6	0,61	52,4	6,8	0,72	25,6	3,4	0,69
6	52,5	**2,1**	0,79	47,9	1,6	0,68	51,9	**4,1**	0,70	25,2	2,1	0,76
5	53,7	2,6	0,74	52,2	2,2	0,68	48,9	5,9	0,72	22,8	**1,4**	0,68
4	**64,7**	2,5	**0,97**	**57,9**	3,0	**0,82**	**74,7**	4,3	**0,74**	**40,1**	2,7	**0,91**
3	54,1	5,8	0,89	50,5	4,9	0,72	66,0	0,088	0,65	38,3	2,8	0,75
2	42,0	7,1	0,88	47,7	0,82	0,66	51,5	0,190	0,54	31,5	3,0	0,67

TABLEAU A.15 – Indices basés sur les divisions hiérarchiques appliquées aux modes propres transformées en intervalles interquartiles.† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Les indices sont maintenant calculés pour les partitionnements obtenus par la méthode des nuées dynamiques. Le Tableau A.16 contient les valeurs calculées pour chaque indice. Tous les indices arrivent à indiquer correctement le nombre de groupes existants pour les deux premiers niveaux de bruit (niveaux n0 et n1). Toutefois, pour les niveaux n2 et n3, seuls les indices CH et Γ permettent de trouver quatre groupes comme étant le partitionnement optimal.

A.2.4.4 Études complémentaires

L'objectif de ce paragraphe est de vérifier si les indices de partitionnement optimal sont capables d'indiquer correctement le nombre de groupes existants (dans cet exemple, deux) lors des études complémentaires de clustering réalisées au paragraphe 4.3.2.

	n0			n1			n2			n3		
	CH	*C*†*	Γ	*CH*	*C*†*	Γ	*CH*	*C*†*	Γ	*CH*	*C*†*	Γ
10	17,6	2,1	0,79	24,3	2,4	0,52	18,6	6,8	0,66	29,6	**1,0**	0,69
9	14,8	2,0	0,76	22,0	2,8	0,56	18,7	2,7	0,62	34,1	1,5	0,62
8	17,5	1,6	0,80	22,9	3,3	0,73	16,2	2,3	0,63	28,7	1,6	0,69
7	19,1	2,5	0,77	23,8	4,2	0,69	11,3	18,6	0,69	33,1	2,5	0,67
6	18,9	1,8	0,83	25,8	2,8	0,70	19,0	**0,5**	0,74	35,8	2,3	0,68
5	17,8	2,0	0,85	24,0	2,2	0,75	17,7	6,1	0,71	33,9	3,7	0,74
4	**26,6**	**1,0**	**0,88**	**35,6**	**2,1**	**0,94**	**38,2**	37,1	**0,90**	**61,8**	3,0	**0,77**
3	23,9	2,3	0,65	30,4	5,5	0,83	25,1	31,7	0,81	49,6	5,9	0,68
2	18,1	8,3	0,41	25,0	7,2	0,79	22,7	58,7	0,77	51,2	9,7	0,63

TABLEAU A.16 – Indices basés sur les nuées dynamiques appliquées aux modes propres transformés en intervalles interquartiles.† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Scénarios D0 et D3

Le Tableau A.17 regroupe les valeurs calculées en utilisant les méthodes des divisions hiérarchiques et des nuées dynamiques appliquées aux signaux transformés en intervalles interquartiles. Pour deux méthodes, seuls les indices CH et Γ considèrent deux clusters comme étant le partitionnement optimal. L'indice C^* signale toujours un nombre de groupes plus élevé.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	30,9	**0,082**	0,61	27,7	**0,072**	0,65
4	33,2	0,108	0,67	20,8	0,072	0,71
3	40,7	0,131	0,7	23,2	0,074	0,77
2	**50,5**	0,22	**0,75**	**31,7**	0,085	**0,82**

TABLEAU A.17 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en intervalles (scénarios D0 et D3).

Dans le cas des fréquences propres, pour les deux méthodes, tous les indices indiquent le nombre correct de clusters (cf. Tableau A.18). Cela confirme le taux de 100% de réussite des classifications obtenu précédemment.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	73,8	12,1	0,79	43,1	3,3	0,66
4	328,6	1,8	0,98	88,2	5,3	0,7
3	389,2	1,5	0,98	171,4	7,1	0,94
2	**564,0**	**0,3**	**0,99**	**177,7**	**2,4**	**0,95**

TABLEAU A.18 – Indices basés sur les méthodes de clustering appliquées aux fréquences transformées en intervalles (scénarios D0 et D3).† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Une analyse similaire est menée en considérant les modes propres. En général, une bonne adéquation des indices CH et Γ est observée, étant donné qu'ils indiquent toujours correctement le nombre de groupes attendu. L'indice C^*, toutefois, varie significativement (cf. Tableau A.19).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	71	10	0,17	128,8	**3,1**	0,38
4	77,3	8,1	0,25	82,6	7,5	0,34
3	167,3	8,3	0,57	131,7	11,1	0,55
2	**223,9**	**1,6**	**1,00**	**242,4**	5,6	**0,69**

TABLEAU A.19 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en intervalles (scénarios D0 et D3).† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Scénarios D1 et D2

La deuxième étude complémentaire réalisée a utilisé l'ensemble des essais correspondant aux états structuraux D1 et D2. Le Tableau A.20 regroupe les valeurs calculées pour les deux méthodes de partitionnement en utilisant la transformation en intervalles interquartiles. Dans ce cas, il est noté que seuls les indices CH et Γ calculés à partir de la méthode des nuées dynamiques indiquent le nombre correct de clusters. Cela est en accord avec les taux de classification correcte plus importants obtenus avec cette méthode que ceux obtenus avec les divisions hiérarchiques (cf. Tableau A.9). Pour cette dernière, aucun indice n'indique le nombre attendu de groupes.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	99,9	**3,8**	0,72	92,9	**3,4**	0,74
4	**105,4**	5	0,68	67,4	4,2	0,71
3	64,6	6,4	**0,84**	85,2	8,9	0,64
2	89,8	10,9	0,79	**96,3**	11,1	**0,83**

TABLEAU A.20 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en intervalles (scénarios D1 et D2).† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Les résultats en utilisant les fréquences propres sont significativement meilleurs que ceux obtenus avec les signaux. En effet, pour toutes les méthodes, les indices CH et Γ signalent le nombre correct de clusters, comme montré dans le Tableau A.21. Par ailleurs, l'indice C^* oscille entre 2 et 4 clusters, montrant que cet indice est trop sensible aux petites variations structurales.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	46,7	42,9	0,3	38,6	7,5	0,61
4	48,7	30,7	0,57	28,8	6,9	0,74
3	59,8	8	0,86	34,5	**6,7**	0,77
2	**67,0**	**2,5**	**0,89**	**40,7**	65,2	**0,89**

TABLEAU A.21 – Indices basés sur les méthodes de clustering appliquées aux fréquences transformées en intervalles (scénarios D1 et D2).† Les valeurs de l'indice C^* sont multipliées par ($\times 10^{-2}$).

Les modes propres sont considérés dans le Tableau A.22 qui détaille les valeurs calculées pour chaque indice. Pour les deux méthodes, seuls l'indice CH et Γ signalent deux clusters comme étant le partitionnement optimal.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	18,2	**0,132**	0,56	33,8	**0,075**	0,65
4	21,9	0,151	0,41	24,6	0,213	0,59
3	26,3	0,212	0,61	35,1	0,163	0,62
2	**36,1**	0,356	**0,77**	**40,1**	0,3	**0,67**

TABLEAU A.22 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en intervalles (scénarios D1 et D2).

Scénarios D0, D1, D2 et D3

La troisième partie de l'étude complémentaire a considéré les 160 essais. Le Tableau A.23 regroupe les valeurs calculées pour les trois indices à partir des méthodes de partitionnement. Il est noté que les indices CH et Γ signalent quatre clusters lorsque la méthode des nuées dynamiques est utilisée. Pour la méthode des divisions hiérarchiques, toutefois, seul l'indice CH indique le nombre correct de groupes existants.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
10	61,4	**3,4**	0,66	64,8	4,9	0,75
9	57,3	3,5	0,71	67,1	**3,3**	0,72
8	62	3,5	0,77	70,2	4	0,71
7	57,6	5	**0,79**	71,3	4,7	0,69
6	71,6	4,7	0,77	69,7	6,5	0,76
5	73,7	6	0,78	77,6	6	0,78
4	**94,5**	7,3	0,78	**126,1**	8	**0,82**
3	83,9	11,4	0,78	90	10,2	0,77
2	85,2	15,9	0,68	87,1	13,3	0,76

TABLEAU A.23 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en intervalles (scénarios D0, D1, D2 et D3).† Les valeurs de l'indice C^* sont multipliées par $(\times 10^{-2})$.

Dans le cas des fréquences propres, tous les indices sont capables de confirmer le partitionnement optimal avec quatre clusters (cf. Tableaux A.24). Il est également noté que l'indice Γ varie très peu, mais que sa valeur maximale est toujours associée au nombre correct de groupes.

En considérant les modes propres, les résultats sont détaillés dans le Tableau A.25. De façon générale, les indices CH et Γ indiquent dans la plupart des cas quatre clusters comme le partitionnement optimal.

A.3 Méthodes supervisées

Dans ce paragraphe, les résultats obtenus à partir du couplage entre l'ADS et les méthodes de classification supervisée sont présentés. Dans un premier temps, ces méthodes sont appliquées aux représentations symboliques des signaux. Dans un deuxième temps, la même procédure est appliquée aux fréquences propres. Finalement, les modes propres sont utilisés pour la discrimination des différents états structuraux.

Paramètres utilisés

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
10	427	2,9	0,88	742	4,4	0,86
9	397	3,2	0,86	688	4,3	0,93
8	440	4,8	0,87	762	4,1	0,96
7	510	4,4	0,89	885	3,1	0,98
6	598	3,6	0,89	1037	3	0,98
5	734	2,1	0,92	1271	2,2	0,99
4	**956**	**0,6**	**1,00**	**1656**	**1,0**	**1,00**
3	701	1,1	0,88	1214	1,7	0,99
2	55	2,3	0,85	95	1,8	0,98

TABLEAU A.24 – Indices basés sur les méthodes de clustering appliquées aux fréquences transformées en intervalles (scénarios D0, D1, D2 et D3).† Les valeurs de l'indice *C** sont multipliées par ($\times 10^{-2}$).

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
10	18,8	4,4	0,47	10,2	**1,5**	0,59
9	17,6	**2,1**	0,51	11,5	1,6	0,6
8	18,9	13,8	0,48	10,8	1,9	0,59
7	17,3	19,7	0,5	12,1	2,6	0,58
6	22,2	19,3	0,52	11,9	2,6	0,6
5	22,6	19,5	**0,53**	13,5	3,3	0,55
4	**42,1**	20,3	0,51	**18,1**	5,3	**0,63**
3	29,1	27,6	0,32	16,4	51,2	0,54
2	26	68,2	0,28	8,7	56,7	0,47

TABLEAU A.25 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en intervalles (scénarios D0, D1, D2 et D3).

Les simulations utilisant les ADB (Arbres de Décision Bayésiens) sont réalisées sous MatLab® à l'aide des routines développées par l'auteur. Pour la méthode basée sur les RN (réseaux de neurones), l'architecture du réseau utilisée est composée d'une couche cachée avec 20 neurones. Par ailleurs, les fonctions d'activation utilisées sont la fonction sigmoïde (cf. Équation 3.41) dans la couche cachée et la fonction Heaviside (cf. Équation 3.40) dans la couche de sortie. Le taux d'apprentissage η est fixé à 0,01 et la constante de moment μ est fixé à 0,0001. Ces paramètres ont été choisis après quelques simulations réalisées au préalable et ont montré une bonne adéquation aux problèmes traités (Cury & Crémona 2010*c*,*b*). Les simulations ont été réalisées à l'aide de la boîte à outils Netlab (Bishop 1995). En ce qui concerne les SVM (Machines à Vecteurs Supports), la fonction noyau RBF (cf. paragraphe 3.2.3.2) est utilisée et le paramètre σ est déterminé de manière itérative lors de la phase de validation. Les simulations ont été réalisées à l'aide de la boîte à outils LibSVM (Chang & Lin 2001).

A.3.1 Analyse de la sensibilité au bruit

Les simulations réalisées dans la suite suivent le même déroulement que celui détaillé dans le chapitre 4. Plusieurs configurations de groupes d'entraînement, validation et test sont proposées. Dans un premier temps, les taux de classification correcte sont calculés par niveau de bruit et en utilisant 30% d'essais pour l'entraînement, 10% pour la validation et 60% pour le test. Cela

signifie que pour chaque niveau de bruit, les 40 essais (10 essais D0, 10 essais D1, 10 essais D2 et 10 essais D3) sont répartis en 12 essais pour l'entraînement, 4 pour la validation [1] et 24 pour le test. Dans un deuxième temps, 40% d'essais (16) sont utilisés pour l'entraînement, 10% (4) pour la validation et 50% (20) pour le test. Finalement, la moitié des essais (20) font partie du groupe d'entraînement, 10% (4) de validation et 40% (16) de test.

Étant donné que ces distributions peuvent être déterminées de façon aléatoire et afin de mener une étude statistique, 10000 simulations différentes ont été réalisées (i.e., 10000 groupes différents d'entraînement, de validation et de test). Les résultats présentés dans la suite correspondent aux taux de classification maximal, moyen et minimal obtenus pour les 10000 simulations considérées. Dans le cas des méthodes de clustering, les taux de classification correcte ont été présentés par niveau d'endommagement. Ici, les taux correspondent à la proportion globale de classifications correcte obtenues. En d'autres termes, le taux affiché est le ratio entre le nombre total d'essais correctement classés (parmi les états D0, D1, D2 et D3) et le nombre total d'essais existants dans chaque analyse (40).

A.3.1.1 Signaux

Les résultats présentés dans le Tableau A.26 correspondent aux trois méthodes de classification appliquées aux signaux, transformés en intervalles interquartiles. De façon générale, la méthode des RN permet d'atteindre des taux de classification correcte les plus élevés. Les SVM présentent également une performance satisfaisante, ayant cependant les taux de classification minimal les plus faibles parmi les différentes méthodes considérées. Dans ce type de simulation, il est possible de dire que les RN permettent de mieux définir les « seuils » de séparation entre les différents états structuraux.

	ADB				**RN**				**SVM**			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	82	82	80	77	96	92	88	88	92	85	88	82
Moyen	67	65	65	61	78	78	75	74	75	72	73	78
Minimal	35	40	38	30	38	38	38	38	14	15	17	14

TABLEAU A.26 – Taux de classification correcte utilisant les signaux transformés en intervalles interquartiles (30% entr., 10% val., 60% test. Valeurs en %).

Le nombre d'essais utilisés pour l'entraînement des algorithmes est maintenant augmenté (cf. Tableaux A.27 et A.28). À l'exception des ADB, il est noté que les résultats sont améliorés pour toutes les méthodes, surtout pour les taux moyen et minimal. Ce résultat montre que ces méthodes arrivent à mieux discriminer les scénarios d'endommagent à mesure que le nombre d'essais utilisés dans l'étape d'apprentissage augmente. De manière générale, il est intéressant de noter que l'augmentation de la proportion d'essais dans le groupe d'entraînement n'a pas montré une amélioration nette des résultats pour cette méthode.

A.3.1.2 Fréquences propres

Les taux de classification évalués pour les différentes simulations réalisées en utilisant les fréquences propres transformées en intervalles interquartiles sont montrés dans les Tableaux A.29, A.30 et A.31. Comme il a été observé pour les méthodes de clustering, l'utilisation des fréquences propres permet toujours de produire des classifications plus pertinentes. Cela est remarqué pour

1. Pour la méthode des arbres de décision Bayésiens, il n'existe pas de groupes de validation. Dans ce cas, la proportion de ce groupe est ajoutée à celle du groupe d'entraînement.

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	84	82	83	80	95	85	85	90	90	85	82	81
Moyen	68	60	67	65	76	77	77	77	74	72	79	78
Minimal	38	39	35	34	40	40	45	35	16	18	13	18

TABLEAU A.27 – Taux de classification correcte utilisant les signaux transformés en intervalles interquartiles (40% entr., 10% val., 50% test. Valeurs en %).

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	85	84	81	82	94	94	94	82	93	85	85	88
Moyen	68	66	64	66	83	82	82	75	75	78	72	78
Minimal	45	42	39	40	31	38	32	25	15	14	15	23

TABLEAU A.28 – Taux de classification correcte utilisant les signaux transformés en intervalles interquartiles (50% entr., 10% val., 40% test. Valeurs en %).

toutes les méthodes, et plus spécifiquement pour les RN qui permettent d'atteindre des pourcentages plus importants. Par ailleurs les classifications sont améliorées lorsque le nombre d'essais dans le groupe d'entraînement augmente, en particulier pour les taux moyen et minimal sont analysés.

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	95	95	92	91	100	100	100	100	90	92	85	80
Moyen	86	84	85	80	98	95	86	90	90	93	88	93
Minimal	50	50	47	44	88	67	58	63	31	12	25	21

TABLEAU A.29 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles interquartiles (30% entr., 10% val., 60% test. Valeurs en %).

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	96	95	91	90	100	100	100	100	91	85	82	85
Moyen	84	83	84	82	99	98	91	91	95	92	95	90
Minimal	58	55	54	50	95	95	70	64	30	30	20	15

TABLEAU A.30 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles interquartiles (40% entr., 10% val., 50% test. Valeurs en %).

	ADB				RN				SVM			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	95	95	92	91	100	100	100	100	100	82	85	83
Moyen	83	81	80	80	100	98	97	94	91	97	91	91
Minimal	66	62	55	57	100	92	75	81	33	31	31	38

TABLEAU A.31 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles interquartiles (50% entr., 10% val., 40% test. Valeurs en %).

A.3.1.3 Modes propres

Les taux de classification correcte obtenus utilisant les modes propres sont détaillés dans les Tableaux A.32, A.33 et A.34. De façon générale, les résultats sont relativement meilleurs que ceux obtenus en utilisant les signaux, et légèrement moins bons que ceux classifications avec les fréquences propres. Il est noté que les RN et SVM ont des taux moyens similaires pour tous les niveaux de bruit. Cela montre que ces méthodes sont moins sensibles au bruit. Cependant, ceci n'est pas le cas des ADB, étant donné que les taux moyens chutent de 67% (sans bruit) à 58% (bruit n3) (cf. Tableau A.32) et de 74% à 66% (cf. Tableau A.34). La méthode des SVM a les taux de classification correcte les plus bas. Toutefois cette méthode est globalement la plus performante.

	ADB				**RN**				**SVM**			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	88	88	85	84	92	88	75	71	88	85	83	83
Moyen	67	67	60	58	70	70	67	63	78	80	77	80
Minimal	21	25	21	22	50	50	52	50	17	18	21	20

TABLEAU A.32 – Taux de classification correcte utilisant les modes propres transformés en intervalles interquartiles (30% entr., 10% val., 60% test. Valeurs en %).

	ADB				**RN**				**SVM**			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	91	90	89	89	100	95	75	75	85	90	82	80
Moyen	68	67	63	61	83	75	65	61	80	85	75	72
Minimal	33	31	33	30	55	35	59	50	20	15	21	18

TABLEAU A.33 – Taux de classification correcte utilisant les modes propres transformés en intervalles interquartiles (40% entr., 10% val., 50% test. Valeurs en %).

	ADB				**RN**				**SVM**			
	n0	n1	n2	n3	n0	n1	n2	n3	n0	n1	n2	n3
Maximal	95	92	91	91	100	87	75	88	94	88	88	81
Moyen	74	74	69	66	88	75	66	71	84	75	81	65
Minimal	30	33	35	31	81	38	35	60	25	19	19	17

TABLEAU A.34 – Taux de classification correcte utilisant les modes propres transformés en intervalles interquartiles (50% entr., 10% val., 40% test. Valeurs en %).

A.3.2 Études complémentaires

Comme dans le cas des méthodes de clustering, trois analyses complémentaires sont maintenant proposées en utilisant différents ensembles d'essais. Dans un premier temps, les 80 essais correspondant aux états D0 (40 essais) et D3 (40 essais) sont considérés. Dans un deuxième temps, la même démarche est appliquée aux états D1 (40 essais) et D2 (40 essais). Finalement, les méthodes de classification sont appliquées à la totalité des essais (160), le but étant de les classer en quatre groupes. L'objectif est d'observer si les méthodes sont capables de distinguer ces deux états structuraux pour quel que soit le niveau de bruit. Les paramètres utilisés sont essentiellement les mêmes qu'au paragraphe A.3, à l'exception de la fonction d'activation utilisée dans la couche cachée des RN. Ici, au lieu de la fonction sigmoïde, la fonction tangente

hyperbolique est employée (cf. Équation 3.42). Ce choix est lié aux résultats obtenus à partir des simulations réalisées préalablement.

A.3.2.1 Scénarios D0 et D3

Dans un premier temps, les signaux sont considérés (cf. Tableau A.35). En général, les SVM suivies des RN permettent toujours d'obtenir les meilleurs résultats. Il est également remarqué que, dans cette analyse, les ADB permettent d'obtenir des pourcentages les plus faibles. Par ailleurs, à mesure que le nombre d'essais dans le groupe d'entraînement augmente, les taux de classification améliorent pour toutes les méthodes.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	66	58	70	72	63	77	77	68	82
Moyen	42	51	55	50	52	50	55	51	65
Minimal	19	42	23	31	45	12	30	38	23

TABLEAU A.35 – Taux de classification correcte utilisant les signaux transformés en intervalles pour les scénarios D0 et D3 (valeurs en %).

Le Tableau A.36 regroupe les pourcentages de classification correcte obtenus lorsque les fréquences propres sont considérée. En général, les méthodes de classification supervisée permettent d'obtenir des résultats pertinents. Bien que les ADB n'arrivent pas à classer 100% des essais, les RN et les SVM le permettent pour pratiquement toutes les simulations.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	87	100	100	90	100	100	90	100	100
Moyen	78	100	99	83	100	100	85	100	100
Minimal	59	100	98	59	100	100	66	100	100

TABLEAU A.36 – Taux de classification correcte utilisant les fréquences transformées en intervalles pour les scénarios D0 et D3 (valeurs en %).

Des résultats satisfaisants sont également obtenus en considérant les déformées modales (cf. Tableau A.37). Dans ce cas, les RN et les SVM arrivent à classer 100% des essais et, en moyenne, cette valeur est supérieure à 90% quel que soit le nombre d'essais dans le groupe d'entraînement.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	85	98	100	85	100	100	90	100	100
Moyen	79	94	95	80	94	95	82	97	95
Minimal	51	83	90	51	88	92	55	92	92

TABLEAU A.37 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les scénarios D0 et D3 (valeurs en %).

A.3.2.2 Scénarios D1 et D2

La deuxième étude complémentaire considère les essais correspondant aux états structuraux D1 et D2. En général, les taux de classification sont de l'ordre de 50%, en moyenne, pour toutes les méthodes, en considérant la transformation des signaux en intervalles (cf. Tableau A.38).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	58	63	61	60	68	73	62	65	69
Moyen	52	48	57	49	50	61	56	50	55
Minimal	42	38	29	35	38	31	50	35	33

TABLEAU A.38 – Taux de classification correcte utilisant les signaux transformés en intervalles pour les scénarios D1 et D2 (valeurs en %).

Le Tableau A.39 regroupe les résultats lorsque les fréquences propres sont considérées. Bien que les taux soient moins importants que ceux obtenus précédemment dans le cas des états D0 et D3, les pourcentages évalués ici restent satisfaisants. Il est noté que les SVM permettent de classer 100% des essais dans les deux dernières simulations.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	88	100	100	94	100	100	98	100	100
Moyen	80	99	100	82	99	100	81	99	100
Minimal	63	94	98	56	98	100	58	97	100

TABLEAU A.39 – Taux de classification correcte utilisant les fréquences transformées en intervalles pour les scénarios D1 et D2 (valeurs en %).

Enfin, la démarche de classification est appliquée aux modes propres dans le Tableau A.40 qui détaille les pourcentages obtenus. Dans ce cas, les ADB sont associés à des taux de classification correcte relativement faibles, de l'ordre de 50%. Les RN et SVM, toutefois, permettent d'atteindre des taux relativement plus importants (entre 77 et 87%). Il est intéressant de noter qu'à mesure que le nombre d'essais dans le groupe d'entraînement augmente, les taux de classification correcte diminuent. Il est possible que les essais utilisés pour la détermination des seuils de séparation n'aient pas été suffisamment représentatifs de chaque classe.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	63	92	90	63	90	92	68	94	94
Moyen	53	77	87	55	80	80	57	81	80
Minimal	44	48	59	44	43	55	43	53	52

TABLEAU A.40 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les scénarios D1 et D2 (valeurs en %).

A.3.2.3 Scénarios D0, D1, D2 et D3

La troisième partie de l'étude complémentaire considère l'ensemble des 160 essais. L'objectif est toujours de séparer les quatre scénarios d'endommagement, en considérant tous les niveaux de bruit ensemble. Les premières analyses considèrent les signaux transformés en intervalles interquartiles. Dans ce cas, il est montré qu'en moyenne, moins de la moitié des essais est classé correctement (cf. Tableau A.41).

Les résultats sont significativement meilleurs en utilisant les fréquences propres. Les taux de classification correcte sont regroupés dans le Tableau A.42. Bien qu'ils soient légèrement inférieurs à ceux obtenus avec l'analyse des essais D0/D3, les taux de classification sont, en moyenne, de l'ordre de 80% pour les ADB et 90% pour les RN et les SVM.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	61	55	60	58	60	60	63	68	65
Moyen	28	28	33	36	38	45	38	37	44
Minimal	10	17	9	12	20	24	18	15	27

TABLEAU A.41 – Taux de classification correcte utilisant les signaux transformés en intervalles pour tous les scénarios d'endommagement (valeurs en %).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	95	100	100	95	100	100	95	100	100
Moyen	85	99	97	87	99	97	86	98	99
Minimal	77	98	94	76	97	95	76	90	98

TABLEAU A.42 – Taux de classification correcte utilisant les fréquences transformées en intervalles pour tous les scénarios d'endommagement (valeurs en %).

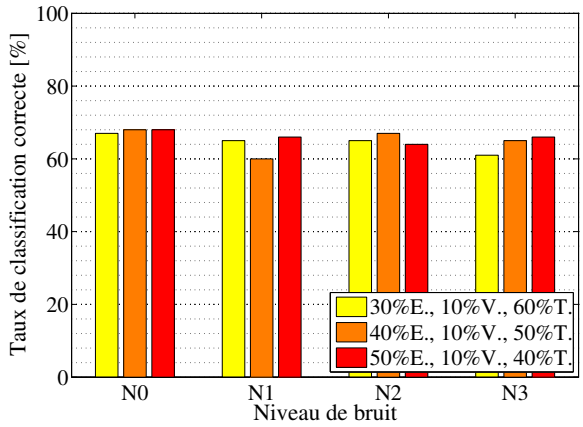
Finalement, les modes propres sont utilisés. En général, les méthodes permettent d'obtenir des pourcentages de l'ordre de 80% environ (cf. Tableau A.43). Il est noté que les résultats sont meilleurs que ceux obtenus par les signaux, mais les taux sont relativement plus faibles que ceux obtenus en considérant les fréquences propres.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	89	92	90	90	88	93	92	91	95
Moyen	80	78	81	79	80	81	82	85	83
Minimal	66	45	50	67	56	48	70	78	71

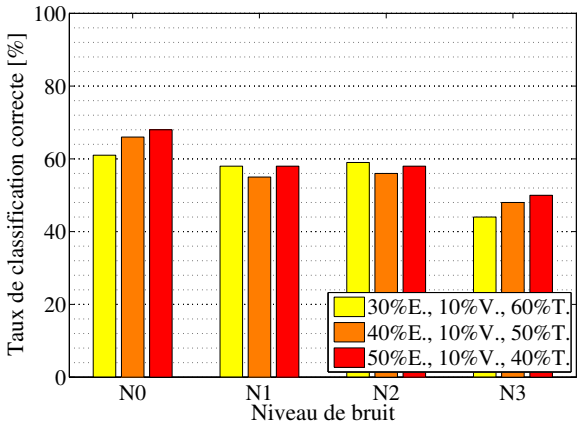
TABLEAU A.43 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour tous les scénarios d'endommagement (valeurs en %).

A.3.3 Bilan des résultats

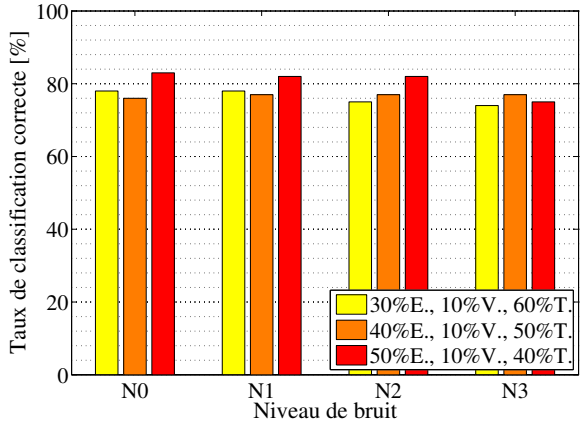
Dans ce paragraphe, un bilan des résultats obtenus par les méthodes de clustering est présenté. Les Figures A.21, A.22 et A.23 regroupent les pourcentages obtenus pour chaque méthode de partitionnement lorsque les signaux, fréquences et modes propres sont utilisés et convertis soit en intervalles interquartiles soit en histogrammes. Pour chacune de ces figures, les taux de classification sont regroupés par niveau de bruit. De cette façon, chaque barre verticale correspond à un niveau d'endommagement différent. La discussion détaillée des résultats est présentée dans le chapitre 4.



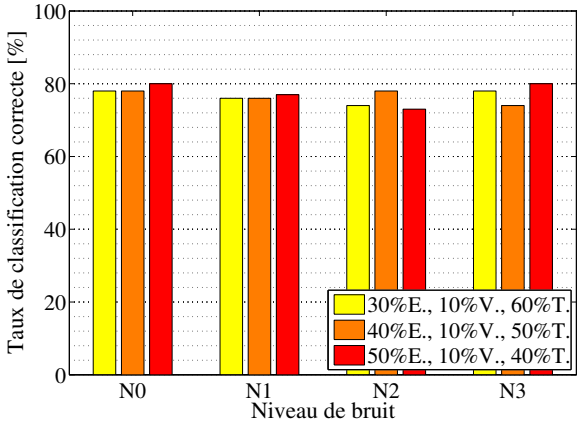
(a) ADB appliqués aux signaux transformés en intervalles interquartiles



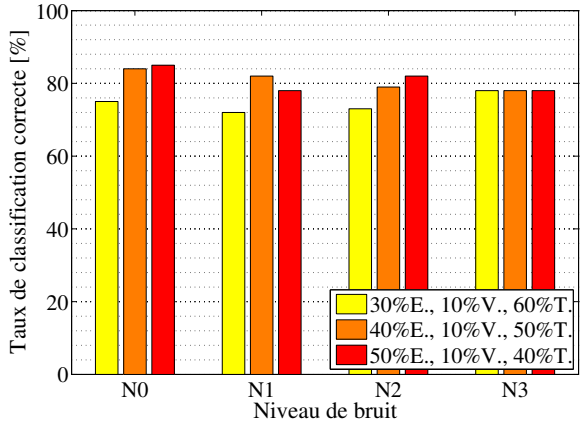
(b) ADB appliqués aux signaux transformés en histogrammes



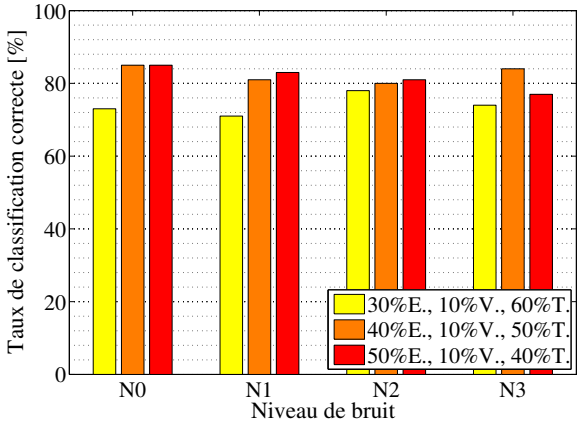
(c) RN appliqués aux signaux transformés en intervalles interquartiles



(d) RN appliqués aux signaux transformés en histogrammes

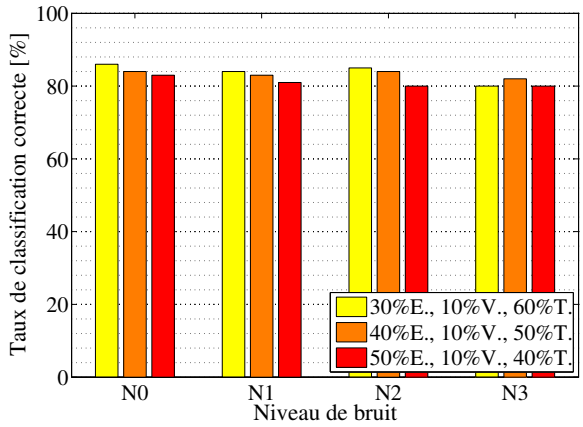


(e) SVM appliquées aux signaux transformés en intervalles interquartiles

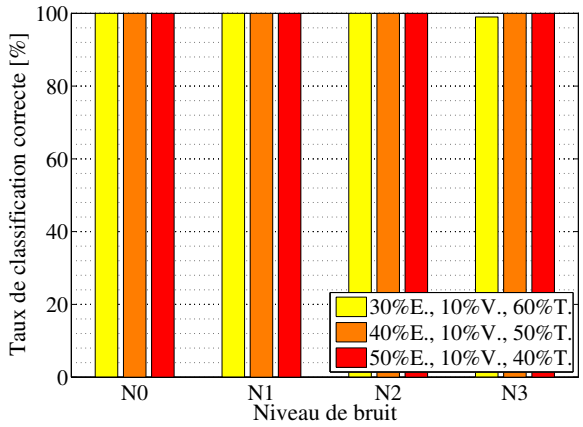


(f) SVM appliquées aux signaux transformés en histogrammes

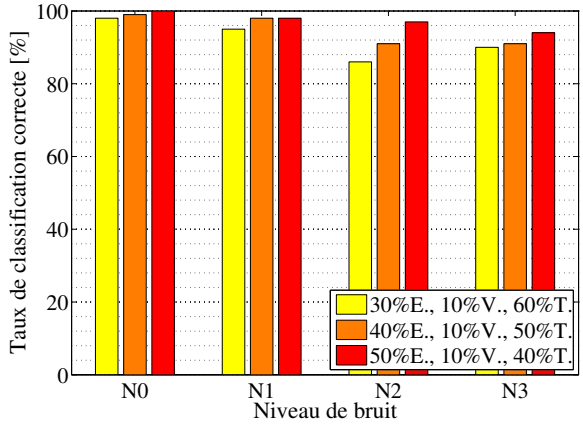
FIGURE A.21 – Taux de classification correcte pour les trois méthodes de classification supervisée utilisant les signaux.



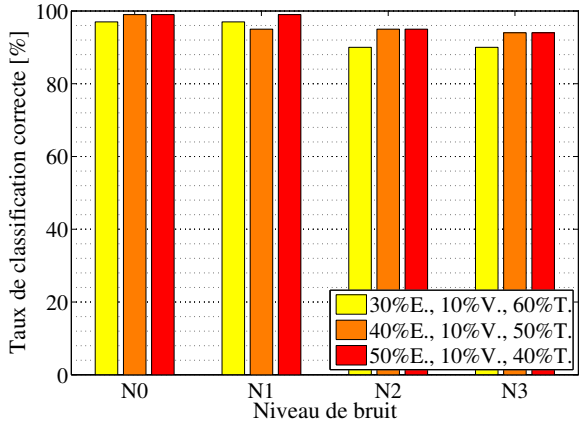
(a) ADB appliqués aux fréquences propres transformés en intervalles interquartiles



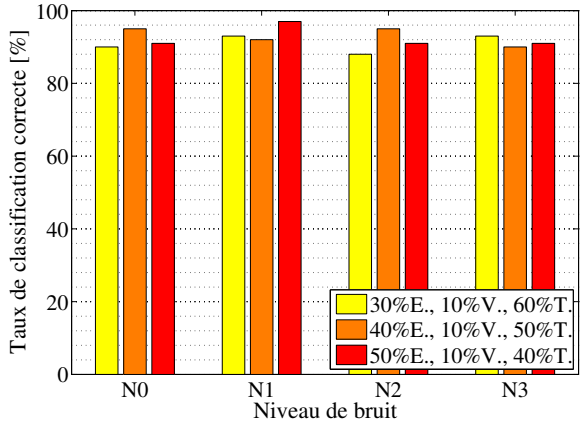
(b) ADB appliqués aux fréquences propres transformés en histogrammes



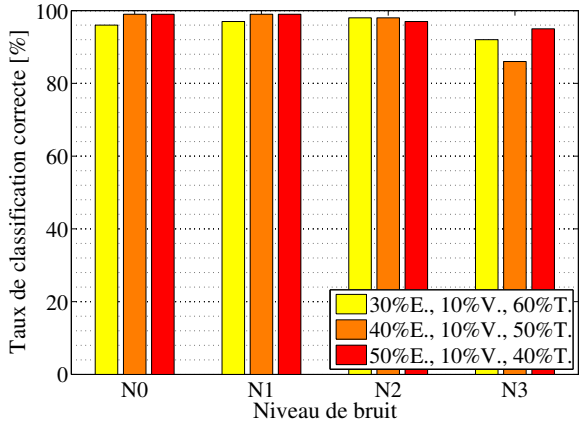
(c) RN appliqués aux fréquences propres transformés en intervalles interquartiles



(d) RN appliqués aux fréquences propres transformés en histogrammes

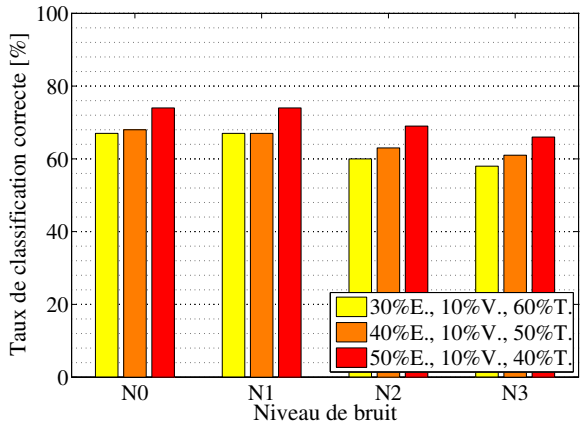


(e) SVM appliquées aux fréquences propres transformés en intervalles interquartiles

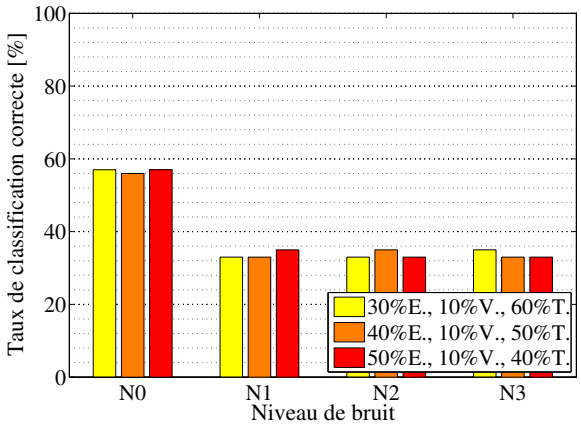


(f) SVM appliquées aux fréquences propres transformés en histogrammes

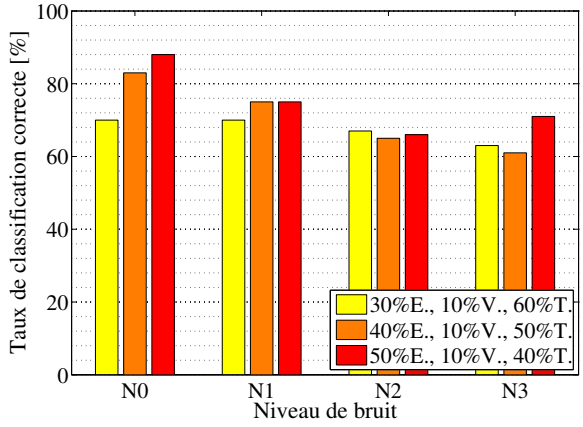
FIGURE A.22 – Taux de classification correcte pour les trois méthodes de classification supervisée utilisant les fréquences propres.



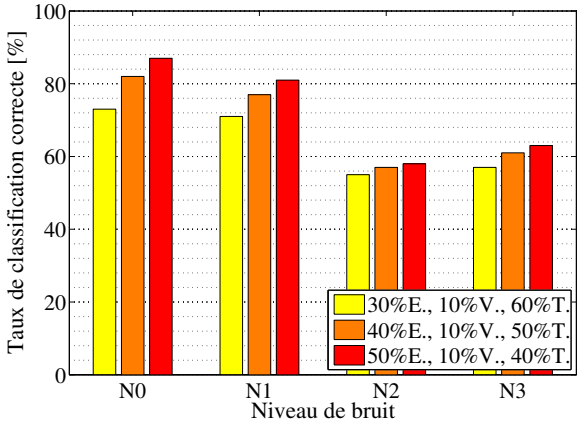
(a) ADB appliqués aux modes propres transformés en intervalles interquartiles



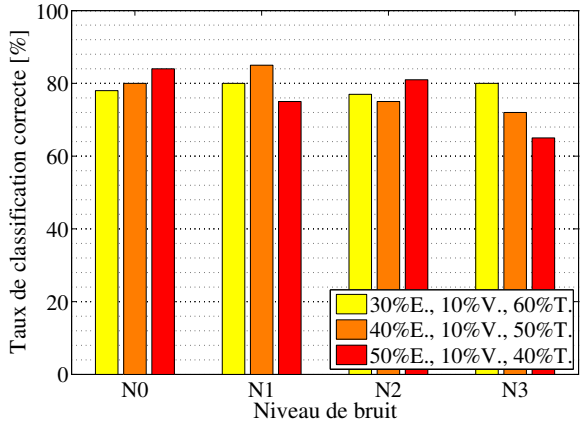
(b) ADB appliqués aux modes propres transformés en histogrammes



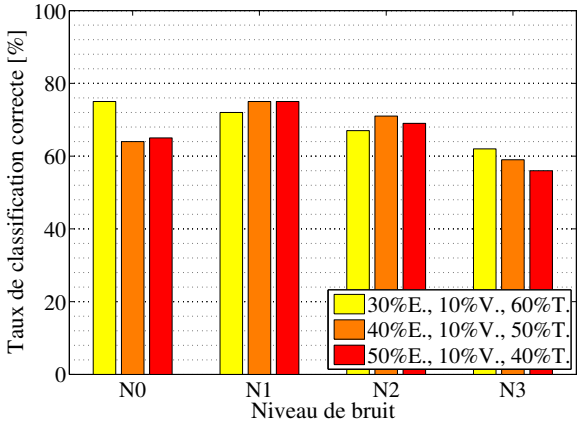
(c) RN appliqués aux modes propres transformés en intervalles interquartiles



(d) RN appliqués aux modes propres transformés en histogrammes



(e) SVM appliquées aux modes propres transformés en intervalles interquartiles



(f) SVM appliquées aux modes propres transformés en histogrammes

FIGURE A.23 – Taux de classification correcte pour les trois méthodes de partitionnement utilisant les modes propres.

Annexe B

Résultats complémentaires - Pont PK 075 + 317

Dans les paragraphes suivants, les résultats complémentaires obtenus par les approches proposées dans le cadre de cette thèse sont présentés. Ici, ces approches sont appliquées aux essais enregistrés à partir des campagnes de mesures réalisées sur le Pont PK 075 + 317, entre les villes de Sens et Soucy.

B.1 Identification Modale

Le Tableau B.1 regroupe les heures d'enregistrement des essais, les valeurs moyennes (calculées à partir des 10 réalisations) et les écarts-types respectifs des fréquences propres identifiées pour chaque essai enregistré avant resserrage.

Essai	Heure	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4	
		$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
TGV1A	10:57	5,88	0,11	8,77	0,08	12,93	0,09	16,86	0,09
TGV2A	11:11	5,79	0,01	8,08	0,03	13,30	0,02	16,22	0,02
TGV3A	11:17	5,82	0,01	8,40	0,13	13,39	0,08	16,42	0,09
TGV4A	11:23	5,80	0,02	8,22	0,02	13,49	0,02	17,23	0,02
TGV5A	11:35	5,92	0,02	8,86	0,15	13,06	0,10	16,99	0,11
TGV6A	12:28	5,93	0,02	9,02	0,05	13,14	0,04	16,99	0,04
TGV7A	12:36	5,89	0,01	8,81	0,08	13,23	0,05	16,99	0,06
TGV8A	13:39	6,03	0,02	9,08	0,16	12,90	0,11	16,92	0,12
TGV9A	13:49	5,88	0,01	8,77	0,03	13,03	0,02	16,85	0,02
TGV10A	14:01	5,90	0,01	8,61	0,04	12,68	0,03	16,65	0,03
TGV11A	14:16	5,91	0,02	8,83	0,05	13,06	0,04	16,92	0,04
TGV12A	14:26	5,94	0,01	8,96	0	12,73	0	17,07	0
TGV13A	14:37	5,90	0,01	8,75	0,08	13,07	0,05	16,80	0,06
TGV14A	14:47	5,94	0,03	9,34	0,14	13,46	0,10	16,86	0,10
TGV15A	14:57	5,89	0	8,79	0,01	13,01	0,07	17,13	0,01

TABLEAU B.1 – Fréquences propres identifiées lors de la première campagne d'essais (avant resserrage). Valeurs en Hz.

Le Tableau B.2 regroupe les heures d'enregistrement des essais, les valeurs moyennes (calculées à partir des 10 réalisations) et les écarts-types respectifs des fréquences propres identifiées pour chaque essai enregistré pendant resserrage.

	Essai	Heure	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4	
			$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
Phase 1	TGV1R	10:47	5,82	0,25	9,29	0,64	13,02	0,29	16,83	0,14
	TGV2R	11:04	5,93	0,15	9,15	1,28	12,94	0,29	17,07	0,06
	TGV3R	11:12	5,98	0,12	8,11	0,62	13,67	0,75	19,71	1,70
Phase 2	TGV4R	14:01	6,35	0,20	8,69	0,21	13,09	0,64	17,23	0,10
	TGV5R	14:16	6,20	0,11	8,37	0,79	14,27	0,42	17,59	0,65
	TGV6R	14:31	6,30	0,10	9,98	2,19	12,91	0,83	16,94	0,44
Phase 3	TGV7R	15:57	6,42	0,20	9,40	1,13	12,85	0,90	17,14	0,14
	TGV8R	16:01	6,35	0,12	8,93	0,87	12,93	0,67	16,78	0,25
	TGV9R	16:16	6,19	0,28	8,17	0,79	14,02	1,97	19,72	0,86
Phase 4	TGV10R	16:27	6,51	0,11	8,94	0,97	13,63	0,49	17,42	0,12
	TGV11R	17:37	6,49	0,15	8,88	1,06	13,51	0,67	17,29	0,25
	TGV12R	17:46	6,36	0,18	8,41	0,87	13,70	0,36	17,42	1,83
	TGV13R	17:50	6,34	0,42	8,20	0,88	12,18	1,54	20,11	1,61

TABLEAU B.2 – Fréquences propres identifiées lors de la deuxième campagne d'essais (pendant resserrage). Valeurs en Hz.

Le Tableau B.3 regroupe les valeurs moyennes (calculées à partir des 10 réalisations) et les écarts-types respectifs des taux d'amortissement identifiés pour chaque phase de l'état pendant resserrage.

Phase	Mode 1		Mode 2		Mode 3		Mode 4	
	$\overline{\xi}$	σ_ξ	$\overline{\xi}$	σ_ξ	$\overline{\xi}$	σ_ξ	$\overline{\xi}$	σ_ξ
1	5,3	3,5	1,6	2,4	1,1	1,7	0,9	3,1
2	4,2	3,1	0,5	1,3	1,6	3,4	3,0	1,5
3	4,3	2,6	1,0	2,0	3,3	4,9	2,3	2,9
4	4,2	2,3	1,5	4,1	3,3	3,5	2,5	7,6
Moyenne	**4,5**	2,9	**1,1**	2,4	**2,3**	3,3	**2,2**	3,8

TABLEAU B.3 – Valeurs moyennes et écarts-types des coefficients d'amortissement (pendant resserrage). Valeurs en (%).

Le Tableau B.4 regroupe les heures d'enregistrement des essais, les valeurs moyennes (calculées à partir des 10 réalisations) et les respectifs écarts-types des fréquences propres identifiées pour chaque essai enregistré après resserrage.

B.2 Analyse classique avec prise en compte des effets thermiques

Dans ce paragraphe, une étude des effets de la variation des températures sur les paramètres modaux est menée. La démarche proposée prend en compte les techniques de régression linéaire et non linéaire présentées dans le paragraphe 3.4. La Figure B.1 illustre les histogrammes pour les quatre premières fréquences propres des états avant et après resserrage, corrigées par le filtre non linéaire.

Le Tableau B.5 regroupe les résultats obtenus avec des tests d'hypothèse en considérant les quatre fréquences propres corrigées par le filtre non linéaire. Il est noté que les tests échouent pour

Essai	Heure	Fréquence 1		Fréquence 2		Fréquence 3		Fréquence 4	
		$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f	$\overline{f}$	σ_f
TGV1B	11:27	6,30	0,16	8,32	0,37	14,28	1,66	17,97	0,81
TGV2B	11:43	6,46	0,18	8,92	0,44	13,10	0,09	17,68	0,82
TGV3B	12:34	6,52	0,25	8,98	0,45	13,03	0,11	17,36	0,69
TGV4B	13:03	6,48	0,16	9,03	0,39	13,05	0,12	17,80	0,53
TGV5B	13:37	6,55	0,12	8,79	0,54	12,98	0,18	17,52	0,99
TGV6B	13:50	6,63	0,05	8,80	0,20	13,22	0,17	17,37	0,18
TGV7B	14:19	6,46	0,20	8,82	0,54	12,93	0,13	17,05	0,69
TGV8B	14:31	6,47	0,10	9,09	0,43	13,14	0,09	17,54	0,63
TGV9B	14:46	6,48	0,16	8,87	0,42	13,15	0,30	17,46	0,66
TGV10B	14:58	6,47	0,23	8,81	0,39	12,91	0,37	17,19	0,72
TGV11B	15:12	6,44	0,21	8,84	0,47	12,89	0,39	17,57	1,21
TGV12B	15:25	6,43	0,12	8,94	0,42	13,00	0,11	17,29	0,24
TGV13B	15:38	6,48	0,10	9,05	0,37	13,01	0,12	17,25	0,43

TABLEAU B.4 – Fréquences propres identifiées lors de la troisième campagne d'essais (après resserrage). Valeurs en Hz.

les trois premières fréquences propres, ce qui montre que leurs distributions sont effectivement différentes avant et après resserrage.

$1^{ère}$ fréq.		$2^{ème}$ fréq.		$3^{ème}$ fréq.		$4^{ème}$ fréq.	
H	p^* (%)	H	p^* (%)	H	p^* (%)	H	p^* (%)
1	0	1	1,05	1	1,63	0	22,8

TABLEAU B.5 – Résultats des tests d'hypothèse de Kolmogorov-Smirnov pour les quatre premières fréquences propres (états avant et après ; régression non linéaire).

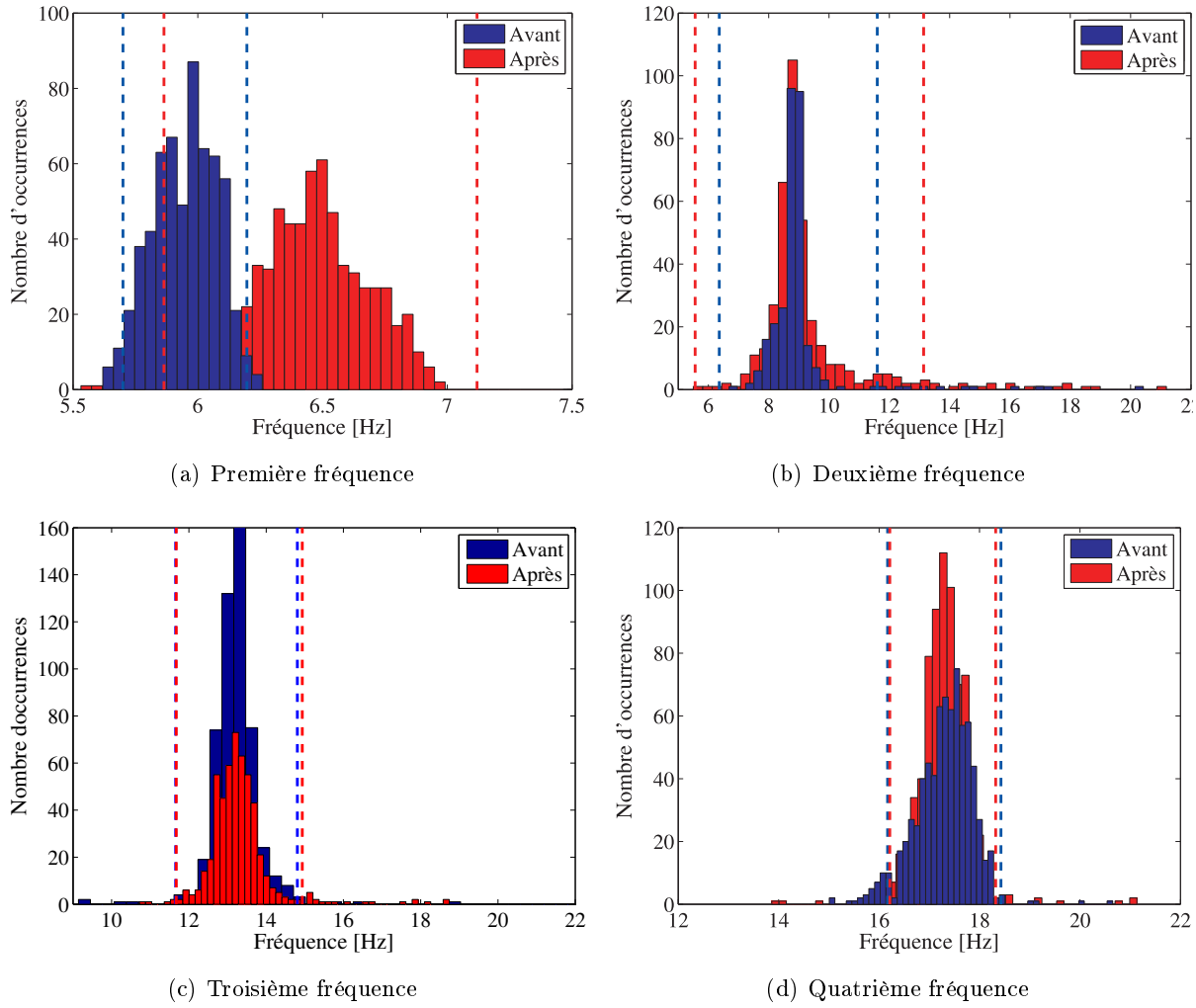
La Figure B.2 illustre les quatre premières déformées modales corrigées par le filtre non linéaire et les coefficients MAC associés.

B.3 ADS couplée aux méthodes de clustering

Dans ce paragraphe, les résultats obtenus à partir des méthodes de clustering appliquées aux signaux et aux paramètres modaux sont présentés. Dans cette annexe, la transformation en intervalles interquartiles est considérée. Toutes les simulations de clustering décrites dans les paragraphes suivants considèrent les quatre premières fréquences propres et déformées modales associées.

B.3.1 Signaux

Les premiers résultats correspondent à l'application de la méthode des divisions hiérarchiques aux signaux. La Figure B.3 montre la classification obtenue en utilisant les signaux transformés en intervalles interquartiles. En général, les taux de classification correcte sont pertinents. Il est observé que les essais de l'état pendant resserrage tendent à se mélanger relativement plus que les essais des autres états. Par exemple, le premier cluster contient pratiquement tous les essais « avant », mais il existe aussi deux essais (TGV10R et TGV11R) qui, théoriquement, ne devraient pas y être. Toutefois, il est important de remarquer qu'aucun essai « avant » se retrouve dans le cluster d'essais « après », et vice-versa. Cette observation est importante, car elle montre



(a) Première fréquence

(b) Deuxième fréquence

(c) Troisième fréquence

(d) Quatrième fréquence

FIGURE B.1 – Comparatif des histogrammes pour les quatre premières fréquences propres des états avant et après resserrage (régression non linéaire).

que l'approche proposée est suffisamment sensible pour détecter les effets du resserrage, ce qui n'apparaissait pas clairement en utilisant l'analyse classique. Dans cet arbre, les variables les plus discriminantes sont les capteurs 10 (situé proche de l'appui) et 7 (situé à mi-travée) pour les première et seconde divisions, respectivement. Par ailleurs, 12 essais sur 15 (80%) de l'état avant sont correctement classés. Pour les états pendant et après, les taux sont de 70% (9/13) et 85% (11/13), respectivement.

La méthode basée sur les agglomérations hiérarchiques est maintenant considérée. Les trois clusters obtenus sont identifiés à la Figure B.4. Il est noté qu'un nombre plus important d'essais a été classé incorrectement. Bien que la plupart des essais « avant » soient classés dans un cluster séparé, il y en a d'autres qui se mélangent dans le cluster de l'état « après ». Le même résultat est observé pour les essais « pendant ». Comme il a été noté dans le chapitre 4, le critère de distance utilisé dans cette méthode peut ne pas être suffisamment sensible pour discriminer les états structuraux. Par conséquent, seuls 9 essais sur 15 (60%) de l'état « avant » sont classés correctement, tandis que 7 sur 13 (54%) pour l'état « pendant » et 8 sur 13 (62%) pour « après ».

Enfin, les nuées dynamiques sont appliquées aux signaux. Il est observé cette méthode permet d'obtenir des meilleurs résultats que ceux obtenus par les divisions hiérarchiques et les agglo-

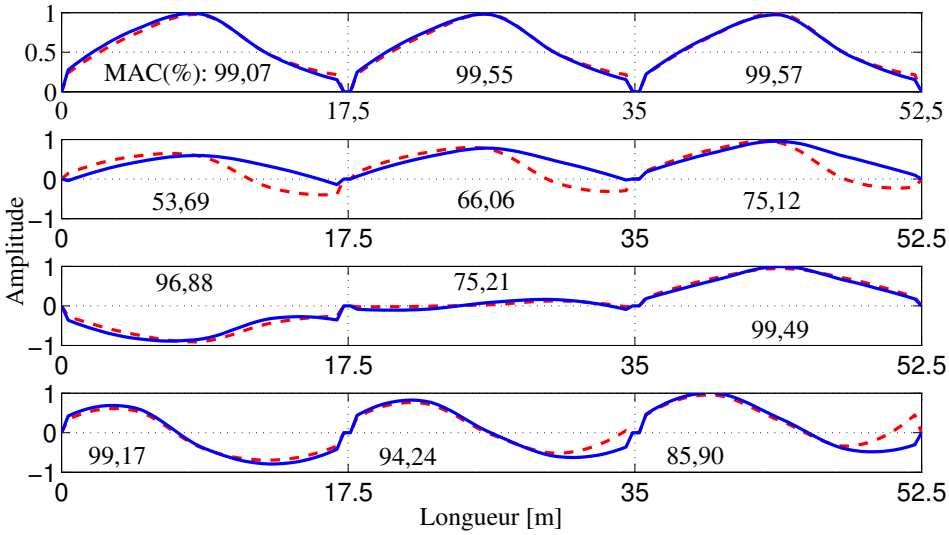


FIGURE B.2 – Comparatif des quatre premiers modes propres avant et après resserrage et calcul de MAC (régression non linéaire).

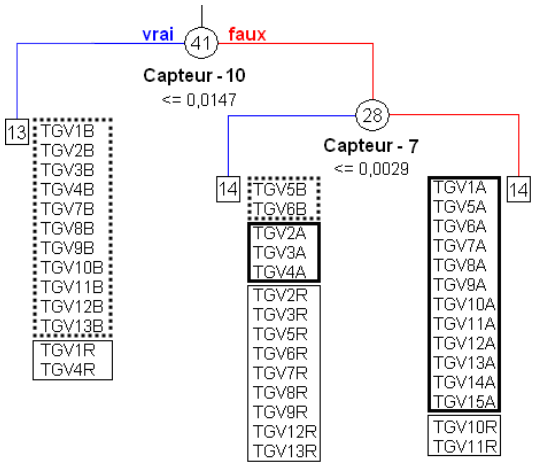


FIGURE B.3 – Classification des signaux en utilisant les divisions hiérarchiques (intervalles).

mérations hiérarchiques. Les taux de classification correcte sont de 73% (11/15) pour le cluster « avant », 85% (11/13) pour l'état « pendant » et 77% (10/13) pour l'état « après ». Pour cette application, il est possible de dire que cette méthode est la plus pertinente.

B.3.2 Fréquences propres

B.3.2.1 Sans prise en compte des effets thermiques

Dans ce paragraphe, les méthodes de partitionnement sont appliquées aux quatre premières fréquences propres transformées en données symboliques. La Figure B.5 illustre l'arbre de division construit et les clusters obtenus pour chacun des états. Comme il a été remarqué précédemment, bien que le cluster « pendant » contienne tous les types d'essais, le cluster « avant » ne contient que des essais avant resserrage. En effet, la classification obtenue en utilisant les fréquences est significativement meilleure que celle avec les signaux. Dans cette analyse, 87% (13/15) des essais

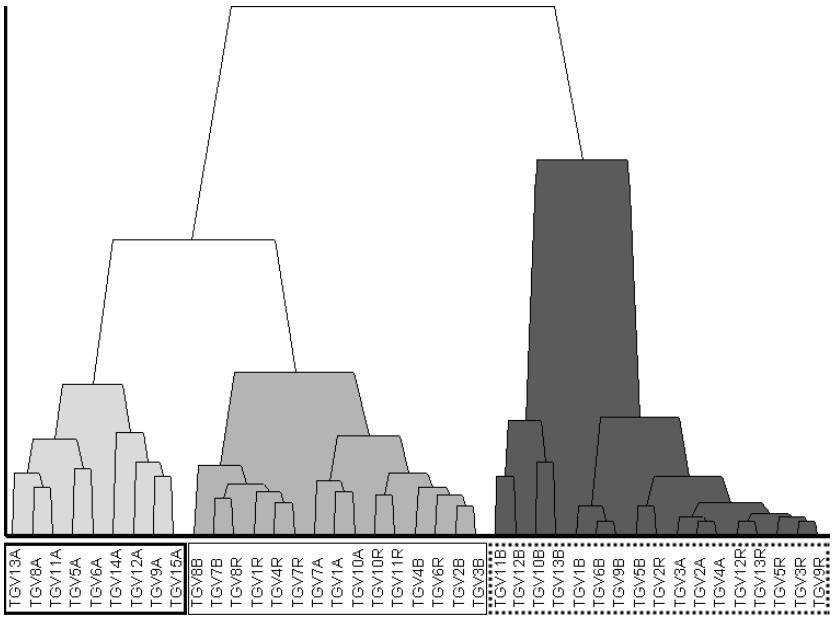


FIGURE B.4 – Classification des signaux en utilisant les agglomérations hiérarchiques (intervalles).

sont correctement classés comme « avant », 85% (11/13) comme « pendant », et 92%(12/13) comme « après ».

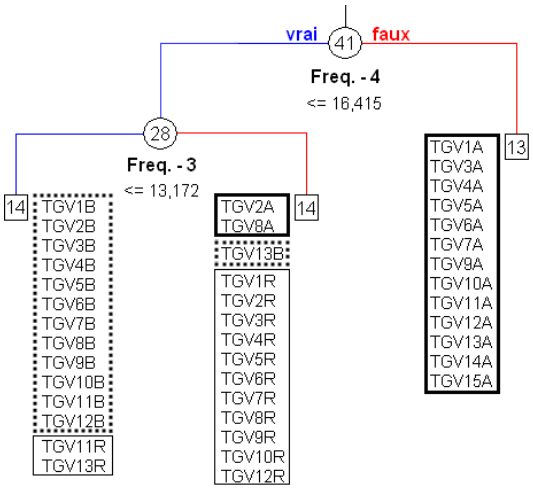


FIGURE B.5 – Classification des fréquences propres en utilisant les divisions hiérarchiques (intervalles).

Par ailleurs, les agglomérations hiérarchiques sont considérées. La Figure B.6 montre la hiérarchie construite à partir de la transformation des fréquences propres en intervalles. Les taux de classification sont relativement moins importants que ceux obtenus avec la méthode des divisions hiérarchiques : 11 essais sur 15 (73%) sont classés correctement comme « avant », 8 sur 13 (62%) pour l'état « pendant » et 12 sur 13 (92%) pour l'état « après ». Il est néanmoins surprenant d'observer que le cluster relatif à l'état « après » se trouve à coté du groupe « avant ». En effet, si la représentation en intervalles interquartiles n'est pas la plus pertinente pour cette méthode, il est possible que les calculs de distance entre les essais soient affectés et donc, que l'assemblage des clusters ne soit pas optimal.

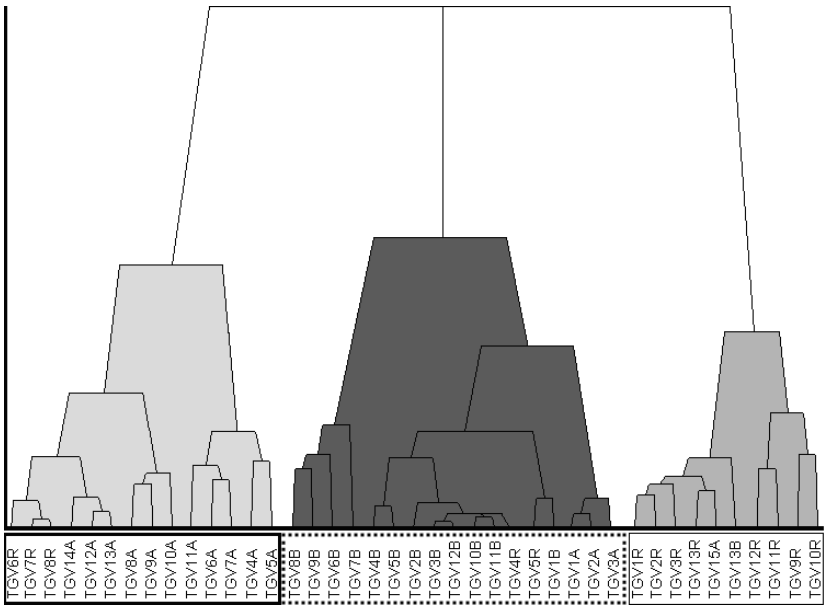


FIGURE B.6 – Classification des fréquences propres en utilisant les agglomérations hiérarchiques (intervalles).

Les résultats obtenus en utilisant les nuées dynamiques sont pertinents. Dans le cas de la transformation en intervalles interquartiles, pratiquement tous les essais de l'état « avant » et « après » sont classés correctement (93% et 92%, respectivement). Pour l'état « pendant », le taux est égal à 77%.

B.3.2.2 Avec prise en compte des effets thermiques

Dans un premier temps, les fréquences propres corrigées par le filtre linéaire transformées en intervalles interquartiles sont considérer pour appliquer la méthode des divisions hiérarchiques. La Figure B.7 montre les clusters obtenus pour chacun des états. Le partitionnement obtenu ne présente pas de différence significative avec que celui évalué sans la correction thermique (cf. Figure B.5). Les taux de classification correcte sont relativement similaires : 87% (13/15) d'essais sont correctement classés comme « avant », 85% (11/13) comme « pendant », et 100%(13/13) comme « après ».

En considérant la méthode des agglomérations hiérarchiques, il est noté que l'utilisation du filtre ne change pas significativement les taux de classification correcte obtenus : dans le cas des intervalles, 80% (12/15) sont classés dans le cluster « avant », 70% (9/13) dans « pendant » et 85%(11/13) dans « après » (cf. Figure B.8). Il existe quelques essais qui sont regroupés différemment pour chaque type de transformation symbolique, mais en général, il est observé une quantité élevée d'essais mélangés.

La méthode des nuées dynamiques est aussi considérée. Les résultats obtenus confirment les remarques précédentes. Le taux de classification correcte est de 87% (13/15) pour le cluster « avant », 70% (9/13) pour l'état « pendant » et 85% (11/13) pour l'état « après ».

Par la suite, des filtres non linéaires (basés sur des réseaux de neurones) sont utilisés pour la correction des fréquences propres. L'objectif de ces simulations est de seulement non corriger les effets liés à la variation de températures, mais aussi de comparer les résultats avec ceux obtenus en utilisant le filtre linéaire.

En premier lieu, les divisions hiérarchiques sont appliquées aux fréquences propres. L'arbre

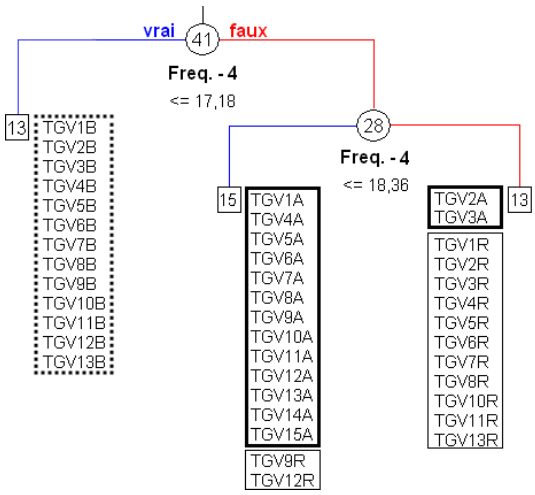


FIGURE B.7 – Classification des fréquences propres en utilisant les divisions hiérarchiques (intervalles ; régression linéaire).

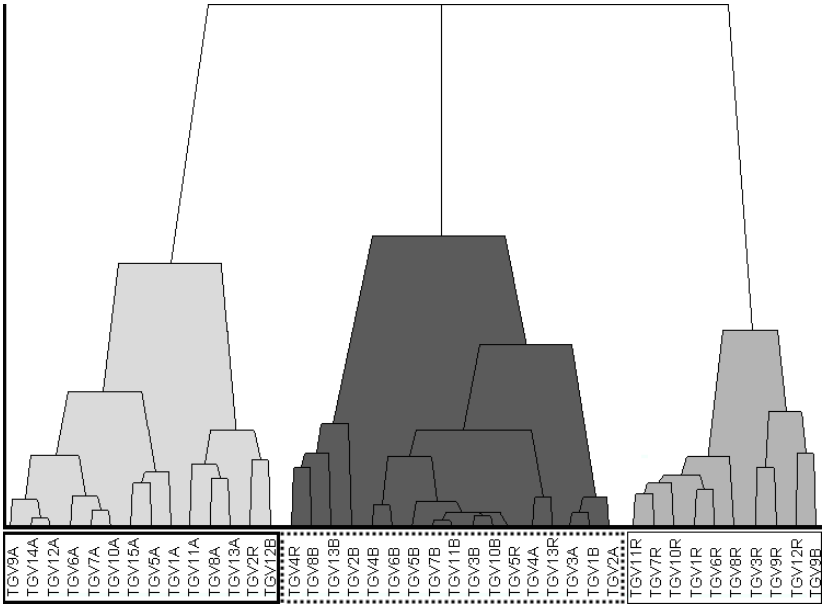


FIGURE B.8 – Méthode des agglomérations hiérarchiques appliquée aux fréquences propres corrigées (intervalles ; régression linéaire).

de décision est montré à la Figure B.9. En général, les résultats sont moins pertinents que ceux obtenus en utilisant le filtre linéaire (cf. Figure B.7). Cela permet d'indiquer qu'une correction des effets thermiques sur les mesures peut ne pas être pertinente dans cette application.

Dans le cas des agglomérations hiérarchiques, les taux de classification correcte sont relativement faibles. En considérant la transformation en intervalles interquartiles, 80% des essais sont correctement classés dans le cluster « avant », 85% dans le cluster « pendant » et seulement 46% dans le cluster « après » (Figure B.10).

Pour la méthode des nuées dynamiques, les taux de classification sont plus importants que ceux obtenus avec le filtre linéaire. Le taux de classification correcte est de 93% (14/15) pour le cluster « avant », 77% (10/13) pour l'état « pendant » et 85% (11/13) pour l'état « après ».

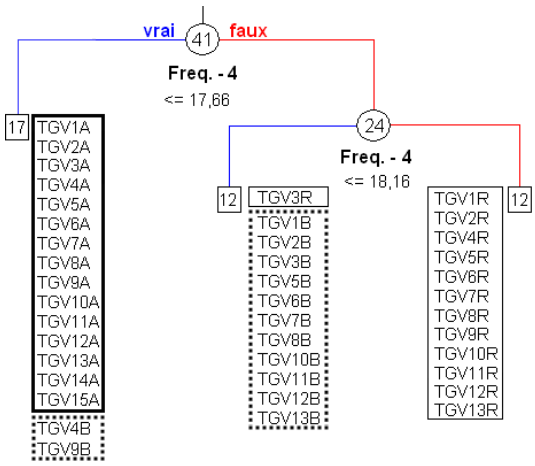


FIGURE B.9 – Classification des fréquences propres en utilisant les divisions hiérarchiques (intervalles ; régression non linéaire).

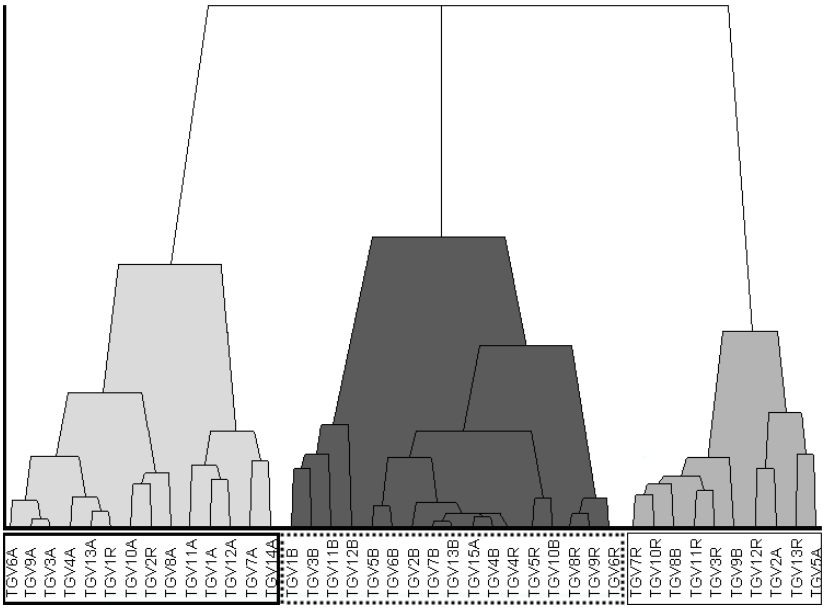


FIGURE B.10 – Méthode des agglomérations hiérarchiques appliquée aux fréquences propres corrigées (intervalles ; régression non linéaire).

B.3.3 Modes propres

B.3.3.1 Sans prise en compte des effets thermiques

Les quatre premiers modes propres sont transformés en données symboliques et utilisés pour appliquer la méthode des divisions hiérarchiques. L'arbre de décision construit à partir de la représentation en intervalles interquartiles est illustré à la Figure B.11. Les résultats ne sont pas aussi pertinents que ceux obtenus en utilisant les fréquences propres. En effet, il est noté que des essais « avant » et « après » se mélangent dans tous les groupes, ce qui montre la faible sensibilité des modes propres par rapport aux effets des travaux réalisés. Dans ce cas, 10 essais sur 15 (67%) se trouvent dans le cluster « avant », 8 sur 13 (62%) dans le groupe « pendant » et 9 sur 13 (70%) dans le cluster de l'état « après ».

La méthode des agglomérations hiérarchiques est également utilisée en considérant la trans-

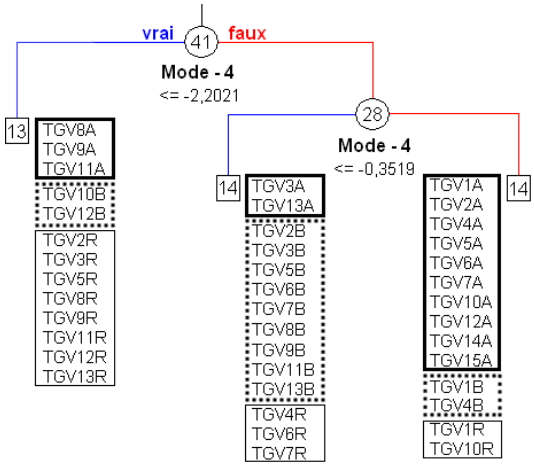


FIGURE B.11 – Classification des modes propres en utilisant les divisions hiérarchiques (intervalles).

formation en intervalles interquartiles. Il est noté (comme pour les signaux) qu'un grand nombre d'essais est mélangé (Figure B.12). Bien que les clusters des états « avant » et « après » ne regroupent pratiquement que leurs essais respectifs, le cluster « pendant » est très hétérogène, en regroupant tous les types d'essais. Les taux de classification correcte sont relativement faibles : 47% (7/15) pour l'état « avant », 62% (8/13) pour « pendant » et 54% (7/13) pour l'état « après ». Il est noté que les dispositions des clusters sont inversées (avant → après → pendant), comme dans l'analyse avec les fréquences propres. Cela met en évidence que cette méthode peut ne pas être la plus pertinente pour cette étude.

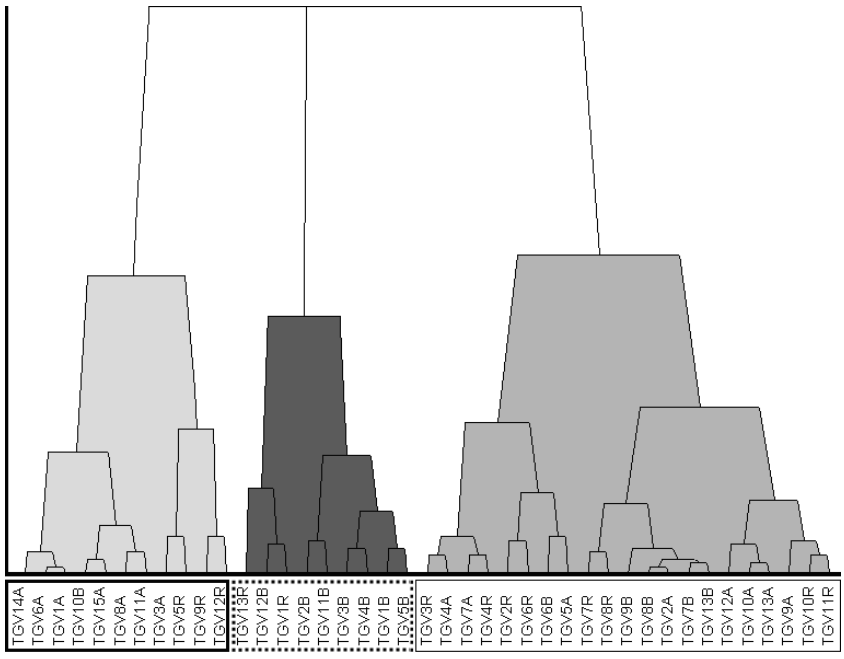


FIGURE B.12 – Classification des modes propres en utilisant les agglomérations hiérarchiques (intervalles).

La méthode des nuées dynamiques est également utilisée. Le taux de classification correcte est de 60% (9/15) pour le cluster « avant », 62% (8/13) pour l'état « pendant » et 62% (8/13) pour l'état « après ».

B.3.3.2 Avec prise en compte des effets thermiques

La démarche proposée est maintenant appliquée aux modes propres à partir de la correction de chacune des composantes modales. Les divisions hiérarchiques sont d'abord appliquées aux intervalles interquartiles (Figure B.13). Bien que l'arbre de décision soit relativement différent de celui construit sans prise en compte des effets thermiques (cf. Figure B.11), les résultats sont exactement les mêmes : 10 essais sur 15 (67%) se retrouvent dans le cluster « avant », 8 sur 13 (62%) dans le groupe « pendant » et 9 sur 13 (70%) dans le cluster de l'état « après ».

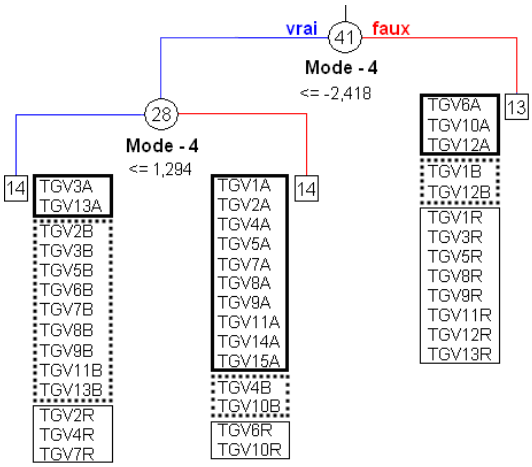


FIGURE B.13 – Classification des modes propres en utilisant les divisions hiérarchiques (intervalles ; régression linéaire).

La méthode des agglomérations hiérarchiques est maintenant considérée en considérant la transformation en intervalles interquartiles. Comme pour les signaux, un grand nombre d'essais sont mélangés (Figure B.14). Les taux de classification correcte sont également faibles : 60% (9/15) pour l'état « avant », 62% (8/13) pour « pendant » et 46% (6/13) pour l'état « après ».

Par ailleurs, la méthode des nuées dynamiques est utilisée. Dans le cas de la transformation en intervalles interquartiles, le taux de classification correcte est égal à 67% (10/15) pour le cluster « avant », 62% (8/13) pour l'état « pendant » et 54% (7/13) pour l'état « après ». En général, les taux de classification évalués ici sont significativement plus faibles que ceux obtenus sans l'approche de correction thermique.

Les simulations suivantes considèrent des filtres non linéaires (basés sur des réseaux de neurones) pour la correction des modes propres. La Figure B.15 illustre l'arbre de décision construit en considérant des intervalles interquartiles. Il est noté une sensible amélioration (par rapport au filtre linéaire) dans la classification des essais, surtout pour les états « avant » (87%) et « après » (85%) resserrage. Toutefois, ces résultats ne permettent pas de conclure si le filtre correction est cohérent, étant donné que les taux de classification ne sont pas significativement différents de ceux déjà obtenus (sans correction thermique).

Dans le cas de la méthode des agglomérations hiérarchiques, les taux de classification correcte sont relativement faibles. Notamment dans le cas des intervalles, un peu plus de la moitié de essais « avant » et « après » est classée correctement (Figure B.16). Dans l'ensemble des résultats, les agglomérations obtenues ne changent pratiquement pas quel que soit le type de filtre utilisé.

Les taux obtenus avec la méthode des nuées dynamiques sont relativement plus élevés que ceux obtenus par le filtre linéaire. Le taux de classification correcte est en effet de 73% (11/15) pour le cluster « avant », 54% (7/13) pour l'état « pendant » et 77% (11/13) pour l'état « après ». Toutefois, il n'est pas observé une amélioration de la classification par rapport à celle obtenue

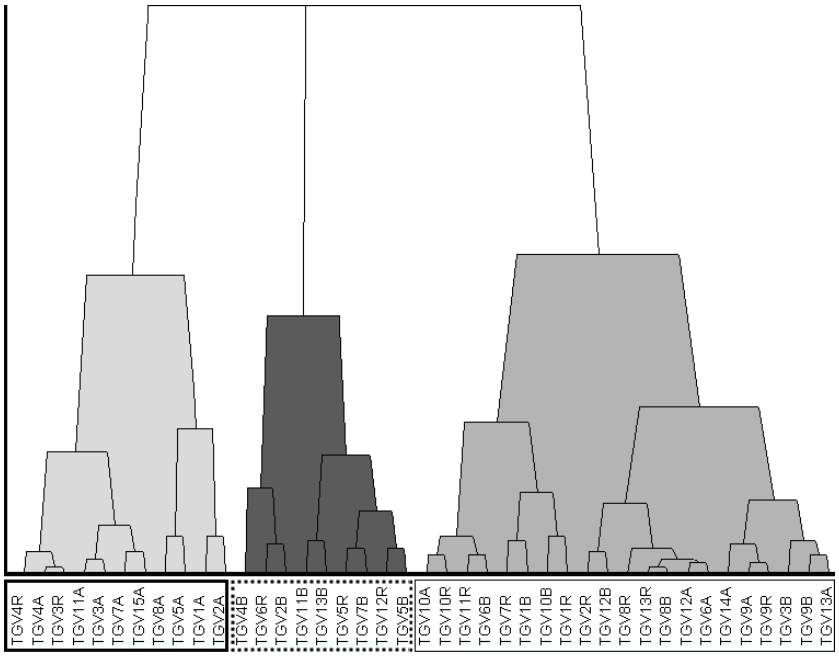


FIGURE B.14 – Méthode des agglomérations hiérarchiques appliquée aux modes propres corrigées (intervalles ; régression linéaire).

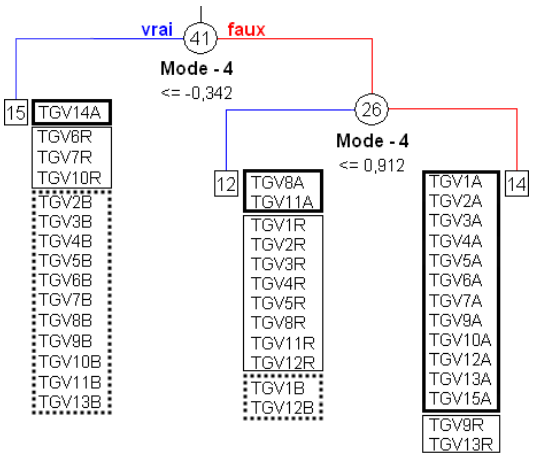


FIGURE B.15 – Classification des modes propres en utilisant les divisions hiérarchiques (intervalles ; régression non linéaire).

sans correction thermique.

B.3.4 Étude complémentaire

Une brève étude complémentaire est maintenant considérée. Les simulations de partitionnement réalisées dans les paragraphes précédents ont considéré les essais correspondant aux trois états structuraux du pont-rail. Étant donné que l'état « pendant resserrage » représente une condition structurale relativement floue, il se peut que cet état ne corresponde pas lui-même à une condition unique de l'ouvrage.

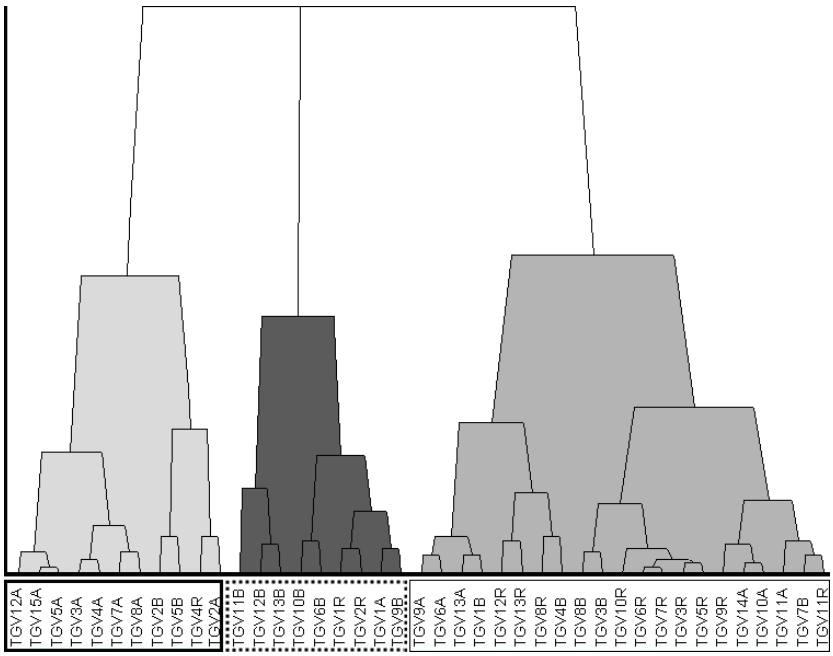


FIGURE B.16 – Méthode des agglomérations hiérarchiques appliquée aux modes propres corrigées (intervalles ; régression non linéaire).

B.3.4.1 Signaux

Dans un premier temps, la méthode des divisions hiérarchiques est appliquée aux signaux transformés en intervalles. La Figure B.17 présente l'arbre de décision obtenu. Il est noté que les taux de classification sont significativement meilleurs. Pour la transformation en intervalles interquartiles, tous les essais « avant » sont classés dans un cluster, tandis que 62% (8/13) le sont dans le cluster « après ».

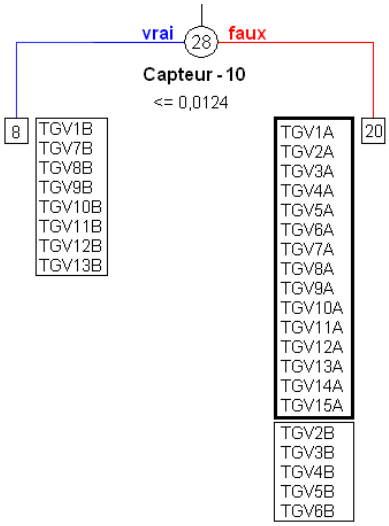


FIGURE B.17 – Classification des signaux en utilisant les divisions hiérarchiques (intervalles).

Dans le cas de la méthode des agglomérations hiérarchiques, 60% (9/15) des essais sont classés correctement à l'état « avant » et 54% (7/13) à l'état « après » (5.29). En effet, peu de différence est remarqué entre ces les hiérarchies obtenues par le type de transformation en

données symboliques. En général, il est observé une quantité élevée d'essais mélangés.

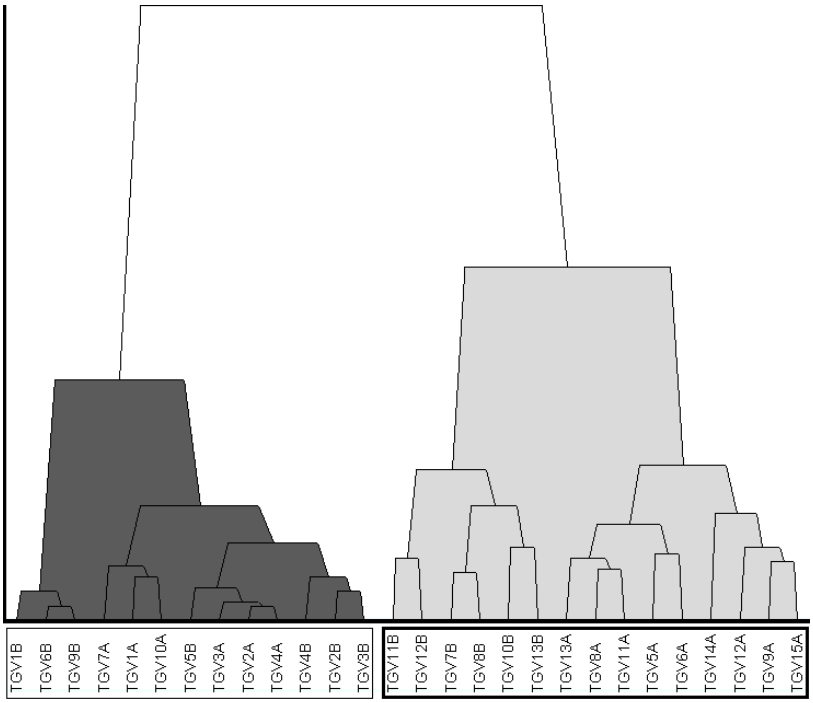


FIGURE B.18 – Classification des signaux en utilisant les agglomérations hiérarchiques (intervalles).

La méthode des nuées dynamiques permet d'obtenir les meilleurs résultats. En effet, tous les essais « avant » sont classés correctement. Pour l'état « après » resserrage, le taux de classification correcte est 69% (9/13).

B.3.4.2 Fréquences propres

Dans un deuxième temps, les représentations symboliques des quatre premières fréquences propres sont utilisées. Comme déjà observé dans les autres applications, les résultats sont également meilleurs que ceux obtenus avec les signaux. Si la méthode des divisions hiérarchiques est appliquée à des intervalles interquartiles, tous les essais « avant » sont classés correctement, tandis que 11 essais sur 13 (85%) le sont pour l'état « après » (Figure B.19).

La Figure B.20 illustre le partitionnement obtenu en utilisant des intervalles en considérant la méthode des agglomérations hiérarchiques. Dans ce cas, les taux de classification correcte sont de 80% (12/15) pour l'état « avant » et 46% (6/13) pour l'état « après ». Cela montre que cette méthode n'est pas capable de bien discriminer ces états structuraux, même si des fréquences propres sont utilisées.

Dans le cas des nuées dynamiques, les résultats sont meilleurs que ceux obtenus par les divisions hiérarchiques. Le taux de classification correcte est de 93% (14/15) pour le cluster « avant », et 92% (12/13) pour l'état « après ».

B.3.4.3 Modes propres

Les méthodes de clustering sont maintenant appliquées aux quatre premiers modes propres. Dans le cas de la transformation en intervalles interquartiles, 80% (12/15) des essais sont bien

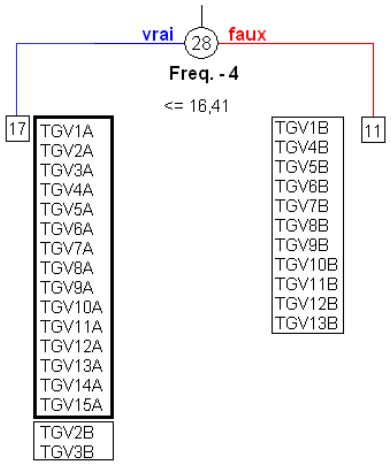


FIGURE B.19 – Classification des fréquences propres en utilisant les divisions hiérarchiques (intervalles).

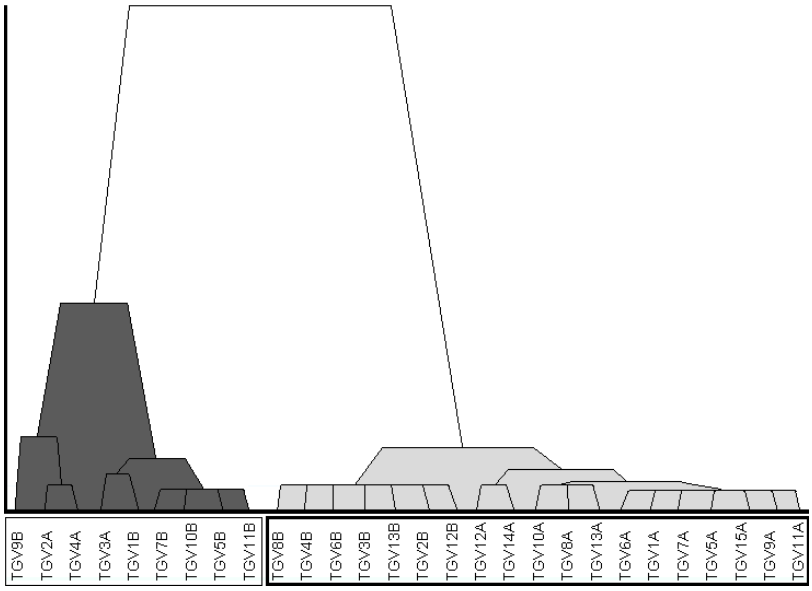


FIGURE B.20 – Classification des fréquences propres en utilisant les agglomérations hiérarchiques (intervalles).

classés dans le cluster « avant » et 100% dans le groupe « après » (cf. Figure B.21).

En utilisant les agglomérations hiérarchiques, bien que les taux de classifications soient meilleures que celux avec des signaux, un nombre significatif d'essais est mélangé. La Figure B.22 illustre la hiérarchie obtenue en considérant des intervalles interquartiles. Les taux de classification correcte sont de 67% (10/15) pour l'état « avant » et 60% (9/13) pour l'état « après ».

La méthode des nuées dynamiques est maintenant considérée. Cette méthode permet généralement d'obtenir les meilleurs résultats (comparés à ceux obtenus avec les autres méthodes). Dans cet exemple, 93% (14/15) des essais sont classés correctement dans le cluster « avant » et 77% (10/13) dans le cluster « après ».

Compte tenu des résultats, les méthodes des divisions hiérarchiques et des nuées dynamiques sont les méthodes les plus adaptées pour la discrimination d'états structuraux.

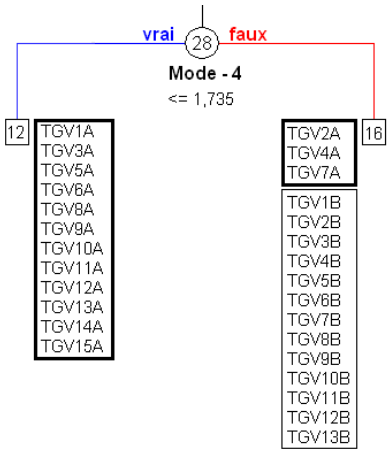


FIGURE B.21 – Classification des modes propres en utilisant les divisions hiérarchiques (intervalles).

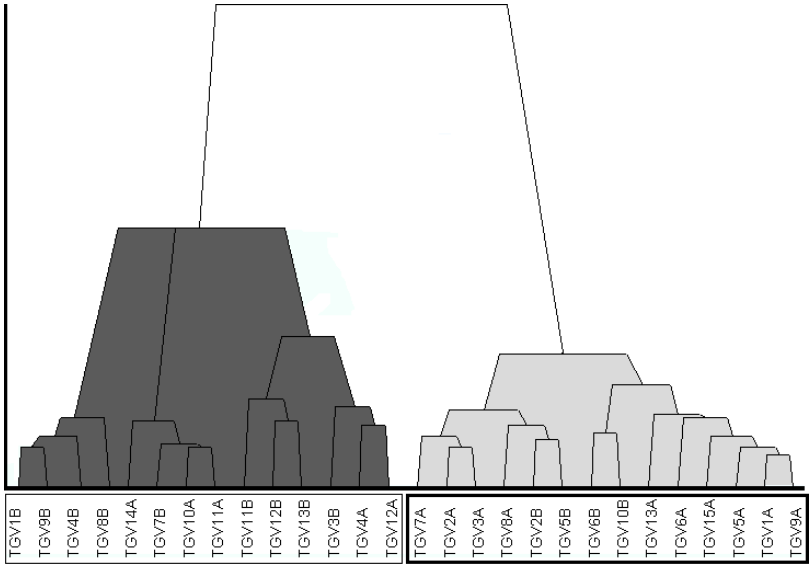


FIGURE B.22 – Classification des modes propres en utilisant les agglomérations hiérarchiques (intervalles).

B.3.4.4 Application aux paramètres modaux corrigés

Comme il a été proposé dans les paragraphes précédents, une étude complémentaire tenant compte uniquement des essais enregistrés dans les états « avant » et « après » resserrage est proposée.

Les premières analyses portent sur la méthode des divisions hiérarchiques appliquées aux paramètres modaux corrigés par le filtre linéaire (cf. Tableaux B.6). En général, les taux de classification correcte obtenus sont pratiquement du même ordre que ceux évalués précédemment.

Les classifications obtenues en utilisant la méthode des agglomérations hiérarchiques sont montrées dans le Tableau B.7. Les pourcentages sont plus faibles que ceux obtenus avec la méthode précédente.

Par ailleurs, les nuées dynamiques sont considérées. Dans ce cas, les résultats sont plus pertinents que la méthode des agglomérations hiérarchiques, mais les taux de classification cor-

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	93 (14/15)	80 (12/15)	100 (15/15)	80 (12/15)
Après	85 (11/13)	100 (13/13)	85 (11/13)	100 (13/13)

TABLEAU B.6 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (régression linéaire ; valeurs en %).

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	80 (12/15)	60 (9/15)	80 (12/15)	67 (10/15)
Après	62 (8/13)	60 (9/13)	46 (6/13)	60 (9/13)

TABLEAU B.7 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (régression linéaire ; valeurs en %).

recte sont relativement plus faibles que ceux obtenus sans utiliser de filtre. Les pourcentages de classification correcte en considérant les fréquences et les modes propres sont donnés dans le Tableau B.8.

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	93 (14/15)	87 (13/15)	93 (14/15)	93 (14/15)
Après	85 (11/13)	77 (10/13)	92 (12/13)	77 (10/13)

TABLEAU B.8 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (régression linéaire ; valeurs en %).

De façon similaire à celle présentée pour le filtre linéaire, une deuxième technique, basée sur la régression non linéaire à partir des réseaux de neurones, est proposée. Dans un premier temps, la méthode des divisions hiérarchiques est appliquée aux fréquences et modes propres. Le Tableau B.9 regroupe les taux de classification obtenus. Dans les deux cas, les résultats ne changent pas significativement. En considérant les paramètres modaux transformés en intervalles interquartiles, les résultats sont légèrement meilleurs.

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	93 (14/15)	87 (13/15)	100 (15/15)	80 (12/15)
Après	77 (10/13)	92 (12/13)	85 (11/13)	100 (13/13)

TABLEAU B.9 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (régression non linéaire ; valeurs en %).

Dans le cas des agglomérations hiérarchiques, les taux de classification correcte sont plus élevés que ceux obtenus par le filtre linéaire (cf. Tableau B.10). Les résultats sont cependant pratiquement du même ordre que ceux évalués sans la prise en compte des effets thermiques.

Le Tableau B.11 regroupe les pourcentages de classification correcte obtenus lorsque les nuées dynamiques sont appliquées aux paramètres modaux corrigés par le filtre non linéaire. Dans le

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	80 (12/15)	67 (10/15)	80 (12/15)	67 (10/15)
Après	77 (10/13)	69 (9/13)	46 (6/13)	60 (9/13)

TABLEAU B.10 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (régression non linéaire ; valeurs en %).

cas des fréquences propres, les résultats sont moins pertinents que ceux obtenus par le filtre linéaire ou encore sans aucun filtre. À l'inverse, les taux sont à la fois plus important pour les déformées modales, notamment pour le cluster « après ».

	Avec correction		Sans correction	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
Avant	87 (13/15)	80 (12/15)	93 (14/15)	93 (14/15)
Après	77 (10/13)	100 (13/13)	92 (12/13)	77 (10/13)

TABLEAU B.11 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (régression non linéaire ; valeurs en %).

B.3.5 Indices de partitionnement optimal

Comme il a été présenté dans l'étude numérique de sensibilité au chapitre 4, les indices de partitionnement optimal introduits dans le paragraphe 3.1.5 sont employés dans cette application expérimentale. L'objectif est de vérifier l'aptitude de ces indices à indiquer le nombre d'états structuraux existants. Comme expliqué précédemment, seules les méthodes des divisions hiérarchiques et des nuées dynamiques sont prises en compte dans cette étude. En effet, ces deux techniques nécessitent la définition du nombre de clusters existants au préalable, ce qui n'est pas le cas des agglomérations hiérarchiques.

B.3.5.1 Signaux

L'analyse du Tableau B.12 permet de vérifier que les indices CH et Γ indiquent de manière consistante le nombre de clusters attendu (i.e., trois), si la transformation en intervalles interquartiles est utilisée. L'indice C^*, toutefois, n'indique jamais le nombre d'états structuraux théoriquement existants. Il est intéressant d'observer que cet indice signale systématiquement un nombre de clusters toujours plus élevé qu'en réalité. Il se peut qu'il soit très sensible à des petites modifications, ce qui peut justifier la mise en évidence d'autres groupes de comportement.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	C^*	Γ	CH	C^*	Γ
5	18,4	0,207	0,65	15,6	0,156	0,61
4	21,5	**0,163**	0,68	19,8	**0,134**	0,58
3	**30,0**	0,178	**0,72**	**32,2**	0,145	**0,66**
2	23,6	0,189	0,67	24,6	0,185	0,64

TABLEAU B.12 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en intervalles (états « avant », « pendant » et « après »).

B.3.5.2 Fréquences propres

Dans ce paragraphe, les indices sont évalués lorsque les fréquences propres sont considérées. Le Tableau B.13 regroupe les valeurs calculées. Dans ce cas, il est noté qu'un bon accord est trouvé parmi les trois indices, pour les deux méthodes de clustering. En effet, si les partitionnements sont obtenus à partir des intervalles, les trois indices indiquent le nombre attendu d'états structuraux.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	26,5	0,13	0,66	29,69	0,2	0,62
4	27,9	0,116	0,79	21,5	0,07	0,62
3	**34,0**	**0,107**	**0,83**	**39,22**	**0,06**	**0,76**
2	22,1	0,124	0,72	24,32	0,08	0,69

TABLEAU B.13 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en intervalles (états « avant », « pendant » et « après »).

Les indices calculés en utilisant les fréquences propres corrigées par le filtre linéaire sont regroupés dans le Tableau B.14. Les indices CH et Γ indiquent généralement trois clusters comme le partitionnement optimal. Toutefois, ces résultats ne sont pas vérifiés pour la méthode des nuées dynamiques appliquées aux intervalles interquartiles. L'indice C^*, comme déjà observé, indique toujours un nombre de clusters supérieur à ceux que signalés par les autres indices.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	14,2	**0,036**	0,47	48,4	**0,019**	0,8
4	13,5	0,16	0,54	48	0,044	**0,85**
3	**17,5**	0,232	**0,65**	**50,7**	0,036	0,83
2	10,6	0,394	0,29	49	0,042	0,82

TABLEAU B.14 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformées en intervalles (régression linéaire).

Une procédure similaire est menée en tenant compte de la correction par le filtre non linéaire (cf. Tableau B.15). Dans le cas des fréquences propres, la détermination du nombre attendu de clusters est dégradée. Ceci peut être expliqué par les résultats moins pertinents du partitionnement obtenu en utilisant les fréquences corrigées. L'indice CH arrive cependant à indiquer correctement le nombre de clusters dans la plupart des cas, suivi de l'indice Γ.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C*†*	Γ	*CH*	*C*†*	Γ
5	12,15	**3,2**	0,85	21,14	1,6	0,81
4	12,16	4,1	**0,87**	23,66	1,4	0,86
3	**13,63**	4,0	0,84	23,34	1,5	**0,88**
2	12,41	4,0	0,84	**27,25**	**1,3**	0,86

TABLEAU B.15 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformées en intervalles (régression non linéaire). † Les valeurs de l'indice C^* sont multipliées par $(\times 10^{-2})$.

B.3.5.3 Modes propres

Les indices sont également calculés dans ce paragraphe pour chaque partitionnement obtenu en utilisant les modes propres. Pour la méthode des nuées dynamiques et en utilisant des intervalles, par exemple, seul indice Γ indique le nombre attendu de clusters. Pour les divisions hiérarchiques, l'indice CH signale également trois groupes comme partitionnement optimal.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	23	**0,030**	0,56	33,2	**0,092**	0,6
4	13,1	0,064	0,84	40,6	0,11	0,58
3	**24,8**	0,123	**0,86**	41,5	0,325	**0,66**
2	24,1	0,147	0,7	**41,6**	0,481	0,63

TABLEAU B.16 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformées en intervalles (états « avant », « pendant » et « après »).

En utilisant les modes propres corrigés par le filtre linéaire, les indices oscillent significativement lorsque la méthode des divisions hiérarchiques est considérée. Dans cette analyse, seul l'indice Γ signale le nombre attendu de groupes quelle que soit la méthode utilisée (cf. Tableau B.17).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	**38,0**	0,034	0,76	34,7	**0,039**	0,48
4	33,6	0,039	0,86	**44,2**	0,047	0,84
3	23,1	**0,025**	**0,98**	38	0,294	**0,93**
2	20,7	0,793	0,92	41,3	0,314	0,89

TABLEAU B.17 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigés » transformés en intervalles (régression linéaire).

Enfin, les indices sont calculés en considérant le filtre non linéaire. En général, les valeurs obtenues avec la méthode des nuées dynamiques sont plus adéquates que celles avec la méthode des divisions hiérarchiques. En effet, en considérant cette dernière appliquée à des intervalles, seul l'indice CH signale trois clusters comme le partitionnement optimal. Dans le cas des nuées dynamiques, ces deux indices présentent un bon accord, en signalant toujours trois clusters (cf. Tableau B.18).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	24,9	**0,089**	**0,76**	18,6	**0,011**	0,81
4	24,5	0,162	0,75	19,8	0,011	0,76
3	**27,7**	0,173	0,73	**22,4**	0,014	**0,87**
2	25,9	0,161	0,72	17,5	0,014	0,85

TABLEAU B.18 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigés » transformés en intervalles et (régression non linéaire).

B.3.5.4 Étude complémentaire

Une brève étude complémentaire est proposée ci-dessous. Les simulations de partitionnement réalisées dans les paragraphes précédents ont considéré les essais correspondant aux trois états structuraux du pont-rail. Dans les paragraphes précédents, une étude complémentaire a été menée en ne considérant que les états « avant » et « après » resserrage des bielles du pont-rail.

Signaux

Le Tableau B.19 regroupe les valeurs calculées pour chaque indice en considérant les signaux. L'indice CH indique deux groupes pour toutes les méthodes, ce qui est en accord avec l'analyse menée. L'indice Γ, toutefois, ne signale pas deux clusters comme étant le partitionnement optimal lorsque les divisions hiérarchiques sont considérées.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	CH	C^*	Γ
5	11	0,258	0,64	19,3	**0,062**	0,62
4	12,4	0,13	0,64	15,9	0,077	0,75
3	15,2	**0,092**	**0,64**	16,6	0,088	0,73
2	**18,0**	0,096	0,64	**21,0**	0,128	**0,80**

TABLEAU B.19 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en intervalles (états « avant » et « après »).

Fréquences propres

L'analyse suivante considère les fréquences propres. Dans le Tableau B.20, il est montré que tous les indices indiquent deux clusters lorsque la méthode des divisions hiérarchiques est appliquée à des intervalles. Dans les autres cas, seul l'indice C^* indique un nombre plus important de groupes (quatre).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	CH	C^*	Γ
5	13,1	0,194	0,48	30,4	0,113	0,49
4	15,5	0,148	0,58	30,6	**0,084**	0,61
3	14	0,136	0,65	31,7	0,099	0,82
2	**17,1**	**0,029**	**0,88**	**32,1**	0,121	**0,83**

TABLEAU B.20 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en intervalles (états « avant », et « après »).

La procédure de correction thermique est également considérée. Le Tableau B.21 regroupe les valeurs calculées à partir des partitionnements obtenus en utilisant les fréquences propres compensées par le filtre thermique. En général, les indices CH et Γ indiquent toujours deux groupes comme étant le nombre optimal. Toutefois, l'indice C^* oscille entre 3 et 5 clusters.

En utilisant le filtre non linéaire, les résultats sont pratiquement identiques. Si les fréquences propres sont considérées, l'indice Γ (évalué pour le partitionnement obtenu avec les nuées dynamiques) n'indique pas le nombre attendu de clusters (cf. Tableau B.22). De plus, en considérant la méthode des divisions hiérarchiques appliquées aux intervalles, l'indice CH ne concorde pas avec les autres indices évalués. En effet, cela est expliqué par les faibles taux de classification correcte obtenus précédemment par cette méthode.

Modes propres

Les indices sont maintenant calculés en utilisant les modes propres (cf. Tableau B.23). Pour

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	18,7	0,02	0,75	32,6	**0,064**	0,68
4	19,5	0,02	0,9	39,2	0,074	0,71
3	25,3	**0,006**	0,9	45,7	0,102	0,74
2	**27,0**	0,011	**0,92**	**51,7**	0,146	**0,80**

TABLEAU B.21 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformés en intervalles (états « avant » et « après » ; régression linéaire).

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	37,1	**0,056**	0,55	27,4	**0,123**	0,52
4	37,7	0,061	0,58	29,6	0,142	0,51
3	**44,3**	0,065	0,59	32,6	0,165	0,53
2	42,4	0,073	**0,63**	**35,4**	0,187	**0,55**

TABLEAU B.22 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres « corrigées » transformés en intervalles (états « avant » et « après » ; régression non linéaire).

les deux méthodes, seuls indices CH et Γ signalent le nombre attendu de groupes.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	16,4	**0,110**	0,63	18,2	0,057	0,65
4	18,1	0,143	0,61	19,3	0,051	0,68
3	14,6	0,155	0,61	21,5	**0,042**	0,62
2	**22,5**	0,186	**0,64**	**27,8**	0,102	**0,72**

TABLEAU B.23 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en intervalles (états « avant », et « après »).

En considérant le filtre linéaire, peu de différence est observée dans les résultats. En effet, seul indice C^*, au lieu d'indiquer cinq clusters pour la méthode des divisions hiérarchiques comme précédemment, signale maintenant quatre clusters (cf. Tableau B.24).

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	25,3	0,084	0,61	31,3	0,041	0,46
4	31,8	**0,063**	0,64	36,1	**0,038**	0,5
3	35,4	0,067	0,65	41,1	0,051	0,53
2	**38,1**	0,079	**0,70**	**45,7**	0,075	**0,58**

TABLEAU B.24 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigées » transformés en intervalles (états « avant » et « après » ; régression linéaire).

En utilisant le filtre non linéaire, les résultats sont pratiquement identiques à ceux obtenus sans correction. Le Tableau B.25 regroupe les valeurs calculées en considérant les déformées propres. Dans ce cas, il est noté un nombre important d'indications incorrectes. En considérant les divisions hiérarchiques, par exemple, l'indice CH n'indique jamais deux clusters comme étant le partitionnement optimal. Par ailleurs, pour la méthode des nuées dynamiques, l'indice Γ signale trois ou cinq clusters.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	44,9	**0,057**	0,65	27,4	0,241	**0,76**
4	45,5	0,062	0,65	29,6	0,136	0,71
3	**58,0**	0,074	0,73	32,6	**0,078**	0,73
2	53,1	0,089	**0,80**	**35,4**	0,09	0,75

TABLEAU B.25 – Indices basés sur les méthodes de clustering appliquées aux modes propres « corrigées » transformés en intervalles (états « avant » et « après » ; régression non linéaire).

B.4 ADS couplée aux méthodes de classification supervisée

Dans ce paragraphe, les résultats obtenus à partir du couplage entre l'ADS et les méthodes de classification supervisée appliquées aux signaux et paramètres modaux, sont présentés. Les simulations réalisées ici suivent la même démarche que celle utilisée dans les applications numériques au chapitre 4, en créant des groupes d'entraînement, de validation et de test à partir de l'ensemble des essais.

B.4.1 Première étude

Le Tableau B.26 regroupe les résultats obtenus par la méthode des ADB appliqués aux signaux et paramètres modaux, transformés en intervalles interquartiles. Dans ce cas, et en considérant les critères précédents, 62% des essais (8/13) représentés par des signaux, sont classés correctement dans les deux classes, tandis que 77% le sont en utilisant les paramètres modaux. Il est intéressant de remarquer que l'essai TGV6R est systématiquement classé incorrectement. Ceci peut être expliqué par le fait que cet essai se trouve dans la transition entre les phases 2 et 3 et présente donc un caractère ambigu.

Essai	Classement espéré	Classement obtenu		
		Signaux	Fréquences	Modes
TGV1R	1	1	1	1
TGV2R	1	1	1	1
TGV3R	1	2	1	1
TGV4R	1	2	1	1
TGV5R	1	1	1	2
TGV6R	1	2	2	2
TGV7R	2	2	2	1
TGV8R	2	2	2	2
TGV9R	2	1	1	2
TGV10R	2	1	1	2
TGV11R	2	2	2	2
TGV12R	2	2	2	2
TGV13R	2	2	2	2

TABLEAU B.26 – Classification d'essais obtenue par les ADB (intervalles).

Maintenant, les procédures de correction thermique sont considérées. De façon générale, il est noté des modifications sur l'ensemble résultats. Dans le cas des fréquences propres, par exemple, l'essai TGV8R est maintenant classé au groupe 1, contrairement au cas sans prise en compte des effets thermiques. En réalité, l'utilisation des filtres thermiques a dégradé les classification

des essais. Par ailleurs, il est observé que l'essai TGV6R est toujours classé au groupe 2 (cf. Tableau B.27).

Essai	Classement espéré	Classement obtenu			
		Filtre linéaire		Filtre non linéaire	
		Fréquences	Modes	Fréquences	Modes
TGV1R	1	1	1	1	1
TGV2R	1	1	1	1	1
TGV3R	1	1	1	1	1
TGV4R	1	2	2	2	1
TGV5R	1	2	1	1	1
TGV6R	1	2	2	2	2
TGV7R	2	2	2	2	1
TGV8R	2	1	2	2	2
TGV9R	2	2	2	1	1
TGV10R	2	2	2	1	2
TGV11R	2	2	2	2	2
TGV12R	2	2	2	2	2
TGV13R	2	2	2	2	2

TABLEAU B.27 – Classification d'essais obtenue par les ADB (correction thermique ; intervalles).

Les taux de classification correcte obtenus en utilisant les RN sont détaillées dans le Tableau B.28. En considérant les signaux, les taux de classification correcte sont similaires à ceux obtenus avec les ADB : 70% (9/13) pour la transformation en intervalles. Les taux sont significativement meilleurs en considérant les paramètres modaux. Dans ce cas, 10 essais sur 13 (77%) sont bien classés en utilisant des intervalles (le même pour les modes propres). En utilisant les modes propres, le taux est de 85%, soit 11 essais sur 13. La remarque par rapport à l'essai TGV6R reste valable en utilisant cette méthode.

Essai	Classement espéré	Classement obtenu		
		Signaux	Fréquences	Modes
TGV1R	1	1	1	1
TGV2R	1	1	1	1
TGV3R	1	1	2	1
TGV4R	1	1	1	2
TGV5R	1	2	1	1
TGV6R	1	2	2	2
TGV7R	2	2	2	1
TGV8R	2	1	2	2
TGV9R	2	2	2	2
TGV10R	2	1	1	1
TGV11R	2	2	2	2
TGV12R	2	2	2	2
TGV13R	2	2	2	2

TABLEAU B.28 – Classification d'essais obtenue par les RN (intervalles).

Des résultats similaires sont obtenus en considérant les procédures de correction thermique. De manière générale, peu de changements est observé. Dans cette analyse, le filtre non linéaire mène à des meilleurs résultats que ceux du filtre linéaire. Dans tous les cas, les taux de classifica-

tion correcte sont légèrement supérieurs à ceux évalués sans prise en compte des effets thermiques (cf. Tableau B.29).

Essai	Classement espéré	Classement obtenu			
		Filtre linéaire		Filtre non linéaire	
		Fréquences	Modes	Fréquences	Modes
TGV1R	1	1	1	1	1
TGV2R	1	1	1	1	1
TGV3R	1	1	1	1	1
TGV4R	1	2	2	1	1
TGV5R	2	1	1	1	1
TGV6R	1	2	2	2	2
TGV7R	2	2	2	2	2
TGV8R	2	2	2	2	2
TGV9R	1	1	2	1	2
TGV10R	2	2	1	2	1
TGV11R	2	2	2	2	2
TGV12R	2	2	2	2	2
TGV13R	2	2	2	2	2

TABLEAU B.29 – Classification d'essais obtenue par les RN (correction thermique ; intervalles).

Par ailleurs, le Tableau B.30 regroupe les résultats obtenus par les SVM. En utilisant des intervalles, le taux est relativement faible : 62% (8/13). Les taux de classification en considérant les paramètres modaux ont diminué par rapport à ceux obtenus avec les RN. Pour les fréquences propres transformées en intervalles interquartiles, 9 essais sur 13 (70%) sont correctement classés. Les taux obtenus pour les modes propres est égal à 77% (10/13).

Essai	Classement espéré	Classement obtenu		
		Signaux	Fréquences	Modes
TGV1R	1	1	1	1
TGV2R	1	1	1	1
TGV3R	1	1	1	1
TGV4R	1	1	1	2
TGV5R	1	2	2	1
TGV6R	1	2	2	2
TGV7R	2	1	1	1
TGV8R	2	2	2	2
TGV9R	2	2	2	2
TGV10R	2	1	1	2
TGV11R	2	1	2	2
TGV12R	2	2	2	2
TGV13R	2	2	2	2

TABLEAU B.30 – Classification d'essais obtenue par les SVM (intervalles).

Comme il a été observé avec les RN, l'utilisation des filtres thermiques couplés à la méthode des SVM améliore légèrement les taux de classification correcte. Il est à noté que le filtre linéaire semble être plus pertinent dans cette analyse. Dans le cas des fréquences propres, par exemple, seuls les TGV6R et TGV7R sont classés incorrectement. Pour les modes propres une classification similaire est obtenue. L'utilisation d'un filtre non linéaire mène à des résultats relativement moins

pertinents, notamment dans le cas des déformées modales (cf. Tableau B.31).

Essai	Classement espéré	Classement obtenu			
		Filtre linéaire		Filtre non linéaire	
		Fréquences	Modes	Fréquences	Modes
TGV1R	1	1	1	1	1
TGV2R	1	1	1	1	1
TGV3R	1	1	1	1	1
TGV4R	1	1	1	2	1
TGV5R	1	2	1	1	2
TGV6R	1	2	2	2	2
TGV7R	2	1	1	2	1
TGV8R	2	2	2	2	2
TGV9R	2	2	1	1	1
TGV10R	2	2	2	2	1
TGV11R	2	2	2	2	2
TGV12R	2	2	2	2	2
TGV13R	2	2	2	2	2

TABLEAU B.31 – Classification d'essais obtenue par les SVM (correction thermique ; intervalles).

B.4.2 Deuxième étude

Cette deuxième étude a pour objectif de vérifier la robustesse et l'efficacité des méthodes de classification supervisée pour la discrimination des trois états structuraux. Pour cela, une étude statistique est menée en faisant varier la taille des groupes d'entraînement et de test. Au total, 10000 simulations sont réalisées pour chaque configuration et les meilleurs taux de classification correcte sont évalués ainsi que les taux moyens et minimaux.

B.4.2.1 Signaux

La simulation suivante correspond à l'application des méthodes de classification aux signaux, transformés en intervalles interquartiles (cf. Tableau B.32). Dans le premier cas, où seulement 30% des essais sont utilisés pour l'entraînement, il est noté que les trois méthodes mènent à des résultats similaires. En effet, seuls les RN et SVM réussissent, à classer parfaitement tous les essais (au moins une fois). Par ailleurs, il apparaît clairement qu'à mesure que le nombre d'essais dans le groupe d'entraînement augmente, les taux de classification correcte sont plus élevés pour toutes les méthodes. Il est également remarqué que dans cette analyse, les SVM ont produit des pourcentages moyens plus importants que ceux avec les autres méthodes.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	89	100	100	90	100	100	90	100	100
Moyen	78	81	77	77	82	85	82	87	91
Minimal	20	40	45	27	43	15	30	62	28

TABLEAU B.32 – Taux de classification correcte utilisant les signaux transformés en intervalles pour les états « avant », « pendant » et « après » (valeurs en %).

B.4.2.2 Fréquences propres

Le Tableau B.33 regroupe les pourcentages de classification correcte obtenus lorsque les fréquences propres sont transformées en intervalles interquartiles. Comme il a été observé dans les méthodes de clustering, les méthodes de classification supervisée permettent d'atteindre des résultats pertinents.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	89	100	100	91	100	100	93	100	100
Moyen	61	88	77	59	90	82	61	94	89
Minimal	22	40	17	25	24	21	28	56	12

TABLEAU B.33 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour les états « avant », « pendant » et « après » (valeurs en %).

Par ailleurs, les méthodes de classification sont appliquées aux fréquences propres corrigées par le filtre linéaire. Le Tableau B.34 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en intervalles interquartiles est considérée. En général, les méthodes de classification supervisée permettent d'atteindre des résultats aussi pertinents que ceux obtenus sans le filtre thermique.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	92	92	92	93	100	95	93	100	100
Moyen	70	74	75	77	84	80	76	83	87
Minimal	41	36	8	36	48	12	29	31	24

TABLEAU B.34 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour les états « avant », « pendant » et « après » (régression linéaire ; valeurs en %).

Le filtre non linéaire est maintenant considérée. Les résultats obtenus en considérant la transformation des fréquences propres en intervalles est détaillée dans le Tableau B.35. Il est noté que les taux de classification pour les ADB sont du même ordre que ceux obtenus sans l'utilisation des lois de correction. Pour les RN et les SVM, toutefois, les résultats sont moins pertinents, ce qui peut indiquer une « perturbation » des valeurs des fréquences propres corrigées.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	85	88	88	84	88	89	84	88	88
Moyen	61	81	83	64	82	82	67	82	83
Minimal	31	15	7	23	28	13	28	31	15

TABLEAU B.35 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour les états « avant », « pendant » et « après » (régression non linéaire ; valeurs en %).

B.4.2.3 Modes propres

Les méthodes de classification supervisée sont maintenant appliquées aux modes propres. Le Tableau B.36 regroupe les taux de classification évalués en considérant la transformation en intervalles interquartiles. Dans ce cas, les RN et les SVM arrivent à classer plus de 70% d'essais

correctement, mais, en moyenne, cette valeur n'est supérieure qu'à 60%, quel que soit le nombre d'essais dans le groupe d'entraînement. Dans cette analyse, les ADB ont obtenu des résultats plus pertinents que les RN, avec les meilleurs taux de classification correcte, de l'ordre de 80%.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	82	67	75	85	71	77	89	75	77
Moyen	40	40	54	44	40	55	50	43	61
Minimal	25	16	23	25	10	25	28	25	18

TABLEAU B.36 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les états « avant », « pendant » et « après » (valeurs en %).

Par ailleurs, les modes propres filtrés par le modèle linéaire sont considérés. Comme remarqué précédemment, les taux de classification correcte sont généralement moins importants que ceux trouvés avec les fréquences propres. De plus, quel que soit le type de transformation en données symboliques utilisé, les résultats obtenus par les ADB et les RN ont évolué par rapport à ceux évalués sans prise en compte des effets thermiques. Toutefois, les taux de classification correcte obtenus par les SVM ont été relativement plus faibles (cf. Tableau B.37).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	77	92	86	77	96	89	83	100	100
Moyen	60	87	61	60	86	59	63	90	57
Minimal	25	22	9	28	30	12	31	25	23

TABLEAU B.37 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les états « avant », « pendant » et « après » (régression linéaire ; valeurs en %).

Les déformées modales corrigées par le filtre non linéaire produit des résultats moins pertinents. En particulier, les taux de classification sont relativement plus faibles que ceux évalués sans filtre thermique dans le cas de l'utilisation des intervalles interquartiles (cf. Tableau B.38).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	75	76	82	77	75	81	81	79	84
Moyen	38	42	45	39	41	50	43	44	55
Minimal	22	17	17	24	19	23	23	26	24

TABLEAU B.38 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les états « avant », « pendant » et « après » (régression non linéaire ; valeurs en %).

B.4.3 Troisième étude

L'objectif de la troisième étude est de reproduire les simulations statistiques, mais en n'utilisant que les essais des états « avant » et « après ». Ce raisonnement a été déjà considéré pour les méthodes de clustering et des résultats relativement meilleurs ont été obtenus. Étant donné que ces états sont plus « éloignés », il est espéré que les méthodes de classification soient capables de mieux les distinguer.

B.4.3.1 Signaux

La première simulation considère l'utilisation des signaux. Il est noté que la transformation en intervalles produit des résultats pertinents, significativement supérieurs à ceux obtenus lors de la deuxième étude. En effet, il est montré dans le Tableau B.39 qu'en moyenne, plus de 80% des essais sont classés correctement. Les meilleurs taux sont obtenus par les RN, suivi par les SVM. Ces méthodes ont déjà montré être plus performantes pour la classification d'essais. Par ailleurs, il est également noté que les pourcentages de classification correcte améliorent à mesure que le nombre d'essais dans le groupe d'entraînement augmente.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	97	100	100	97	100	100	95	100	100
Moyen	76	92	88	77	94	88	83	96	90
Minimal	32	59	15	41	71	28	37	73	43

TABLEAU B.39 – Taux de classification correcte utilisant les signaux transformés en intervalles pour les états « avant » et « après » (valeurs en %).

B.4.3.2 Fréquences propres

Le Tableau B.40 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en intervalles interquartiles est considérée. Les trois méthodes de classification supervisée permettent d'atteindre des résultats pertinents, avec des taux moyens de l'ordre de 90%. De plus, toutes les méthodes ont réussi à classer, au moins une fois, tous les essais correctement.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	100	100	100	100	100	100	100	100	100
Moyen	89	96	92	90	97	93	93	99	96
Minimal	40	59	28	43	82	45	49	86	62

TABLEAU B.40 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour les états « avant » et « après » (valeurs en %).

Les taux de classification correcte en considérant le filtre linéaire sont regroupés dans le Tableau B.41. Les méthodes de classification supervisée permettent généralement d'atteindre des résultats moins pertinents que ceux obtenus sans le filtre thermique (cf. Tableau B.40).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	92	92	92	92	92	92	92	92	92
Moyen	86	93	89	87	94	90	90	96	93
Minimal	38	55	26	40	77	42	46	81	58

TABLEAU B.41 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour les états « avant » et « après » (régression linéaire ; valeurs en %).

Par ailleurs le fréquences propres corrigées par le filtre non linéaire sont considérées (cf. Tableau B.42). Les taux de classification correcte sont légèrement plus importants que ceux avec le filtre linéaire. Les résultats ne montrent pas une amélioration significative par rapport à ceux sans correction thermique.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	95	95	95	95	95	95	95	95	95
Moyen	83	89	86	84	90	87	87	92	89
Minimal	37	55	26	40	76	42	46	80	58

TABLEAU B.42 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour les états « avant » et « après » (régression non linéaire ; valeurs en %).

B.4.3.3 Modes propres

En utilisant la transformation en intervalles, un peu plus de la moitié des essais est classée correctement, en moyenne. La méthode la plus performante est celle basée sur les SVM, qui arrive à classer environ 60% d'essais correctement (cf. Tableau B.43).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	75	71	77	76	76	81	80	91	85
Moyen	55	48	62	55	52	67	59	55	73
Minimal	27	21	18	30	24	21	22	9	36

TABLEAU B.43 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les états « avant » et « après » (valeurs en %).

Le Tableau B.44 regroupe les taux de classification correcte en considérant le filtre linéaire. Les méthodes de classification supervisée ne permettent pas globalement d'atteindre des résultats aussi pertinents que ceux obtenus sans le filtre thermique (cf. Tableau B.43).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	71	67	73	72	72	76	75	86	80
Moyen	51	45	58	51	48	62	55	51	68
Minimal	25	20	17	28	23	20	21	8	34

TABLEAU B.44 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les états « avant » et « après » (régression linéaire ; valeurs en %).

Dans le cas du filtre non linéaire, bien que les taux moyens de classification correcte soient légèrement plus importants que ceux avec le filtre linéaire, il est noté que les taux maximaux sont plus faibles(cf. Tableau B.45). Dans tous les cas, l'utilisation des filtres n'apporte pas une amélioration des résultats.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	68	64	70	69	69	74	73	83	77
Moyen	52	45	59	52	49	63	56	52	69
Minimal	26	20	17	29	23	20	21	9	34

TABLEAU B.45 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour les états « avant » et « après » (régression non linéaire ; valeurs en %).

B.5 Affectation de nouveaux essais

B.5.1 Affectation basée sur les divisions hiérarchiques

Dans ce paragraphe, les arbres construits à partir de la méthode des divisions hiérarchiques sont utilisés pour l'affectation des nouveaux essais.

B.5.1.1 Signaux

Dans le cas des signaux, la Figure B.3, illustre le partitionnement obtenu en utilisant les signaux transformés en intervalles interquartiles. Dans ce cas, il est observé que les variables les plus discriminantes sont les capteurs 7 et 10, avec les valeurs de coupures 0,029 m· s^{-2} et 0,0147 m· s^{-2}, respectivement. Ici, tous les essais sont affectés dans un nouveau cluster. Cela signifie qu'en analysant uniquement les réponses dynamiques, la structure a un comportement différent de ceux déjà connus (i.e., « avant », « pendant » et « après » resserrage). La loi log-normale a été ajustée aux distances évaluées menant à un seuil égal à 2,77.

B.5.1.2 Fréquences propres

Le Tableau B.46 regroupe les taux de classification calculés en considérant l'analyse avec les fréquences propres. Dans ce cas, les résultats sont significativement différents. En utilisant des intervalles interquartiles, il est observé que pour les deux premières campagnes de mesures après resserrage, la plupart des essais sont revenus à l'état « pendant » les travaux. Par ailleurs, pour les deux séries suivantes, le classement obtenu indique l'existence d'un nouveau comportement (ici nommé « NC »), ce qui pourrait être représenté par un état intermédiaire, entre « pendant » et « avant » travaux, par exemple. De plus, pour les essais enregistrés en 2006, il est noté que la majorité des essais est classée dans l'état « avant ». Cela signifie que le procédé de resserrage, au bout de 3 ans, n'est plus efficace. La distribution ajustée et la distance seuil sont, respectivement (Weibull ; 4,44).

Campagne	Avant	Pendant	Après	NC
Déc-04	25	**42**	14	19
Mai-Juin/05	**30**	**30**	19	21
Jul-Oct/05	23	23	14	**40**
Nov-Déc/05	30	18	8	**44**
Jan-Mar/06	**45**	29	11	15
Moyenne	**36**	31	13	20

TABLEAU B.46 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux fréquences propres (intervalles ; valeurs en %).

Le Tableau B.47 regroupe les taux de classification calculés en utilisant le filtre linéaire. La plupart des essais ne sont jamais classés dans le groupe qui indique un nouveau comportement. Il est observé que pour les deux premières campagnes de mesures après resserrage, la plupart des essais sont classés dans le groupe « après » ou « pendant » les travaux. Par ailleurs, pour les deux séries suivantes, il est noté que la majorité des essais est classée à l'état « avant ». Cela montre encore une fois que le procédé de resserrage, au bout de 3 ans, n'est plus efficace. La distribution ajustée est la loi log-normale et la distance seuil est égale à 2,16.

Le filtre non linéaire est maintenant appliqué aux fréquences propres. Le Tableau B.48 regroupe les pourcentages d'affectation calculés à partir des fréquences propres transformées en

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	24	**33**	22	21
Jul-Oct/05	22	27	**33**	18
Nov-Déc/05	27	**35**	19	19
Jan-Mar/06	31	**32**	14	23
Moyenne	26	**32**	22	20

TABLEAU B.47 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux fréquences propres (régression linéaire ; valeurs en %).

intervalles interquartiles. Il est noté que les résultats sont légèrement différents de ceux obtenus en utilisant le filtre linéaire. Ces différences peuvent être expliquées par le mauvais ajustement obtenu par le filtre linéaire. Pratiquement aucun essai n'est classé dans le groupe « après », mais plutôt dans celui correspondant à l'état « pendant ». La loi d'ajustement utilisée est la loi de Weibull (avec un seuil égal à 4,19)

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	23	**28**	23	25
Jul-Oct/05	24	**30**	24	23
Nov-Déc/05	26	**27**	20	27
Jan-Mar/06	**28**	25	19	28
Moyenne	25	**28**	22	26

TABLEAU B.48 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux fréquences propres (régression non linéaire ; valeurs en %).

B.5.1.3 Modes propres

En utilisant les modes propres, les résultats sont ambigus (cf. Tableau B.49). Dans le cas de la transformation en intervalles interquartiles, il est indiqué que soit les essais sont considérés comme appartenant sà l'état « pendant » resserrage (lors des trois premières séries de mesures) soit à l'état « avant », lors des deux dernières campagnes. La loi ajustée est la loi log-normale et la distance seuil est égale à 3,18.

Campagne	Avant	Pendant	Après	NC
Déc-04	5	**71**	0	24
Mai-Juin/05	4	**61**	0	36
Jul-Oct/05	9	**62**	0	29
Nov-Déc/05	**38**	37	0	25
Jan-Mar/06	**40**	37	0	23
Moyenne	19	**54**	0	27

TABLEAU B.49 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux modes propres (intervalles ; valeurs en %).

Le Tableau B.50 regroupe les résultats obtenus en considérant les modes propres corrigés par le filtre linéaire. En général, la plupart des essais ne sont pratiquement jamais classés dans le groupe « après », quel que soit le type de transformation symbolique utilisé. Dans le cas des intervalles interquartiles, la loi ajustée est la loi de Weibull (avec un seuil égal à 2,52).

L'affectation obtenue avec les divisions hiérarchiques appliquées aux modes propres corrigés par le filtre non linéaire est détaillée dans le Tableau B.51. La plupart des essais de la deuxième

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	12	**35**	33	20
Jul-Oct/05	18	25	21	**36**
Nov-Déc/05	29	**32**	12	27
Jan-Mar/06	**33**	24	15	28
Moyenne	23	**29**	20	28

TABLEAU B.50 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux modes propres (régression linéaire ; valeurs en %).

campagne de mesures sont affectés dans l'état « après ». La loi d'ajustement utilisée est la loi log-normale (seuil égal à 3,85).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	22	30	**36**	12
Jul-Oct/05	27	**41**	12	20
Nov-Déc/05	**31**	26	25	18
Jan-Mar/06	**36**	30	29	5
Moyenne	29	**32**	26	14

TABLEAU B.51 – Affectation des nouveaux essais en utilisant les divisions hiérarchiques appliquées aux modes propres (régression non linéaire ; valeurs en %).

B.5.2 Affectation basée sur les nuées dynamiques

B.5.2.1 Signaux

La deuxième démarche pour l'affectation des nouveaux essais considère l'application de la méthode des nuées dynamiques. En utilisant les signaux, les résultats sont similaires à ceux obtenus avec la méthode des divisions hiérarchiques : pratiquement tous les essais sont affectés au cluster NC, ce qui indique un comportement différent de ceux existants (cf. Tableau B.52). La loi ajustée est la loi log-normale. Le seuil de distances est égal à 2,66.

Campagne	Avant	Pendant	Après	NC
Déc-04	0	0	0	100
Mai-Juin/05	0	0	0	100
Juil-Oct/05	0	0	0	100
Nov-Déc/05	4	0	0	96
Jan-Mar/06	1	0	0	99
Moyenne	3	0	0	**97**

TABLEAU B.52 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux signaux (intervalles ; valeurs en %).

B.5.2.2 Fréquences propres

Le Tableau B.53 regroupe les taux d'affectation calculés et il est remarqué que la plupart de essais sont classés dans le cluster « avant » resserrage. Cela signale la perte d'efficacité de la procédure menée et montre que l'approche proposée peut fournir une analyse plus détaillée du

phénomène (ce qui n'était pas possible en observant la Figure 5.40). Dans le cas de la transformation en intervalles interquartiles, la loi de Weibull a été ajustée aux distances intra-cluster, ayant 3,09 comme valeur seuil.

Campagne	Avant	Pendant	Après	NC
Déc-04	**62**	10	24	5
Mai-Juin/05	**59**	0	35	7
Jul-Oct/05	**74**	1	22	4
Nov-Déc/05	**75**	2	13	10
Jan-Mar/06	**76**	9	11	5
Moyenne	**69**	4	21	6

TABLEAU B.53 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux fréquences propres (intervalles ; valeurs en %).

Le Tableau B.54 regroupe les taux d'affectation calculés pour les fréquences propres corrigées par le filtre linéaire. La plupart des essais sont classés dans le cluster « pendant » resserrage (différemment de l'approche sans prise en compte des effets thermiques, qui a indiqué le cluster « avant » resserrage ; cf. Tableau B.53). L'affectation basée sur la transformation en intervalles interquartiles est faite dans le cluster « pendant ». La loi ajustée est la loi log-logistique (seuil égal à 1,76).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	24	**33**	21	21
Jul-Oct/05	24	**34**	26	15
Nov-Déc/05	28	**32**	18	21
Jan-Mar/06	**33**	29	14	24
Moyenne	28	**32**	20	20

TABLEAU B.54 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux fréquences propres (régression linéaire ; valeurs en %).

Les fréquences propres corrigées par le filtre non linéaire sont maintenant analysées. En général, les essais sont affectés soit dans le cluster « pendant » soit dans le cluster « avant » (cf. Tableau B.55). La loi ajustée est la loi log-normale et la distance seuil est égal à 2,97.

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	19	33	**35**	13
Jul-Oct/05	28	**31**	29	12
Nov-Déc/05	**31**	27	17	25
Jan-Mar/06	**40**	20	19	21
Moyenne	**30**	28	25	18

TABLEAU B.55 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux fréquences propres (régression non linéaire ; valeurs en %).

B.5.2.3 Modes propres

La procédure d'affectation est maintenant appliquée aux modes propres (cf. Tableau B.56). La loi ajustée est la loi log-normale avec une distance seuil égale à 2,50. Comme observé pour les fréquences propres, la majorité des essais est classée dans le groupe correspondant à l'état

« avant » de l'ouvrage. Il est aussi remarqué que 24% des essais sont classés en moyenne comme « pendant », dans le cas de la transformation en intervalles interquartiles. Toutefois, pratiquement aucun essai n'est affecté au cluster « après ».

Campagne	Avant	Pendant	Après	NC
Déc-04	**81**	5	0	14
Mai-Juin/05	**67**	12	0	21
Jul-Oct/05	**57**	29	2	11
Nov-Déc/05	**46**	42	2	10
Jan-Mar/06	**59**	30	1	10
Moyenne	**62**	24	1	13

TABLEAU B.56 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux modes propres (intervalles ; valeurs en %).

En considérant les modes propres corrigés par le filtre linéaire, les résultats sont hétérogènes (cf. Tableau B.57). Ici, les essais sont considérés soit dans l'état « avant » resserrage, soit « pendant » soit encore à un nouvel état. La loi ajustée est la loi log-logistique (seuil égal à 1,14).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	15	28	25	**33**
Jul-Oct/05	18	**30**	26	26
Nov-Déc/05	26	**33**	16	25
Jan-Mar/06	**30**	27	16	27
Moyenne	22	**30**	21	28

TABLEAU B.57 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux modes propres (régression linéaire ; valeurs en %).

Les taux d'affectation calculés en considérant les modes propres compensés par le filtre non linéaire sont regroupés dans le Tableau B.58. Comme observé ci-dessus, les essais sont classés dans l'un de quatre clusters possibles. La loi d'ajustement utilisée dans le cas des intervalles est la loi log-logistique (seuil égal à 2,06).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	18	33	**35**	14
Jul-Oct/05	24	**27**	26	23
Nov-Déc/05	30	**33**	15	22
Jan-Mar/06	**34**	21	20	25
Moyenne	27	**29**	24	21

TABLEAU B.58 – Affectation des nouveaux essais en utilisant les nuées dynamiques appliquées aux modes propres (régression non linéaire ; valeurs en %).

B.5.3 Affectation basée sur le calcul de distances

La troisième approche d'affectation des nouveaux essais ne considère aucune méthode de clustering.

B.5.3.1 Signaux

Dans un premier temps, les signaux transformés en données symboliques sont utilisés. Dans ce cas, tous les essais sont affectés au nouveau cluster NC. Ce résultat est en accord avec les classifications déjà obtenues en utilisant les divisions hiérarchiques et les nuées dynamiques. La loi ajustée pour les intervalles interquartiles est la loi log-logistique (seuil égal à 3,18).

B.5.3.2 Fréquences propres

Le calcul des distances est maintenant réalisé en considérant les fréquences propres. Dans ce cas, l'affectation est toujours réalisée soit dans le cluster « pendant » lors des trois premières séries de mesures, soit dans le cluster « avant », pour les dernières campagnes (cf. Tableau B.59). Ainsi, les effets des travaux de resserrage ne sont plus vérifiés. La loi ajustée dans le cas des intervalles interquartiles est la loi de Weibull (seuil égal à 4,03).

Campagne	Avant	Pendant	Après	NC
Déc-04	19	**38**	29	14
Mai-Juin/05	23	**44**	2	31
Jul-Oct/05	32	**46**	6	16
Nov-Déc/05	**50**	21	19	10
Jan-Mar/06	**38**	35	21	6
Moyenne	32	**37**	15	15

TABLEAU B.59 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux fréquences propres (valeurs en %).

Le Tableau B.60 regroupe les pourcentages d'affectation pour chacun des états structuraux, en considérant les fréquences propres corrigées par le filtre linéaire. L'affectation est toujours réalisée soit dans le cluster « pendant » lors des trois premières séries de mesures, soit dans le cluster « avant », pour la dernière campagne. Ces résultats montrent que les effets du resserrage ne sont effectifs. La loi ajustée est la loi de Weibull (seuil égal à 5,04).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	11	**60**	8	21
Jul-Oct/05	**43**	30	25	2
Nov-Déc/05	**62**	27	11	0
Jan-Mar/06	**39**	23	0	38
Moyenne	**39**	35	11	15

TABLEAU B.60 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux fréquences propres (régression linéaire ; valeurs en %).

La démarche est appliquée en considérant le filtre non linéaire. Dans ce cas, les taux d'affectation changent significativement. Maintenant, la plupart des essais sont classées dans le cluster « pendant » (dans le cas des intervalles), comme détaillé dans le Tableau B.61. La loi d'ajustement utilisée est la loi de Weibull (seuil égal à 4,28).

B.5.3.3 Modes propres

L'analyse avec les modes propres est maintenant considérée. Mis à part les essais de la série de mesures réalisée en décembre 2004 qui ont été classés dans le cluster « après », tous les autres

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	25	**55**	11	9
Jul-Oct/05	32	**49**	7	12
Nov-Déc/05	33	**43**	18	6
Jan-Mar/06	**41**	28	22	9
Moyenne	33	**44**	15	9

TABLEAU B.61 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux fréquences propres (régression non linéaire ; valeurs en %).

essais sont affectés au cluster « avant » resserrage (cf. Tableau B.62). La loi ajustée est la loi log-logistique (seuil égal à 1,20).

Campagne	Avant	Pendant	Après	NC
Déc-04	19	10	**57**	14
Mai-Juin/05	**39**	31	9	21
Jul-Oct/05	**47**	22	12	18
Nov-Déc/05	**50**	15	10	25
Jan-Mar/06	**57**	22	5	16
Moyenne	**42**	20	19	19

TABLEAU B.62 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux modes propres (valeurs en %).

L'analyse avec les modes propres corrigés est aussi considérée. En utilisant la transformation en intervalles, mis à part les essais de la série de mesures réalisée entre mai et juin 2005, qui ont été classés dans le cluster « après », tous les autres essais sont affectés soit au cluster « pendant », soit à celui « après » resserrage (cf. Tableau B.63). L'ajustement est réalisé en utilisant la loi log-logistique (seuil égal à 2,83).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	15	31	**32**	22
Jul-Oct/05	21	**28**	25	26
Nov-Déc/05	31	**35**	15	19
Jan-Mar/06	**33**	29	16	22
Moyenne	25	**31**	22	22

TABLEAU B.63 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux modes propres (régression linéaire ; valeurs en %).

Enfin, les résultats évalués en utilisant les modes propres (filtre non linéaire) sont donnés dans le Tableau B.64. Les résultats ne concordent pas totalement avec ceux obtenus en utilisant le filtre linéaire, ce qui remet encore une fois la validité de son utilisation dans le cas étudié. La loi d'ajustement utilisée dans le cas des intervalles est la loi log-normale (seuil égal à 1,95).

B.5.4 Affectation basée sur les méthodes de classification supervisée

À titre de complément, une approche alternative est considérée dans ce paragraphe. Cette démarche consiste à utiliser les méthodes de classification supervisée pour l'affectation des nouveaux essais dans l'une des classes connues (i.e., « avant », « pendant » et « après » resserrage).

Campagne	Avant	Pendant	Après	NC
Mai-Juin/05	16	33	13	**38**
Jul-Oct/05	35	**37**	18	10
Nov-Déc/05	**31**	29	19	21
Jan-Mar/06	31	**34**	18	17
Moyenne	28	**33**	17	22

TABLEAU B.64 – Affectation des nouveaux essais basée sur l'évaluation de distances appliquée aux modes propres (régression non linéaire ; valeurs en %).

B.5.4.1 Signaux

Les affectations obtenues pour chaque méthode en utilisant les signaux transformés en intervalles interquartiles sont regroupées dans le Tableau B.65. Toutes les méthodes indiquent que 100% des nouveaux essais sont affectés dans la classe correspondante à l'état « pendant » renforcement.

Campagne	**ADB**			**RN**			**SVM**		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Déc/04	21	34	**45**	0	100	0	0	100	0
Mai-Juin/04	8	**89**	3	0	100	0	0	100	0
Jul-Oct/05	23	**75**	2	0	100	0	0	100	0
Nov-Déc/05	14	**81**	5	0	100	0	0	100	0
Jan-Mar/06	33	**59**	8	0	100	0	0	100	0
Moyenne	19	**76**	5	0	**100**	0	0	**100**	0

TABLEAU B.65 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux signaux transformés en intervalles interquartiles (valeurs en %).

B.5.4.2 Fréquences propres

Cette démarche est maintenant appliquée aux fréquences propres. La tendance déjà observée en utilisant les signaux est encore vérifiée. En considérant la transformation en intervalles interquartiles, les trois méthodes affectent les nouveaux essais dans la classe « pendant » (cf. Tableau B.66).

Campagne	**ADB**			**RN**			**SVM**		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Déc/04	37	**44**	19	0	100	0	35	**37**	28
Mai-Juin/04	39	**43**	18	0	100	0	28	**68**	4
Jul-Oct/05	**45**	22	33	0	100	0	30	**45**	25
Nov-Déc/05	22	**56**	22	0	100	0	25	**55**	20
Jan-Mar/06	33	**47**	20	0	100	0	12	**77**	11
Moyenne	35	**42**	22	0	**100**	0	26	**56**	18

TABLEAU B.66 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux fréquences propres transformées en intervalles interquartiles (valeurs en %).

Le Tableau B.67 regroupe les affectations obtenues pour chaque méthode en utilisant les fréquences propres transformées en intervalles interquartiles (filtre linéaire). Dans ce cas, la méthode basée sur les RN indique que 100% des nouveaux essais sont affectés dans la classe correspondante

à l'état « pendant » renforcement. Pour les ADB et les SVM, ceci est aussi le cas, à l'exception de la dernière campagne de mesures où l'affectation est réalisée à la classe « avant » les travaux.

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	**43**	35	22	0	100	0	33	**34**	33
Jul-Oct/05	**42**	33	25	0	100	0	31	**44**	25
Nov-Déc/05	34	**38**	28	0	100	0	**35**	33	32
Jan-Mar/06	**39**	31	30	0	100	0	**39**	30	31
Moyenne	**40**	34	26	0	**100**	0	**35**	35	30

TABLEAU B.67 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux fréquences propres transformées en intervalles interquartiles (régression linéaire ; valeurs en %).

Dans le cas des fréquences propres corrigées par le filtre non linéaire, les RN indiquent toujours l'affectation de tous les essais dans le cluster « pendant ». Les ADB et les SVM, toutefois, tendent à partager les affectation entre les groupes « pendant » et « après » resserrage (cf. Tableau B.68).

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	**45**	22	33	0	100	0	23	**44**	33
Jul-Oct/05	**48**	33	19	0	100	0	29	**39**	32
Nov-Déc/05	37	28	**35**	0	100	0	**37**	30	33
Jan-Mar/06	**39**	25	36	0	100	0	**44**	25	31
Moyenne	**42**	27	31	0	**100**	0	33	**35**	32

TABLEAU B.68 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux fréquences propres transformées en intervalles interquartiles (régression non linéaire ; valeurs en %).

B.5.4.3 Modes propres

Le Tableau B.69 regroupe l'affectation obtenue en utilisant les modes propres transformés en intervalles interquartiles. Les résultats obtenus avec les méthodes des RN et des SVM sont les mêmes. Dans les deux cas, 100% des essais sont classés comme « pendant ». En utilisant les ADB, cela n'est vérifié que pour les trois dernières campagnes de mesures. Pour les deux premières, toutefois, la plupart des essais sont affectés à la classe « après ».

Campagne	ADB			RN			SVM		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Déc/04	25	35	**40**	0	100	0	0	100	0
Mai-Juin/04	29	35	**36**	0	100	0	0	100	0
Jul-Oct/05	30	**37**	33	0	100	0	0	100	0
Nov-Déc/05	30	**44**	26	0	100	0	0	100	0
Jan-Mar/06	21	**49**	30	0	100	0	0	100	0
Moyenne	27	**40**	33	0	**100**	0	0	**100**	0

TABLEAU B.69 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux modes propres transformés en intervalles interquartiles (valeurs en %).

De plus, les modes propres corrigés par le filtre linéaire sont considérés. En utilisant des intervalles interquartiles, les résultats obtenus par les méthodes des RN et des SVM sont les mêmes. Dans les deux cas, 100% des essais sont classés comme « pendant » (cf. Tableau B.70). En utilisant les ADB, cela n'est vérifié que pour les deux dernières campagnes de mesures.

Campagne	**ADB**			**RN**			**SVM**		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	**41**	34	25	0	100	0	0	100	0
Jul-Oct/05	29	**44**	27	0	100	0	0	100	0
Nov-Déc/05	**41**	29	30	0	100	0	0	100	0
Jan-Mar/06	**39**	26	35	0	100	0	0	100	0
Moyenne	38	33	**29**	0	**100**	0	0	**100**	0

TABLEAU B.70 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux modes propres transformés en intervalles interquartiles (régression linéaire ; valeurs en %).

Finalement, en considérant les modes propres compensés par le filtre linéaire, les affectations sont exactement les mêmes pour les RN et les SVM (cf. Tableau B.71). Les deux méthodes indiquent la classification des essais dans le cluster « pendant ». Dans le cas des ADB, l'affectation des essais est soit réalisée dans le groupe « pendant », soit dans le groupe « avant ».

Campagne	**ADB**			**RN**			**SVM**		
	Avant	Pendant	Après	Avant	Pendant	Après	Avant	Pendant	Après
Mai-Juin/05	35	**39**	26	0	100	0	0	100	0
Jul-Oct/05	29	**41**	30	0	100	0	0	100	0
Nov-Déc/05	26	**39**	35	0	100	0	0	100	0
Jan-Mar/06	**37**	28	35	0	100	0	0	100	0
Moyenne	32	**37**	32	0	**100**	0	0	**100**	0

TABLEAU B.71 – Affectation des nouveaux essais basée sur les méthodes de classification appliquées aux modes propres transformés en intervalles interquartiles (régression non linéaire ; valeurs en %).

Annexe C

Résultats complémentaires - Pont PI-57 sur l'Oise

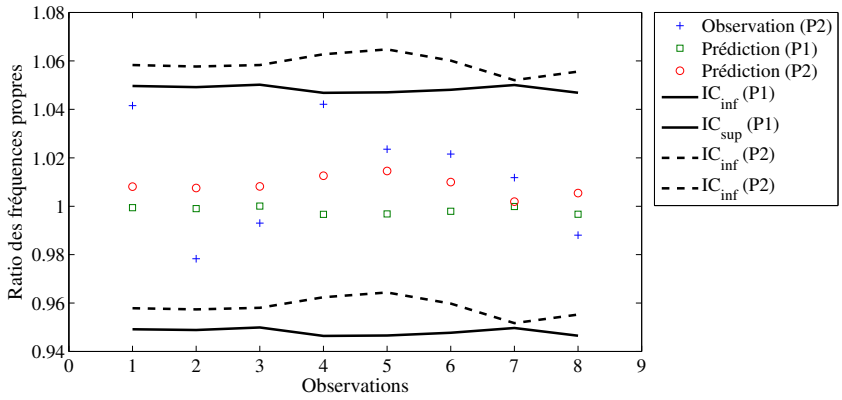
Dans cette annexe, des résultats complémentaires obtenus pour le pont PI-57 sur l'Oise sont présentés. Ces résultats sont obtenus à partir de la transformation des signaux et des paramètres modaux en intervalles interquartiles. Chaque paragraphe correspond à une analyse exploitée dans le chapitre 6.

C.1 Étude des effets thermiques

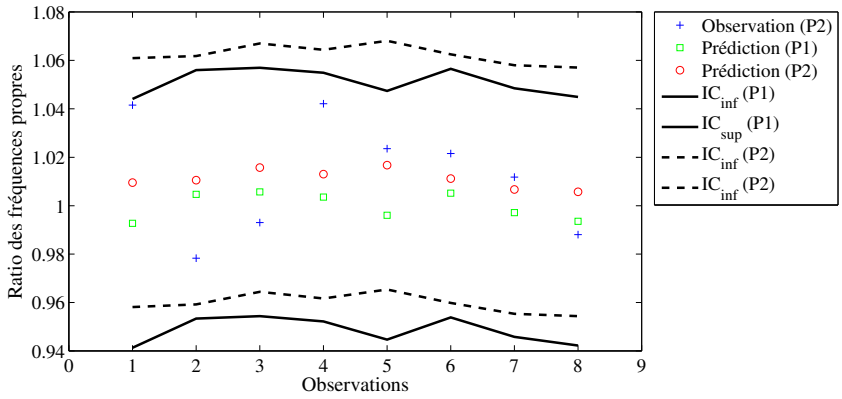
Comme il a été proposé dans l'application expérimentale sur le pont-rail au chapitre 5, une étude des effets thermiques sur les paramètres modaux est considérée. Comme décrit dans le paragraphe 6.1, sept sondes ont été placées à l'intérieur et l'extérieur du caisson du pont afin d'enregistrer les températures durant les deux campagnes de mesures. Dans ce paragraphe, deux études sont proposées. Dans un premier temps, la procédure de « correction » des paramètres modaux est mise en œuvre. Des modèles de régression linéaire et non linéaires sont créés et les méthodes de clustering et de classification supervisée sont appliquées aux fréquences et modes propres « compensés ». Le but est d'observer comment les taux de classification correcte changent ou non, en considérant cette approche. Dans un deuxième temps, les techniques de régression sont utilisées pour la définition de modèles de prédiction. L'objectif est de détecter des variations anormales sur les fréquences propres (i.e., qui ne sont pas liées aux changements de température) et de vérifier, d'une part, les modifications sur les déformées modales et d'autre part, comment ces essais se retrouvent par rapport à ceux considérés comme « normaux ».

Les Figures C.1 et C.2 illustrent un comparatif des prédictions réalisées pour les ratios des deuxième et troisième fréquences propres, respectivement, identifiées lors de la phase 2, à la température de référence. Pour ces deux fréquences propres, les prédictions des deux modèles de régression linéaire sont significativement proches (cf. Figures C.1(a) et C.2(a)). De plus, elles se situent dans les bandes de confiance. Une analyse similaire est menée en considérant des modèles de régression non linéaire (cf. Figures C.1(b) et C.2(b)). Comme dans le cas du modèle linéaire, les prédictions sont assez proches et se retrouvent toutes à l'intérieur des intervalles de confiance évalués.

La Figure C.3 illustre l'évolution au cours du temps des trois premières fréquences propres corrigées par le modèle non linéaire. Comme dans le cas du modèle linéaire (cf. Figure 6.13), il est difficile d'apprécier les différences entre ces fréquences propres et celles sans correction thermique



(a) Comparatif des prédictions (modèles linéaires)



(b) Comparatif des prédictions (modèles non linéaires)

FIGURE C.1 – Comparatif des prédictions pour la deuxième fréquence propre évaluées par des modèles de régression linéaires et non linéaires.

à partir d'une observation simple sur leurs évolutions.

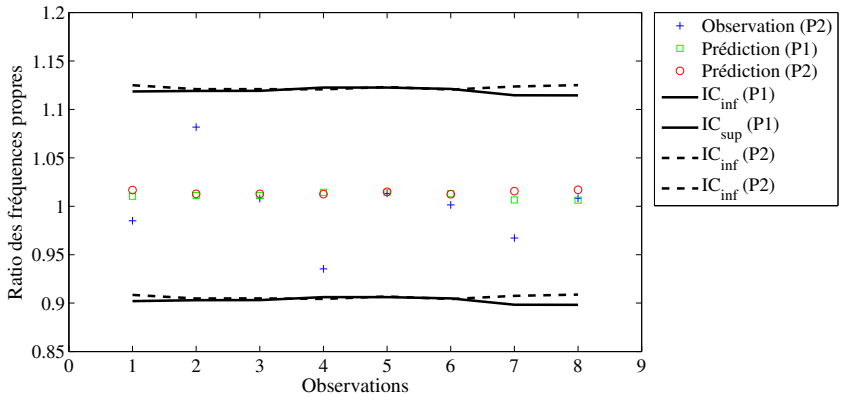
C.1.1 Étude des modèles de prédiction

Dans ce paragraphe, une approche originale basée sur des modèles de régression et l'ACP (Analyse en Composantes Principales) couplés à l'ADS est proposée. L'objectif principal est de pouvoir détecter des modifications qui peuvent représenter un comportement anormal de l'ouvrage, en éliminant les effets liés aux variations de température.

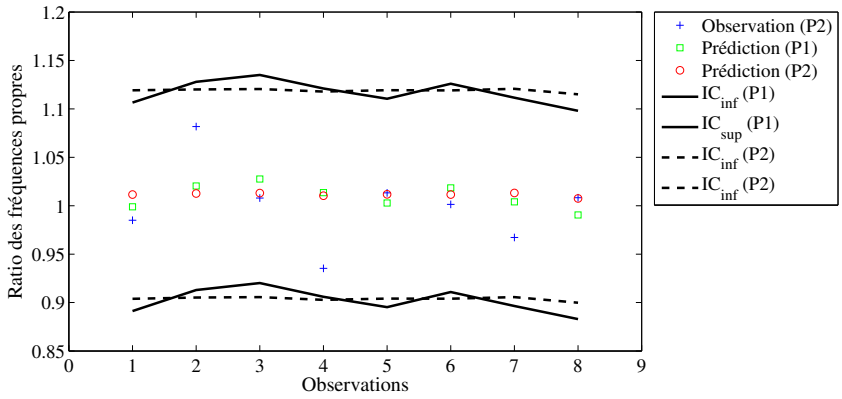
C.1.1.1 Affectation des essais en complément des modèles de régression linéaire

L'approche d'affectation des essais est d'abord considérée dans ce paragraphe en utilisant la première fréquence propre. Le Tableau C.1 regroupe les taux d'affectation calculés. Quel que soit le paramètre modal considéré ou la transformation symbolique utilisée, les essais ne sont, pour la plupart, pas affectés au cluster de la phase 1. Dans le cas des fréquences propres, la loi ajustée est la loi de Weibull (seuil égal à 3,79). Pour les modes propres, la loi ajustée est la loi log-logistique (seuil égal à 3,65).

L'affectation des essais est maintenant effectuée lorsque la deuxième fréquence/mode propre sont considérés. Les résultats sont détaillés dans le Tableau C.2. En considérant les fréquences propres transformées en intervalles interquartiles, un peu plus de la moitié des essais n'est pas

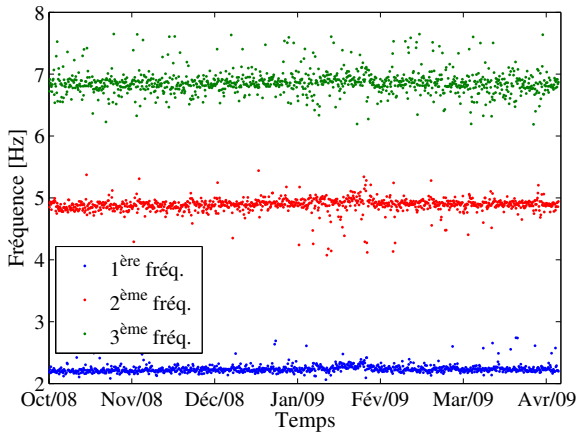


(a) Comparatif des prédictions (modèles linéaires)

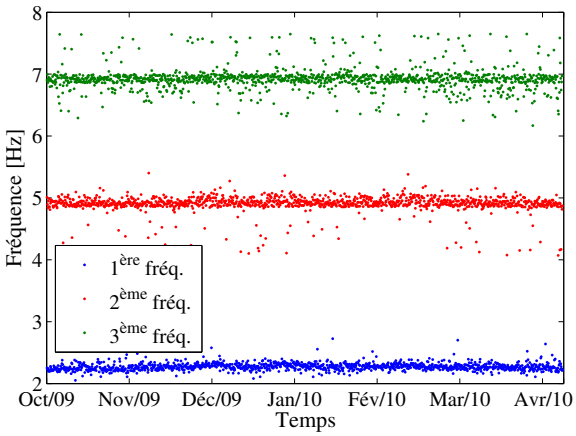


(b) Comparatif des prédictions (modèles non linéaires)

FIGURE C.2 – Comparatif des prédictions pour la troisième fréquence propre évaluées par des modèles de régression linéaires et non linéaires.



(a) Évolution des fréquences propres dans la phase 1



(b) Évolution des fréquences propres dans la phase 2

FIGURE C.3 – Évolution des trois premières fréquences propres en fonction du temps (régression non linéaire).

classé dans le cluster C(P1). Cela indique que le modèle construit pour la deuxième fréquence propre peut ne pas discriminer correctement les effets thermiques d'une modification structurale. En général, le couplage entre les modèles de régression et l'approche d'affectation ne permet pas

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	33%	28%
Non classé	67%	72%

TABLEAU C.1 – Affectation des essais en utilisant la première fréquence/mode propre transformés en intervalles (modèle linéaire).

de différencier ces effets. La loi ajustée est la loi log-normale (seuil égal à 2,63). Les résultats sont pratiquement du même ordre en considérant les déformées modales. Dans ce cas, la loi log-normale est ajustée aux distances entre essais (seuil égal à 1,35).

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	42%	49%
Non classé	58%	51%

TABLEAU C.2 – Affectation des essais en utilisant la deuxième fréquence/mode propre transformés en intervalles (modèle linéaire).

Les résultats de l'affectation des essais pour la troisième fréquence propre sont détaillées dans le Tableau C.3. Comme il a été observé pour la deuxième paire des paramètres modaux, pratiquement la moitié des essais n'est pas affectée au cluster de la phase 1. Cela confirme que le modèle de régression linéaire n'est pas capable de discriminer correctement les effets thermiques d'une modification structurale. La loi ajustée dans le cas de la transformation en intervalles interquartiles est la loi de Weibull (seuil égal à 4,46). Les résultats sont pratiquement du même ordre en considérant les déformées modales. Dans ce cas, la loi log-normale est ajustée aux distances entre essais (seuil égal à 2,95).

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	47%	56%
Non classé	53%	44%

TABLEAU C.3 – Affectation des essais en utilisant la troisième fréquence/mode propre transformés en intervalles (modèle linéaire).

C.1.1.2 Affectation des essais en complément des modèles de régression non linéaire

La démarche pour l'affectation des essais est considérée pour la première fréquence propre. Le Tableau C.4 regroupe les pourcentages d'affectation calculés. Comme il a été observé au paragraphe C.1.1.1, la plupart des essais ne sont pas affectés à C(P1). Ceci implique que le modèle de régression n'est pas suffisant pour discriminer les effets des travaux. En appliquant cette démarche aux modes propres, les résultats sont relativement similaires. Toutefois, il reste toujours des essais qui ne sont pas distingués par l'approche d'affectation. Étant donné que le cluster C(P1) est toujours le même, les lois de probabilités ajustées (et les valeurs seuils associées) sont indiques à celles déterminées pour le modèle linéaire.

Les taux d'affectation évalués pour chacun des clusters pour la deuxième fréquence/mode propre sont détaillés dans le Tableau C.5. Un peu plus de la moitié des essais est affecté au cluster qui représente un nouvel état (i.e., différent de celui considéré comme l'état avant les travaux de renforcement). Cela montre que les modèles de régression non linéaire appliqués à la deuxième fréquence propre peuvent mieux caractériser les modifications structurales par rapport à ceux construits en utilisant la première fréquence. Dans le cas des modes propres, près de 60%

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	38%	31%
Non classé	62%	69%

TABLEAU C.4 – Affectation des essais en utilisant la première fréquence/mode propre transformés en intervalles (modèle non linéaire).

des essais sont classés ailleurs, en indiquant que l'approche d'affectation sert de complément aux modèles de régression. Toutefois, il reste encore un nombre important d'essais pour lesquels les effets thermiques ne peuvent pas être distingués des ceux causés par les travaux.

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	44%	40%
Non classé	56%	60%

TABLEAU C.5 – Affectation des essais en utilisant la deuxième fréquence/mode propre transformés en intervalles (modèle non linéaire).

Le Tableau C.6 regroupe les taux d'affectation évalués pour chacun des clusters (C(P1) et celui qui représente un nouvel état) lorsque la troisième fréquence/mode propre est considéré. Contrairement à ce qui été observé dans le cas du modèle linéaire, un peu plus de la moitié des essais est affecté dans le cluster de la phase 1. Cela montre que l'utilisation de l'approche d'affectation n'apporte pas une meilleure discrimination des effets thermiques. Ceci est également observé en considérant les modes propres.

Cluster	**Fréquences propres**	**Modes propres**
C(P1)	53%	52%
Non classé	47%	48%

TABLEAU C.6 – Affectation des essais en utilisant la troisième fréquence/mode propre transformés en intervalles (modèle non linéaire).

C.1.1.3 Affectation des essais en complément de l'Analyse en Composantes Principales

Les simulations suivantes considèrent l'approche basée sur l'ACP (Analyse en Composantes Principales) introduite au paragraphe 3.5. Ici, les trois premières fréquences propres identifiées lors des phases 1 et 2 d'instrumentation sont considérées pour le calcul de l'indice d'endommagement κ.

Comme il a été considéré pour les techniques de régression, l'approche d'affectation d'essais est maintenant considérée. Ici, trois clusters sont utilisés : celui correspondant à la phase 1 (avant), à la phase 2 (après) et finalement celui correspondant à un nouvel état. Dans un premier temps, les 71 essais considérés comme anormaux pendant la phase 1 sont analysés. En cela, ils sont affectés soit au cluster de la phase 1 (C(P1)) soit dans un autre groupe. Dans un deuxième temps, la même procédure est appliquée aux 129 essais de la phase 2. Cette démarche est aussi mise en œuvre en utilisant les résultats obtenus par l'ACP sur 2 composantes principales. Le Tableau C.7 regroupe les taux d'affectation calculés. Il est noté qu'un grand nombre d'essais sont affectés à leur cluster original. Ceci montre que la démarche basée sur l'ACP suscite un nombre important de fausses alarmes (cf. Figure 6.26). Pour toutes les simulations, pratiquement 90% des essais ne sont pas considérés comme anormaux, d'après l'approche d'affectation des nouveaux essais.

Cluster	**1 CP**	**2 CP**
C(P1)	92%	95%
Non classé	8%	5%
C(P2)	85%	88%
Non classé	15%	12%

TABLEAU C.7 – Affectation des essais en utilisant les fréquences propres transformées en intervalles interquartiles (norme Euclidienne).

Le Tableau C.8 regroupe les résultats de l'approche d'affectation en considérant la norme de Mahalanobis. Un nombre important d'essais n'est pas considéré comme étant anormal. De plus, ceux qui l'ont été sont essentiellement les mêmes que ceux déjà signalés par les autres méthodes. De façon générale, l'utilisation de l'ACP peut mener à l'occurrence de fausses alarmes.

Cluster	**1 CP**	**2 CP**
C(P1)	84%	89%
Non classé	16%	11%
C(P2)	90%	94%
Non classé	10%	6%

TABLEAU C.8 – Affectation des essais en utilisant les fréquences propres transformées en intervalles interquartiles (norme Mahalanobis).

C.2 Analyses classiques

La Figure C.4 montre les histogrammes des trois premières fréquences propres corrigées par le filtre non linéaire, en comparant les phases 1 et 2. Il est noté que tous les intervalles de confiance se chevauchent, ce qui ne permet pas de distinguer clairement les deux états structuraux.

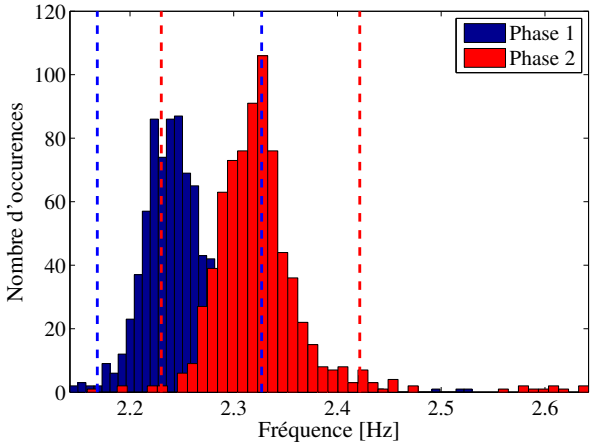
Les déformées propres corrigées par le filtre non linéaire (et les coefficients MAC associés) sont illustrées dans la Figure C.5.

C.3 ADS couplée aux méthodes de clustering

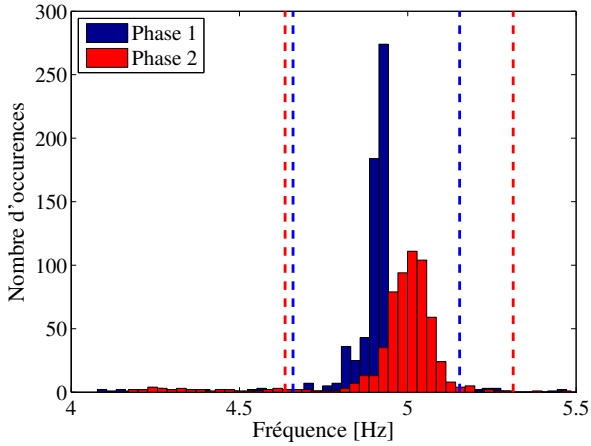
De façon similaire à celle proposée dans les chapitres 4 et 5, l'ADS est couplée aux méthodes de partitionnement de données pour la discrimination des deux états structuraux du pont PI-57. L'objectif de cette étude est d'appliquer les approches proposées dans le cadre de cette thèse pour discriminer les essais enregistrés dans chacune des phases en deux groupes distincts. Les méthodes de clustering sont appliquées aux signaux et aux trois premiers des paramètres modaux transformés en intervalles interquartiles.

C.3.1 Analyse sur l'ensemble des essais

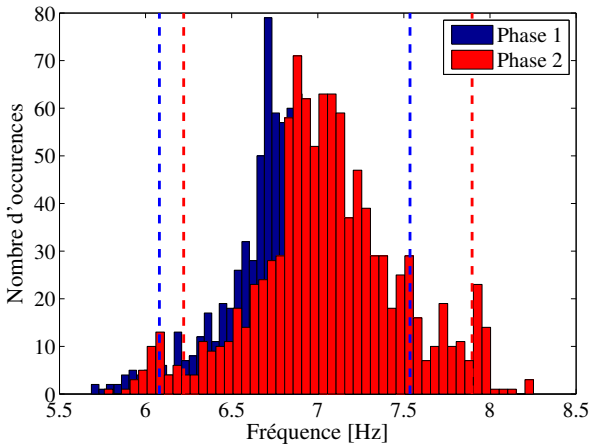
Comme il a été décrit précédemment, deux campagnes de mesures ont été réalisées. La première, entre le 14 octobre 2008 et le 4 avril 2009, a permis d'enregistrer un total de 1174 fichiers. La deuxième a eu lieu pendant la même période, durant les années de 2009 et 2010. Pour celle-ci 1316 fichiers de mesure ont été acquis. Dans le paragraphes suivants, les analyses de clustering considèrent l'ensemble des 2490 essais.



(a) Première fréquence propre



(b) Deuxième fréquence propre



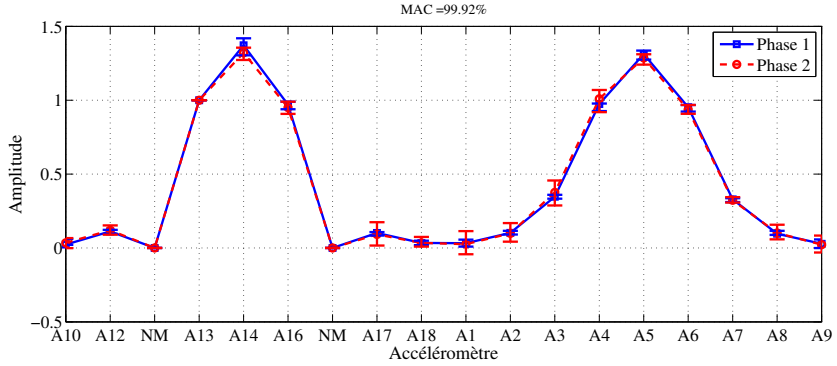
(c) Troisième fréquence propre

FIGURE C.4 – Comparatif des histogrammes pour les trois premières fréquences propres corrigées par le filtre non linéaire (phases 1 et 2).

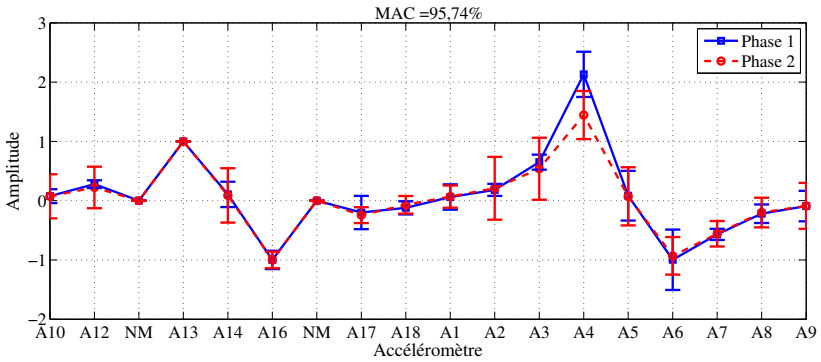
C.3.1.1 Signaux

La méthode des divisions hiérarchiques est d'abord appliquée aux signaux transformés en intervalles interquartiles. Il est évident que le partitionnement optimal est celui capable de séparer l'ensemble des essais en deux groupes parfaitement distincts, chacun ne contenant que des essais du même type. Étant donné le nombre conséquent d'essais utilisés dans cette simulation, les sorties graphiques ne sont pas présentées. Dans cette analyse, 56% des essais sont classés correctement dans le cluster correspondant à la première phase, et 54% dans le cluster de la phase 2. La variable la plus discriminante en considérant des intervalles interquartiles est le capteur 9, situé proche de l'appui. Il est intéressant de noter que c'est un capteur qui enregistre des accélérations relativement faibles (contrairement de ceux situés à mi-travée, par exemple). Cela montre que le critère de coupure des groupes n'est pas nécessairement lié à l'amplitude des vibrations dans cet exemple.

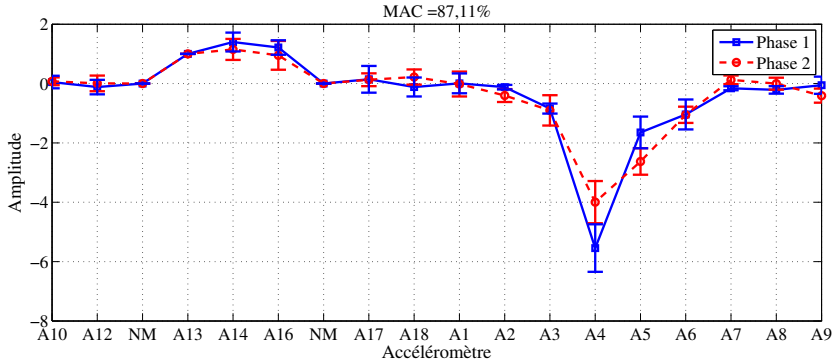
Par ailleurs, la méthode des agglomérations hiérarchiques est considérée. Comme déjà remarqué, cette méthode tend à mélanger plus fréquemment des états structuraux, menant à des taux de classification correcte moins importants par rapport aux autres méthodes. Dans cette analyse, et en tenant compte de la transformation en intervalles, seuls 42% des essais sont correctement



(a) Premier mode propre



(b) Deuxième mode propre



(c) Troisième mode propre

FIGURE C.5 – Comparatif des trois premiers modes propres corrigés par le non filtre linéaire et calcul de MAC (phases 1 et 2).

classés dans le groupe correspondant à la première phase d'acquisition et 41% dans celui de la deuxième phase.

L'analyse des partitionnements obtenus par la méthode des nuées dynamiques est maintenant considérée. Comme pour les divisions hiérarchiques, les taux de classification correcte sont légèrement plus importants : 58% pour le cluster de la phase 1 et 55% pour le cluster de la phase 2.

C.3.1.2 Fréquences propres

Les résultats suivants considèrent l'application des méthodes de clustering aux trois premières fréquences propres transformées en données symboliques. Dans un premier temps, la méthode des divisions hiérarchiques est utilisée. Dans le cas de la transformation en intervalles, les taux sont relativement faibles : 75% pour l'état avant et 72% pour l'état après. La variable discriminante est la deuxième fréquence propre.

Dans le cas des agglomérations hiérarchiques, les résultats sont, en général, moins pertinents. En considérant des intervalles, seuls 65% des essais sont correctement classés dans le cluster de la phase 1, tandis que 68% le sont pour la phase 2.

Par ailleurs, la méthode basée sur les nuées dynamiques est considérée. Ici, les taux de classification sont plus élevés que ceux avec la méthode des agglomérations hiérarchiques, mais du même ordre que ceux obtenus avec les divisions hiérarchiques. En utilisant la transformation en intervalles interquartiles, les pourcentages de classification correcte sont 75% pour le premier groupe (phase 1) et 76% pour le deuxième (phase 2).

C.3.1.3 Modes propres

La méthode des divisions hiérarchiques est maintenant appliquée aux trois premiers modes propres. Les résultats sont en général meilleurs que ceux obtenus en utilisant les signaux, mais moins pertinents que ceux obtenus avec les fréquences propres. En effet, cette tendance a été vérifiée lors des applications numérique et expérimentale (pont-rail au chapitre 5). Dans ce cas, en utilisant des intervalles, 69% des essais sont correctement classés dans le groupe de la phase 1 et 68% dans celui de la phase 2. La variable la plus discriminante est le deuxième mode propre.

Les agglomérations hiérarchiques sont maintenant considérées. Comme déjà remarqué, les taux sont généralement plus faibles que ceux avec les autres méthodes de partitionnement. Ici, les pourcentages de classification correcte sont de 60% dans le cluster de la phase 1 et 63% dans celui de la phase 2.

La méthode des nuées dynamiques est appliquée aux modes propres. Dans le cas de la transformation en intervalles interquartiles, les taux sont plus élevés que ceux obtenus avec les méthodes précédentes. Ici, 72% des essais correspondant à la phase 1 d'acquisition sont correctement classés et 73% le sont dans le cluster de l'état « après » renforcement.

C.3.2 Application aux paramètres modaux corrigés

Comme dans l'application expérimentale sur le pont-rail au chapitre 5, la deuxième partie de l'étude pour le pont sur l'Oise consiste à appliquer les trois premières paires de paramètres modaux « compensés » des effets thermiques aux méthodes de clustering. Dans un premier temps, les analyses de clustering considèrent l'ensemble de tous les essais. Une remarque importante est qu'il n'y pas eu d'enregistrement de température pendant le mois d'octobre 2008. Les analyses suivantes ne considèrent donc que les paramètres modaux identifiés entre les mois de novembre et avril (pour les deux phases d'instrumentation). En cela, seuls les fréquences et modes propres identifiés pendant cette période sont utilisés. En premier lieu les paramètres modaux sont corrigés par les filtres linéaire et non linéaire. En second lieu, les paramètres corrigés sont appliqués aux méthodes de clustering. Les analyses suivantes considèrent toujours les trois premières fréquences propres et déformées modales associées.

C.3.2.1 Analyse sur l'ensemble des essais

Les taux de classification correcte obtenus par les divisions hiérarchiques appliquées aux paramètres modaux sont détaillés dans le Tableau C.9. Dans le cas des fréquences propres, il est noté que les résultats sont relativement moins pertinents que ceux obtenus sans la correction thermique. Les taux obtenus en utilisant les modes propres sont pratiquement similaires, avec toutefois une légère augmentation des taux de classification correcte. Les résultats montrent que la correction des paramètres modaux est, en effet, pertinente, étant donné que les taux de classification sont, en général, plus importants. Dans cette analyse, la variable discriminante est la première fréquence propre et le deuxième mode propre.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	74%	71%	75%	69%
C(P2)	69%	68%	72%	68%

TABLEAU C.9 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (intervalles ; filtre linéaire).

Le Tableau C.10 regroupe les résultats obtenus avec les agglomérations hiérarchiques lorsque les fréquences et les modes propres sont utilisés. Des légères augmentations des taux de classification correcte sont remarquées. Dans le cas des fréquences propres et en considérant des intervalles, 69% des essais sont correctement classés dans le cluster de la phase 1 (contre 65% de l'analyse précédente) et 66% le sont pour la phase 2 (contre 68%). Dans le cas des modes propres, 62% des essais sont correctement classés dans chacun des groupes.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	69%	62%	65%	60%
C(P2)	66%	62%	68%	63%

TABLEAU C.10 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (intervalles ; filtre linéaire).

De plus, la méthode basée sur les nuées dynamiques est appliquée aux paramètres modaux. Ici, l'augmentation des taux de classification est également observée dans les deux cas. Pour les fréquences propres transformées en intervalles interquartiles, par exemple, 77% des essais sont correctement classés dans le cluster de la phase 1 (contre 75% sans prise en compte des effets thermiques). Pour les modes propres, les résultats sont significativement plus pertinents (73% pour la phase 1 et 75% pour la phase 2, contre 72% et 73%, respectivement). Le Tableau C.11 regroupe la totalité des résultats.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	77%	73%	75%	72%
C(P2)	72%	75%	76%	73%

TABLEAU C.11 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (intervalles ; filtre linéaire).

La correction des paramètres modaux par le filtre non linéaire est maintenant considérée. Le Tableau C.12 regroupe les taux de classification obtenus en utilisant la méthode des divisions hiérarchiques. Dans le cas des fréquences propres, les pourcentages sont significativement plus importants que ceux obtenus avec le filtre linéaire. Cela montre que l'utilisation du filtre non linéaire est plus pertinente et que la prise en compte des corrections thermiques est essentielle dans cette étude. Les résultats obtenus en utilisant les déformées modales sont relativement meilleurs que les précédents (cf. Tableau C.11), bien que la différence soit ici plus faible. Dans cette analyse, la première fréquence propre et le deuxième mode propre sont les variables discriminantes en utilisant des intervalles interquartiles.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	75%	70%	75%	69%
C(P2)	72%	70%	72%	68%

TABLEAU C.12 – Taux de classification correcte obtenus en utilisant les divisions hiérarchiques appliquées aux paramètres modaux (intervalles ; filtre non linéaire).

La méthode des agglomérations hiérarchiques est maintenant appliquée aux paramètres modaux corrigés par le filtre non linéaire (cf. Tableau C.13). Comme observé avec la méthode précédente, les pourcentages de classification correcte ont aussi changé. Les classifications obtenues ici sont généralement meilleures que celles obtenues sans la considération des effets thermiques.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	71%	65%	65%	60%
C(P2)	64%	60%	68%	63%

TABLEAU C.13 – Taux de classification correcte obtenus en utilisant les agglomérations hiérarchiques appliquées aux paramètres modaux (intervalles ; filtre non linéaire).

Enfin, la méthode des nuées dynamiques est considérée. Le Tableau C.14 regroupe les taux de classification correcte. Les pourcentages sont sensiblement plus importants que ceux obtenus avec le filtre linéaire (ou encore sans l'utilisation des filtres). Dans le cas des fréquences propres, près de 80% des essais sont correctement classés. Par ailleurs, les pourcentages obtenus en utilisant les déformées modales sont légèrement moins élevés.

	Avec correction		**Sans correction**	
	Fréquences propres	**Modes propres**	**Fréquences propres**	**Modes propres**
C(P1)	79%	70%	75%	72%
C(P2)	71%	74%	76%	73%

TABLEAU C.14 – Taux de classification correcte obtenus en utilisant les nuées dynamiques appliquées aux paramètres modaux (intervalles ; filtre non linéaire).

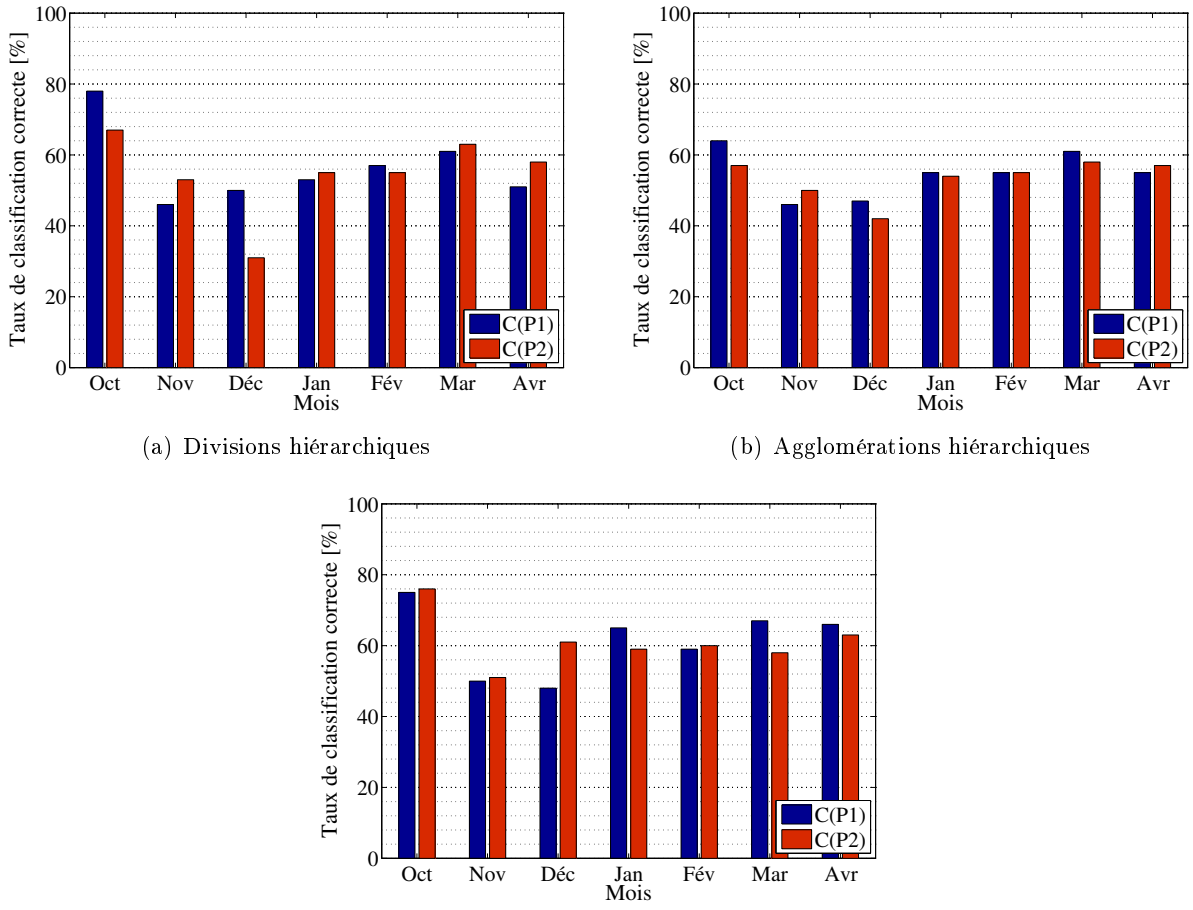
C.3.3 Comparaison des phases 1 et 2 mois par mois

Une analyse plus fine est proposée dans ce paragraphe. Au lieu de faire une étude globale où tous les essais des phases 1 et 2 sont pris en compte en même temps, un traitement de données

mois par mois est proposé.

C.3.3.1 Signaux

Les résultats obtenus en utilisant les méthodes de partitionnement appliquées aux signaux sont montrés à la La Figure C.6. Les barres verticales représentent les taux de classification correcte évalués pour chaque mois. Les pourcentages de classification correcte ont augmenté sensiblement par rapport aux résultats obtenus dans le paragraphe précédent. Cela met en évidence que la procédure utilisée auparavant, où plusieurs essais (enregistrés dans différentes périodes de l'année) sont considérés de manière simultanée, peut ne pas être la plus pertinente pour la détection des modifications structurales.



(a) Divisions hiérarchiques

(b) Agglomérations hiérarchiques

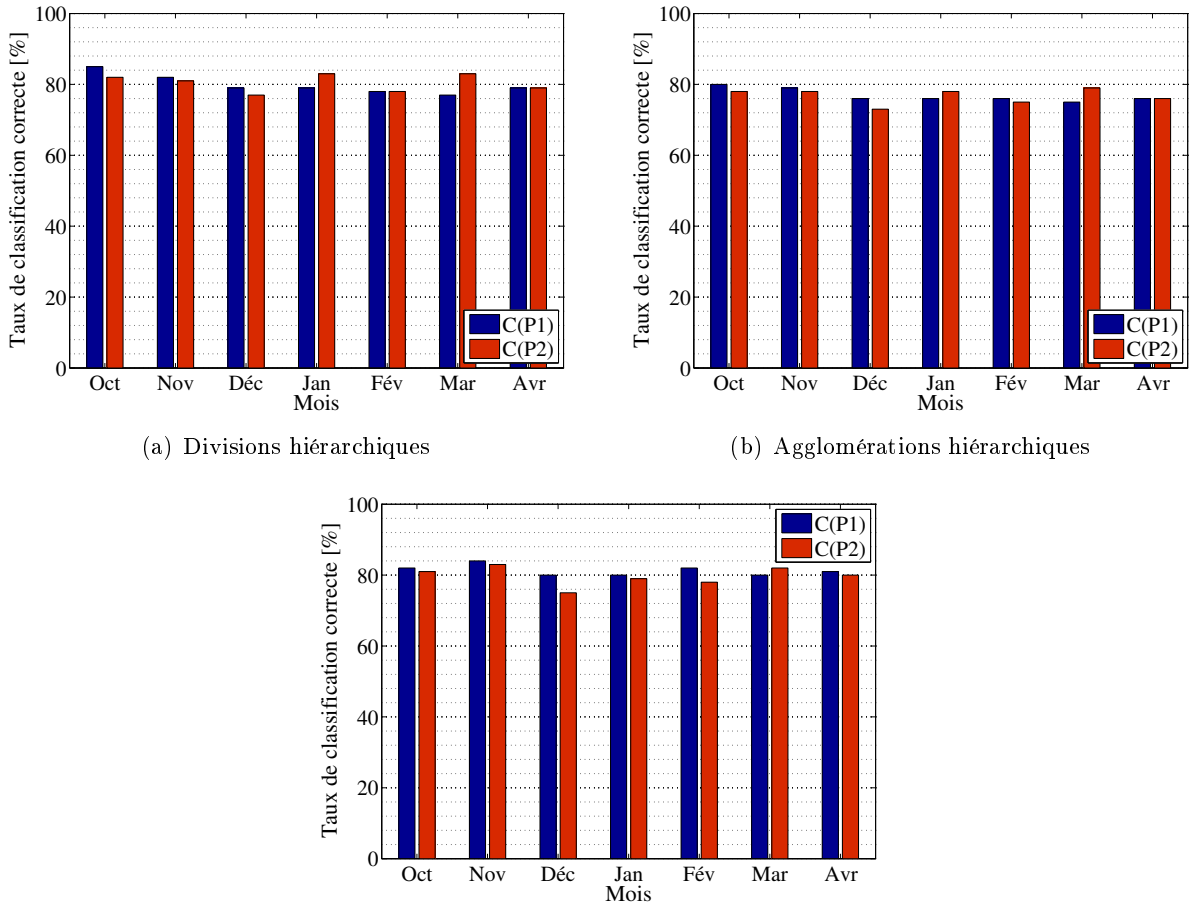
(c) Nuées dynamiques

FIGURE C.6 – Méthodes de partitionnement appliquées aux signaux transformés en intervalles interquartiles.

En général, les méthodes des divisions hiérarchiques et des nuées dynamiques sont celles capables de classer les essais avec les meilleurs taux de réussite, encore qu'ils soient relativement faibles. Par ailleurs, le Tableau C.15 regroupe les variables discriminantes déterminées dans chaque analyse en considérant les divisions hiérarchiques.

C.3.3.2 Fréquences propres

La Figure C.7 présente les résultats obtenus en utilisant les trois premières fréquences propres. Les taux de classification correcte sont plus élevés que ceux obtenus en utilisant les signaux. Les méthodes des divisions hiérarchiques et des nuées dynamiques permettent d'obtenir les meilleurs taux de classification correcte. Le Tableau C.15 regroupe les variables discriminantes déterminées dans chaque analyse en considérant les divisions hiérarchiques. Dans la plupart des cas, la première fréquence propre a été choisie comme celle la plus discriminante.



(a) Divisions hiérarchiques

(b) Agglomérations hiérarchiques

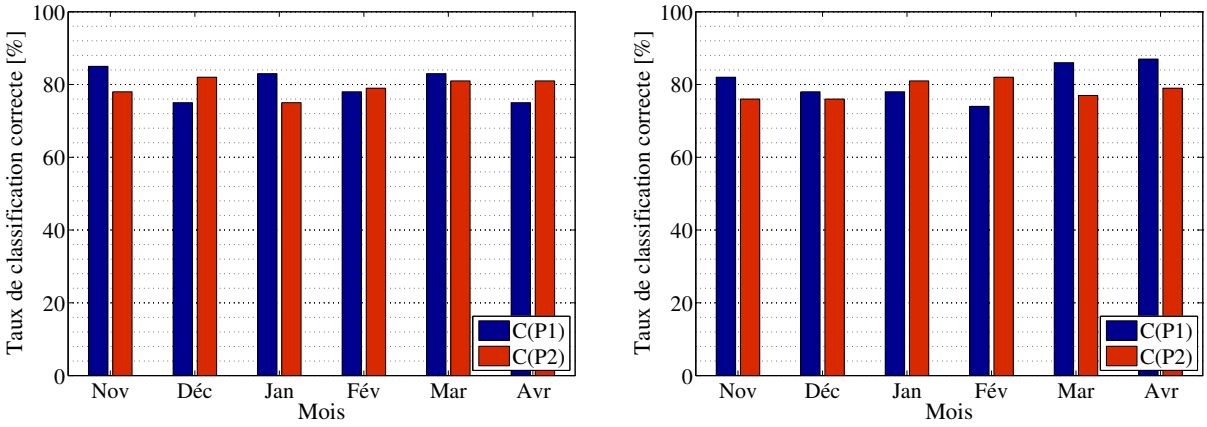
(c) Nuées dynamiques

FIGURE C.7 – Méthodes de partitionnement appliquées aux fréquences propres transformées en intervalles interquartiles.

La démarche proposée pour le filtrage des paramètres modaux est considérée et les procédés de clustering sont à nouveau mis au point. L'objectif est de vérifier si l'utilisation des filtres thermiques peut améliorer les taux de classification obtenus auparavant.

La Figure C.8 illustre le comparatif entre la méthode des divisions hiérarchiques appliquées aux fréquences propres corrigées par le filtre linéaire et non linéaire. Il est noté que dans la plupart des mois, les taux de classification correcte sont plus importants en utilisant ce dernier (cf. Figure C.7(a)).

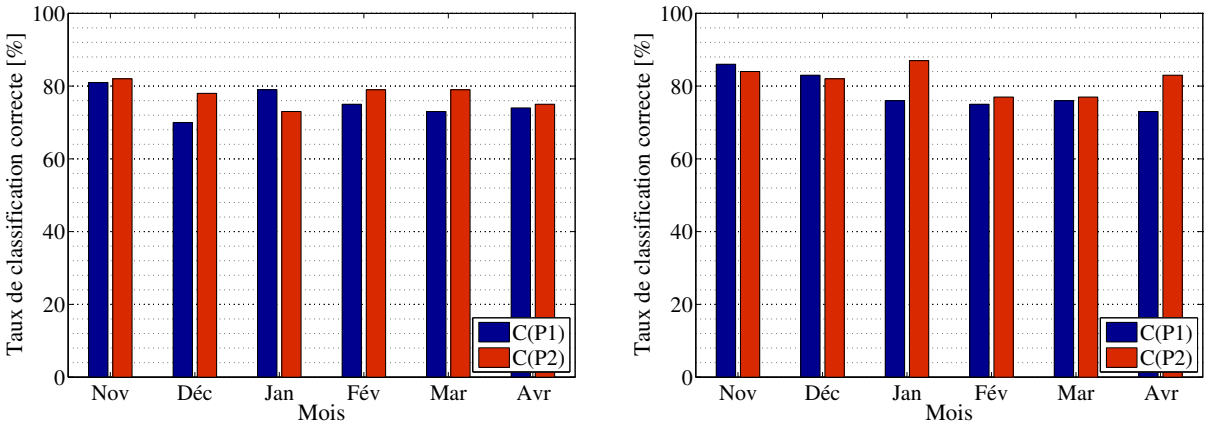
La méthode des agglomérations hiérarchiques est aussi considérée. La Figure C.9 illustre les résultats obtenus. Comme il a été observé dans l'analyse de l'ensemble des essais, les taux de classification correcte sont légèrement plus importants en utilisant le filtre linéaire. Si le filtre non linéaire est pris en compte, les résultats sont significativement meilleurs. Dans ce cas, le taux



(a) Filtre linéaire (b) Filtre non linéaire

FIGURE C.8 – Méthode des divisions hiérarchiques appliquées aux fréquences propres « corrigées » (filtres linéaire et non linéaire)

moyen pour tous les mois est de l'ordre de 77% (contre 75% pour le filtre linéaire).



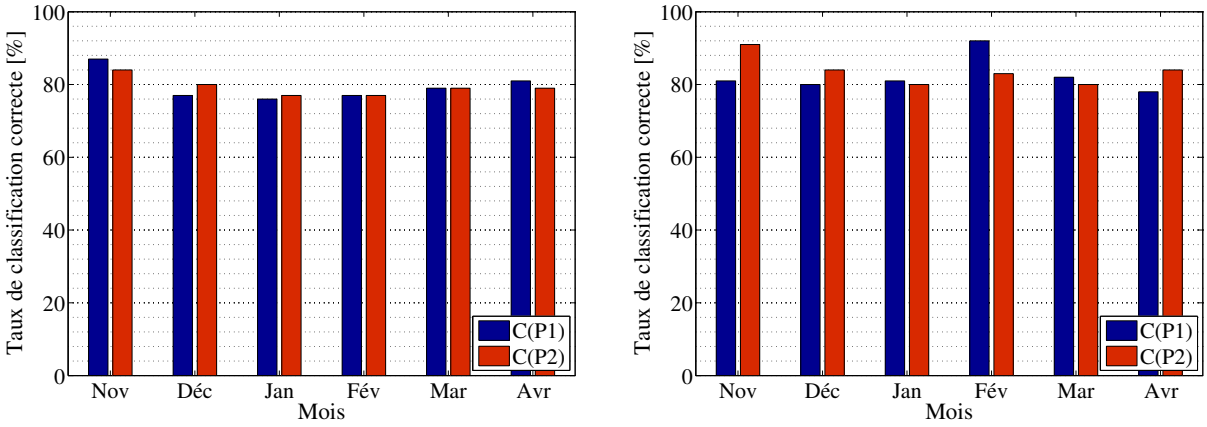
(a) Filtre linéaire (b) Filtre non linéaire

FIGURE C.9 – Méthode des agglomérations hiérarchiques appliquées aux fréquences propres corrigées (filtres linéaire et non linéaire)

Par ailleurs, les classifications obtenues en utilisant les nuées dynamiques sont présentées à la Figure C.10. Les résultats sont significativement meilleurs en considérant les deux filtres thermiques.

C.3.3.3 Modes propres

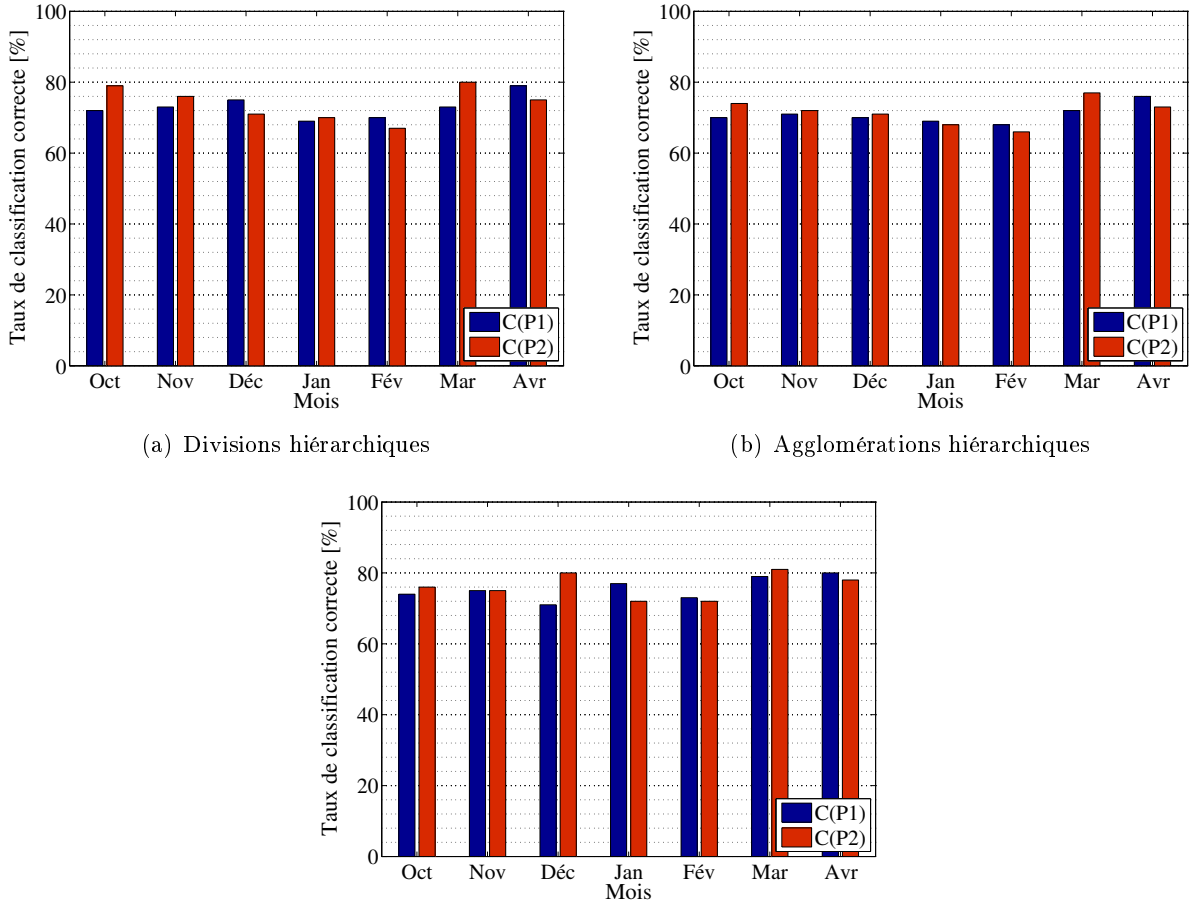
Les résultats obtenus en utilisant les trois premiers modes propres sont montrés à la Figure C.11. Des taux relativement plus élevés de classification correcte sont également observés dans ce cas (par rapport à l'analyse de l'ensemble des essais). En général, les méthodes des divisions hiérarchiques et des nuées dynamiques sont celles capables de classer les essais avec les meilleurs taux de réussite. Le Tableau 6.20 regroupe les variables discriminantes déterminées dans chaque analyse en considérant les divisions hiérarchiques. Il est noté que le deuxième mode propre est celui le plus souvent utilisé pour la discrimination des clusters.



(a) Filtre linéaire

(b) Filtre non linéaire

FIGURE C.10 – Méthode des nuées dynamiques appliquées aux fréquences propres corrigées (filtres linéaire et non linéaire)



(a) Divisions hiérarchiques

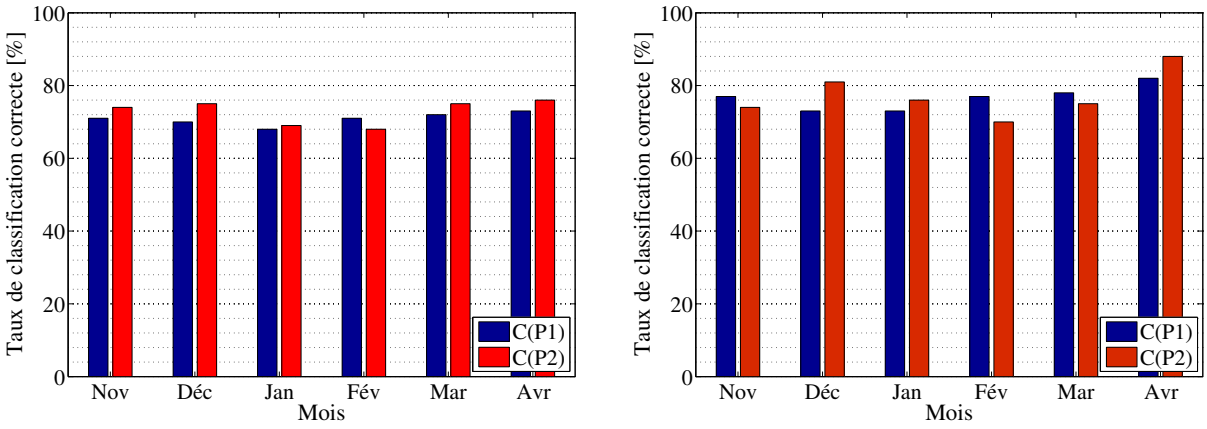
(b) Agglomérations hiérarchiques

(c) Nuées dynamiques

FIGURE C.11 – Méthodes de partitionnement appliquées aux modes propres transformés en intervalles interquartiles.

La procédure de compensation thermique est maintenant appliquée aux déformées modales. La Figure C.12 montre les pourcentages évalués pour les deux filtres. Les taux de classification

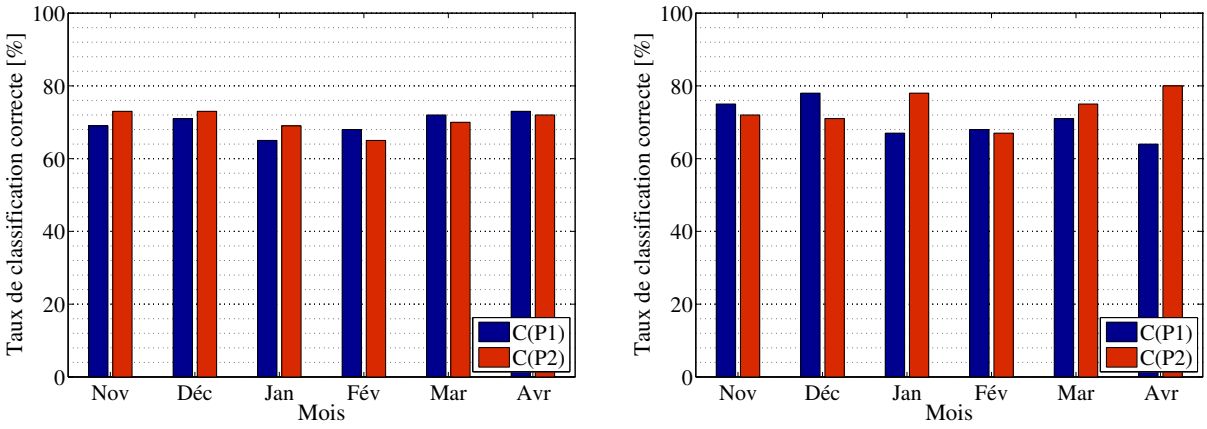
correcte sont relativement plus élevés que ceux obtenus sans prise en compte de la variation thermique. En général, le filtre non linéaire a produit des meilleurs résultats.



(a) Filtre linéaire (b) Filtre non linéaire

FIGURE C.12 – Méthode des divisions hiérarchiques appliquées aux modes propres « corrigés » (filtres linéaire et non linéaire)

Les classifications obtenues en utilisant la méthode des agglomérations hiérarchiques sont montrées à la Figure C.13. Les taux de classification correcte obtenus par le filtre linéaire sont du même ordre que ceux évalués sans aucun filtre. Par ailleurs, l'utilisation d'un filtre non linéaire mène à des pourcentages relativement plus importants, ce qui montre la pertinence de l'utilisation de cette technique.



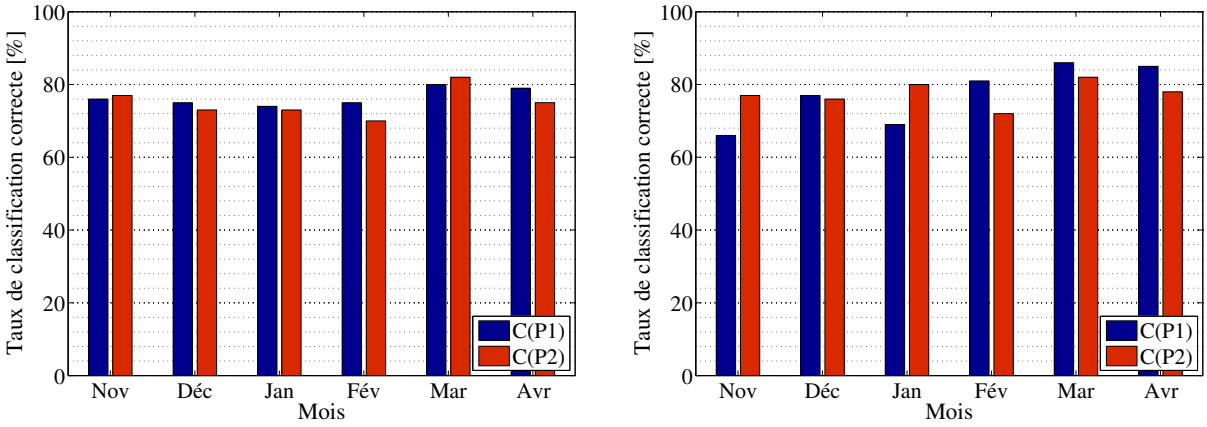
(a) Filtre linéaire (b) Filtre non linéaire

FIGURE C.13 – Méthode des agglomérations hiérarchiques appliquées aux modes propres corrigés (filtres linéaire et non linéaire)

De plus, la Figure C.14 illustre les résultats obtenus par les nuées dynamiques. En général, les taux évalués en utilisant le filtre non linéaire sont plus importants que ceux avec le filtre linéaire.

C.3.3.4 Variables discriminantes pour la méthode des divisions hiérarchiques

Les variables les plus discriminantes pour la méthode des divisions hiérarchiques sont regroupées dans le Tableau C.15.



(a) Filtre linéaire

(b) Filtre non linéaire

FIGURE C.14 – Méthode des nuées dynamiques appliquées aux modes propres corrigés (filtres linéaire et non linéaire)

	Variable discriminante		
	Capteur	Fréquence propre	Mode propre
Oct.	A2	1	2
Nov.	A10	1	2
Déc.	A9	2	1
Jan.	A12	2	2
Fév.	A18	2	1
Mar.	A10	1	2
Avr.	A3	3	2

TABLEAU C.15 – Variables discriminantes déterminées pour chaque mois analysé (intervalles interquartiles).

Les variables les plus discriminantes pour la méthode des divisions hiérarchiques sont regroupées dans le Tableau C.16. Dans ce tableau, les résultats pour les deux types de régression sont donnés.

	Variables discriminantes			
	Régression linéaire		Régression non linéaire	
	Fréquence propre	Mode propre	Fréquence propre	Mode propre
Oct.	2	2	2	2
Nov.	1	2	1	2
Déc.	2	1	1	2
Jan.	1	1	1	1
Fév.	2	2	1	2
Mar.	1	1	1	1
Avr.	3	3	3	1

TABLEAU C.16 – Variables discriminantes déterminées pour chaque mois analysé (intervalles interquartiles).

C.3.4 Indices de partitionnement optimal

Dans ce paragraphe, les indices introduits dans le paragraphe 3.1.5 sont employés afin de déterminer le nombre de clusters caractérisant le partitionnement optimal. Le nombre maximal de clusters est défini comme étant égal a 5.

C.3.4.1 Analyse sur l'ensemble des essais

Signaux

Les indices sont d'abord calculés en utilisant les deux méthodes de clustering appliquées aux signaux, transformés en intervalles interquartiles (cf. Tableau C.17). Les indices CH et Γ indiquent toujours le nombre de clusters attendu (i.e., deux). De plus, l'indice C^* n'indique jamais le nombre d'états structuraux théoriquement existants. Comme déjà remarqué, cet indice signale systématiquement un nombre de clusters toujours plus élevé que dans la réalité.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	377	**0,113**	0,66	276	**0,088**	0,67
4	456	0,134	0,64	313	0,092	0,67
3	605	0,159	0,66	329	0,112	0,64
2	**1028**	0,144	**0,77**	**436**	0,125	**0,71**

TABLEAU C.17 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés en intervalles (ensemble total des essais).

Fréquences propres

Les fréquences propres sont maintenant utilisées. L'analyse du Tableau C.18 permet de vérifier que les indices CH et Γ présentent un bon accord entre les deux méthodes de clustering, indiquant toujours deux clusters comme étant le partitionnement optimal. L'indice C^* oscille cependant entre 4 et 5 clusters.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	*C**	Γ	*CH*	*C**	Γ
5	22,4	**0,074**	0,61	35,7	0,028	0,6
4	28,3	0,086	0,62	36,6	**0,024**	0,64
3	30,8	0,125	0,63	47,4	0,091	0,71
2	**35,3**	0,214	**0,72**	**58,0**	0,195	**0,78**

TABLEAU C.18 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en intervalles (ensemble total des essais).

Pour compléter cette étude, les indices de partitionnement optimal sont calculés en considérant les clusters obtenus à partir de chacune des méthodes de clustering appliquées aux paramètres modaux corrigés. De façon similaire à celle présentée, le nombre maximal de clusters est défini comme étant égal a 5. L'objectif est d'observer si les indices indiquent, même après la correction des effets thermiques, l'existence de deux états structuraux : avant et après les travaux de renforcement.

Dans un premier temps, les indices sont évalués pour les méthodes des divisions hiérarchiques et des nuées dynamiques appliquées aux fréquences propres corrigées par le filtre linéaire (cf. Tableau C.19). Les indices CH et Γ indiquent toujours le même nombre de clusters (deux).

L'indice C^*, toutefois, signale cinq groupes. L'indice Γ a une variation très faible dans le cas des divisions hiérarchiques, ce qui ne facilite pas son interprétation.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	*CH*	C^*	Γ
5	41,7	**0,023**	0,43	36,8	**0,045**	0,45
4	46,8	0,046	0,60	36,3	0,048	0,46
3	55,7	0,097	0,61	43,9	0,074	0,54
2	**61,3**	0,084	**0,62**	**51,2**	0,084	**0,63**

TABLEAU C.19 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en intervalles (filtre linéaire).

Dans un deuxième temps, les indices sont calculés pour les partitionnement obtenus en utilisant le filtre non linéaire. Les résultats sont identiques à ceux obtenus sans filtre thermique (cf. Tableau C.20).

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	C^*	Γ	*CH*	C^*	Γ
5	58,5	**0,037**	0,33	27,4	**0,041**	0,57
4	53,6	0,05	0,47	17,1	0,045	0,67
3	60,4	0,13	0,46	23,1	0,087	0,59
2	**65,8**	0,135	**0,49**	**29,7**	0,106	**0,70**

TABLEAU C.20 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées en intervalles (filtre non linéaire).

Modes propres

Le Tableau C.21) regroupe les indices calculés les partitionnements obtenus en utilisant les modes propres. En général, l'indice *CH* est celui qui présente le meilleur accord dans toutes les simulations réalisées. De plus, l'indice Γ n'indique deux groupes que pour la méthode des nuées dynamiques. Ceci peut être justifié par les meilleurs taux de classification correcte obtenus par cette méthode vis-à-vis de ceux évalués pour la méthode des divisions hiérarchiques. Par ailleurs, l'indice C^* signale deux clusters uniquement pour la méthode des nuées dynamiques. Pour les autres simulations, cet indice indique 5 groupes.

	Divisions hiérarchiques			**Nuées dynamiques**		
	CH	$C^{*\dagger}$	Γ	*CH*	$C^{*\dagger}$	Γ
5	136,6	**6,8**	0,69	147,8	11,6	0,72
4	103	12,8	0,71	187	6,2	0,76
3	131,9	17,8	**0,76**	204,3	6,1	0,79
2	**145,0**	32,2	0,74	**211,5**	**4,7**	**0,83**

TABLEAU C.21 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en intervalles (ensemble total des essais). † Les valeurs de l'indice C^* sont multipliées par $(\times 10^{-2})$.

De plus, les partitionnements obtenus en utilisant les modes propres corrigés sont également considérés. Le Tableau C.22 regroupe les indices calculés pour les méthodes des divisions hiérarchiques et des nuées dynamiques en considérant le filtre linéaire. Dans ce cas, quelques différences par rapport aux indices évalués sans le filtre sont observées. L'indice C^* n'indique jamais deux clusters comme étant le partitionnement optimal, d'ailleurs, cet indice oscillant entre quatre et cinq groupes.

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	35,8	**0,048**	0,62	40,5	0,051	0,59
4	45,2	0,064	0,74	45,3	**0,049**	0,63
3	54,8	0,103	0,76	53,3	0,059	0,68
2	**60,7**	0,094	**0,78**	**60,7**	0,072	**0,78**

TABLEAU C.22 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en intervalles (filtre linéaire).

Dans le cas de l'application d'un filtre non linéaire, les résultats sont légèrement différents de ceux obtenus sans filtre thermique, dans le sens où CH et Γ concordent dans pratiquement toutes les simulations. Ici, l'indice CH indique trois clusters comme étant le partitionnement optimal. Finalement, l'indice C^* oscille entre trois et quatre clusters (cf. Tableau C.23).

	Divisions hiérarchiques			Nuées dynamiques		
	CH	*C**	Γ	*CH*	*C**	Γ
5	46,1	**0,068**	0,72	42,7	0,061	0,64
4	53,6	0,081	0,83	45,3	**0,057**	0,64
3	62,3	0,144	0,85	54,1	0,074	0,71
2	**68,8**	0,14	**0,89**	**62,6**	0,09	**0,78**

TABLEAU C.23 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés en intervalles (filtre non linéaire).

C.3.4.2 Comparaison des phases 1 et 2 mois par mois

Dans le paragraphe C.3.3, une étude complémentaire a été menée en considérant un traitement d'essais mois par mois.

Signaux

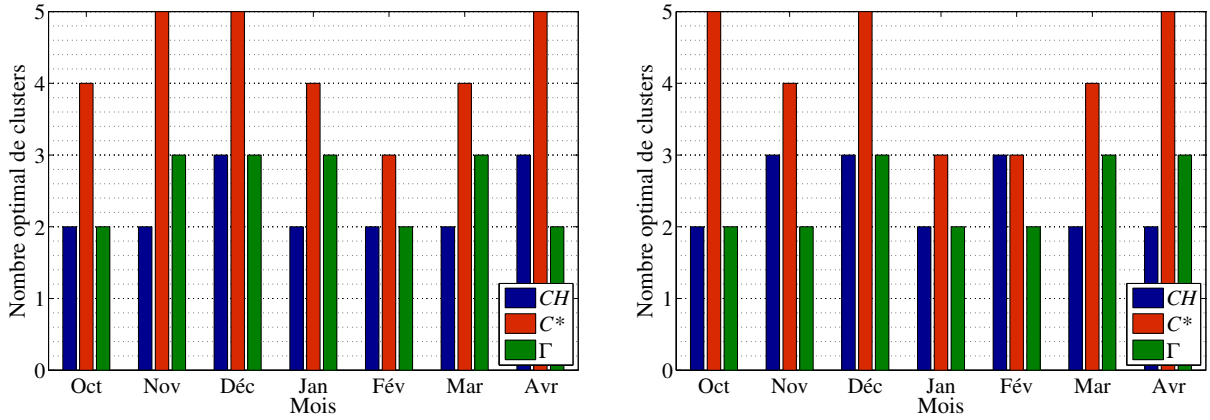
La Figure C.15 illustre les résultats obtenus mois par mois, en utilisant les divisions hiérarchiques et les nuées dynamiques appliquées aux signaux. Dans le premier cas, il est noté que les indices CH et Γ indiquent deux ou trois clusters. Dans le cas de la méthode des nuées dynamiques, il est observé que l'indice Γ oscille entre 2 et 4 clusters. Comme il a été remarqué pour la méthode des divisions hiérarchiques, tous les indices indiquent un nombre de clusters supérieur à 2 pour le mois de décembre.

Fréquences propres

La Figure C.16 montre les résultats obtenus en utilisant les méthodes de clustering appliquées aux fréquences propres. Si la méthode des divisions hiérarchiques est considérée, les indices CH et Γ indiquent deux clusters quelle que soit la paire de mois analysée. Dans le cas de la méthode des nuées dynamiques, il est noté que l'indice Γ indique trois clusters pour le mois de décembre (comme avec les signaux) et janvier. L'indice CH, toutefois, est cohérent dans tous les mois.

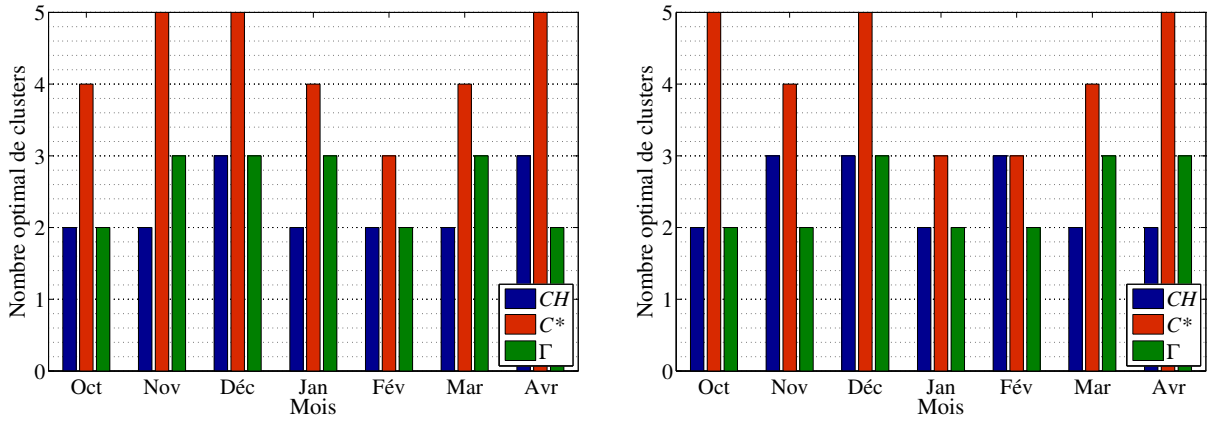
L'objectif maintenant est de vérifier si la démarche utilisée pour la correction thermique des paramètres modaux peut influencer les résultats présentés dans la Figure C.16.

Le Tableau C.24 synthétise les nombres optimaux de clusters déterminés pour chacun des indices, évalués mois après mois et en considérant les deux types de filtres. Il est noté qu'en utilisant le filtre linéaire, les indices CH et Γ signalent toujours deux clusters comme étant le partitionnement optimal, quel que soit le type de transformation symbolique utilisée. Toutefois,



(a) Divisions hiérarchiques (b) Nuées dynamiques

FIGURE C.15 – Indices basés sur les méthodes de clustering appliquées aux signaux transformés intervalles (analyse mois par mois).



(a) Divisions hiérarchiques (b) Nuées dynamiques

FIGURE C.16 – Indices basés sur les méthodes de clustering appliquées aux fréquences propres transformées intervalles (analyse mois par mois).

l'indice C^* ne concorde jamais avec les autres. Par ailleurs, si le filtre non linéaire est considéré, les résultats sont similaires. En effet, la seule exception est remarquée pour le mois de mars, où l'indice Γ indique trois groupes. Par ailleurs, l'indice C^* signale correctement le nombre de clusters pour le mois de novembre.

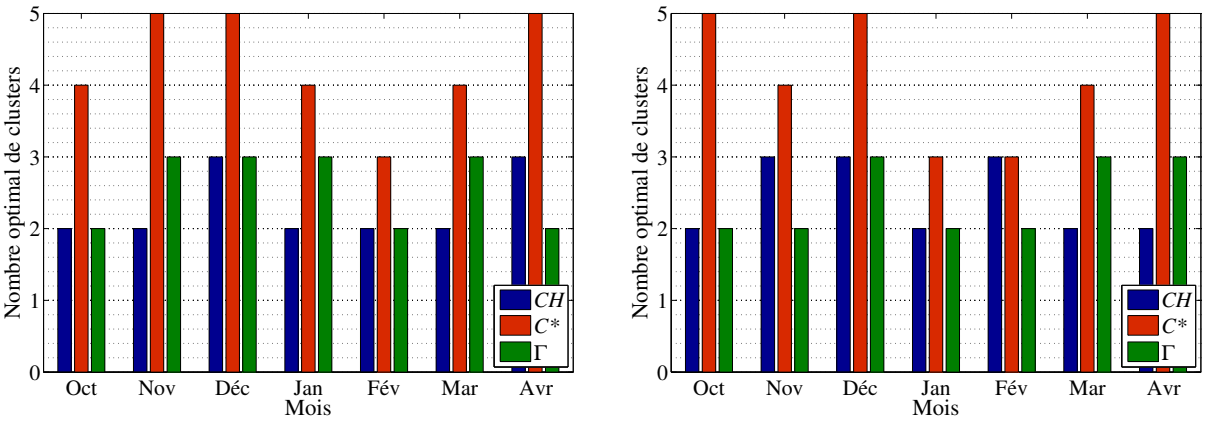
Modes propres

Enfin, les modes propres sont considérés. La Figure C.17(a) montre les résultats obtenus avec la méthode des divisions hiérarchiques. En général, les indices CH et Γ concordent dans la plupart des simulations réalisées en indiquant deux clusters. L'indice C^* ne signale jamais un partitionnement optimal avec deux clusters. La Figure C.17(b) montre les résultats obtenus en utilisant la méthodes des nuées dynamiques. Il est encore observé que pour le mois de décembre, aucun indice n'indique un partitionnement optimal avec deux clusters. Pour les autres mois, les indices CH et Γ indiquent, en général, le nombre de groupes attendu.

Le Tableau C.25 regroupe les résultats obtenus en utilisant les divisions hiérarchiques appliquées aux déformées modales. Les indices CH et Γ signalent pratiquement toujours un partitionnement optimal avec deux clusters. Par ailleurs, l'indice C^* semble osciller moins dans ces

Mois	Divisions hiérarchiques						Nuées dynamiques					
	Filtre linéaire			Filtre non linéaire			Filtre linéaire			Filtre non linéaire		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
Nov	**2**	3	**2**	**2**	**2**	**2**	**2**	3	**2**	**2**	5	**2**
Déc	**2**	3	**2**	**2**	3	**2**	**2**	4	**2**	**2**	4	**2**
Jan	**3**	4	**2**	**2**	3	**2**	**2**	5	3	**2**	5	**2**
Fév	**2**	4	**2**	**2**	4	**2**	**2**	4	**2**	**2**	3	**2**
Mar	**2**	5	**2**	**2**	4	**2**	**2**	3	**2**	**2**	3	**2**
Avr	**2**	3	**2**	**2**	3	**2**	**2**	4	**2**	**2**	3	**2**

TABLEAU C.24 – Nombres optimaux de clusters basé sur les méthodes de clustering appliquées aux fréquences propres (intervalles).



(a) Divisions hiérarchiques

(b) Nuées dynamiques

FIGURE C.17 – Indices basés sur les méthodes de clustering appliquées aux modes propres transformés intervalles (analyse mois par mois).

dernières simulations.

Mois	Divisions hiérarchiques						Nuées dynamiques					
	Filtre linéaire			Filtre non linéaire			Filtre linéaire			Filtre non linéaire		
	CH	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ	*CH*	*C**	Γ
Nov	**2**	**2**	**2**	**2**	3	**2**	**2**	3	**2**	**2**	3	**2**
Déc	**2**	3	**2**	3	3	3	**2**	4	**2**	**2**	3	**2**
Jan	3	3	**2**	**2**	4	**2**	**2**	5	3	**2**	3	**2**
Fév	**2**	4	**2**	**2**	5	**2**	**2**	3	**2**	**2**	5	**2**
Mar	**2**	5	3	**2**	5	**2**	3	3	**2**	3	3	3
Avr	**2**	5	**2**	**2**	4	**2**	**2**	5	**2**	**2**	3	**2**

TABLEAU C.25 – Nombres optimaux de clusters basé sur les méthodes de clustering appliquées aux modes propres (intervalles).

C.4 ADS couplée aux méthodes de classification supervisée

Comme il a été déjà proposé dans les chapitre 4 et 5, une analyse utilisant le couplage entre l'ADS aux méthodes de classification supervisée est proposée. Cette approche est appliquée aux

signaux et aux trois premiers paramètres modaux identifiés.

C.4.1 Analyse sur l'ensemble des essais

Cette première étude considère la totalité des essais enregistrés. Ici, une étude statistique est menée en faisant varier les dimensions des groupes d'entraînement et de test. Au total, 10000 simulations sont réalisées pour chaque configuration proposée et les meilleurs taux de classification correcte sont évalués aussi bien que les maximaux et minimaux.

C.4.1.1 Signaux

Les méthodes de classification sont d'abord appliquées aux signaux. Le Tableau C.26 regroupe les résultats obtenus. En général, les RN et les SVM sont plus robustes pour la classification des essais que celles basées sur les ADB. Des taux moyen de classification correcte élevés sont obtenus. Ces résultats montrent que les mesures dynamiques peuvent effectivement être exploitées si des outils pertinents pour leur manipulation sont utilisés.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	94	100	100	93	100	100	95	100	100
Moyen	83	95	94	86	99	97	86	99	96
Minimal	78	90	92	76	98	94	70	99	91

TABLEAU C.26 – Taux de classification correcte utilisant les signaux transformés en intervalles pour l'ensemble total des essais (valeurs en %).

C.4.1.2 Fréquences propres

Le Tableau C.27 regroupe les pourcentages de classification correcte obtenus lorsque la transformation en intervalles interquartiles est appliquée aux fréquences propres. Les taux de classification correcte sont pratiquement du même ordre que ceux obtenus en utilisant les signaux. Les méthodes basées sur les RN et SVM arrivent toujours à mieux classer les essais, quel que soit le nombre d'observations utilisées dans le groupe d'entraînement. Les taux moyens évalués pour les ADB sont relativement plus importants que ceux obtenus avec les signaux.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	89	100	100	93	100	100	94	100	100
Moyen	88	97	92	87	94	93	86	96	93
Minimal	67	89	77	71	92	79	79	90	81

TABLEAU C.27 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour l'ensemble total des essais (valeurs en %).

L'utilisation des filtres thermiques pour la correction des fréquences propres est maintenant considérée. Le Tableau C.28 regroupe les pourcentages de classification correcte obtenus lorsque le filtre linéaire est utilisé. En général, les méthodes de classification supervisée permettent d'atteindre des résultats aussi pertinents que ceux obtenus sans le filtre thermique. Ceci est notamment observé pour les méthodes des RN et des SVM. Par ailleurs, bien que les ADB et les RN n'arrivent pas à classer 100% des essais, les SVM le font pour quelques simulations.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	87	98	100	86	96	99	92	96	97
Moyen	84	96	91	83	92	92	83	96	91
Minimal	68	91	75	70	88	80	76	82	80

TABLEAU C.28 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour l'ensemble total des essais (filtre linéaire ; valeurs en %).

L'utilisation du filtre non linéaire est maintenant considérée. Le Tableau C.29 regroupe les résultats obtenus. Les taux de classification pour les ADB sont significativement meilleurs que ceux évalués sans l'utilisation des lois de correction. Ces résultats montrent la pertinence de l'utilisation du filtre non linéaire pour la correction des fréquences propres.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	84	91	95	85	93	95	86	93	95
Moyen	83	91	90	82	96	94	81	94	96
Minimal	76	90	75	76	93	83	78	87	84

TABLEAU C.29 – Taux de classification correcte utilisant les fréquences propres transformées en intervalles pour l'ensemble total des essais (filtre non linéaire ; valeurs en %).

C.4.1.3 Modes propres

Les méthodes de classification supervisée sont maintenant appliquées aux modes propres. Les résultats obtenus sont détaillés dans le Tableau C.30. Les taux moyens de classification correcte sont pratiquement du même ordre que ceux évalués pour les fréquences propres. En général, les RN et SVM sont les méthodes qui permettent d'obtenir les taux de classification correcte les plus importants.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	90	96	100	89	97	100	91	97	100
Moyen	82	95	95	80	96	95	82	97	96
Minimal	77	95	79	79	95	82	80	95	89

TABLEAU C.30 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour l'ensemble total des essais (valeurs en %).

De plus, des simulations sont réalisées en utilisant les modes propres corrigés par le filtre linéaire. Comme remarqué précédemment, les taux de classification correcte sont pratiquement du même ordre que ceux trouvés avec les fréquences propres et sensiblement plus élevés que ceux obtenus sans le filtre. Toutefois, les taux de classification correcte obtenus par les RN ont été relativement plus faibles (cf. Tableau C.31).

Les taux de réussite calculés en considérant les déformées modales corrigées par le filtre non linéaire sont regroupés dans le Tableau C.32. Dans ce cas, les améliorations apportées par l'utilisation du filtre sont moins évidentes. Il est remarqué que les taux moyens restent plutôt plus faibles que ceux évalués sans correction thermique.

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	86	93	97	88	97	95	88	92	96
Moyen	80	92	93	79	95	94	79	92	93
Minimal	67	87	79	69	91	80	72	88	85

TABLEAU C.31 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour l'ensemble total des essais (filtre linéaire ; valeurs en %).

	30% E., 10% V., 60% T.			40% E., 10% V., 50% T.			50% E., 10% V., 40% T.		
Méthode	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**	**ADB**	**RN**	**SVM**
Maximal	87	95	99	86	97	94	90	95	98
Moyen	84	92	91	83	97	94	80	96	94
Minimal	70	90	78	72	92	80	76	86	84

TABLEAU C.32 – Taux de classification correcte utilisant les modes propres transformés en intervalles pour l'ensemble total des essais (filtre non linéaire ; valeurs en %).

C.4.2 Comparaison des phases 1 et 2 mois par mois

La démarche basée sur l'analyse mois par mois des essais est considérée en utilisant les méthodes de classification supervisée. L'objectif est d'observer si les taux de classification correcte peuvent être améliorés. Dans un soucis de clarté du texte, les résultats sont présentés de façon synthétique au travers de figures. Étant donné le nombre conséquent de simulations réalisées, seuls les taux moyens évalués pour chaque méthode sont montrés.

Les Figures C.18 et C.19 regroupent les pourcentages moyens évalués en considérant 40% et 50% des essais pour l'entraînement, respectivement, en considérant la transformation en histogrammes.

Les Figures C.20, C.21 et C.22 regroupent les pourcentages moyens évalués en considérant 30%, 40% et 50% des essais pour l'entraînement, respectivement, en considérant la transformation en intervalles interquartiles.

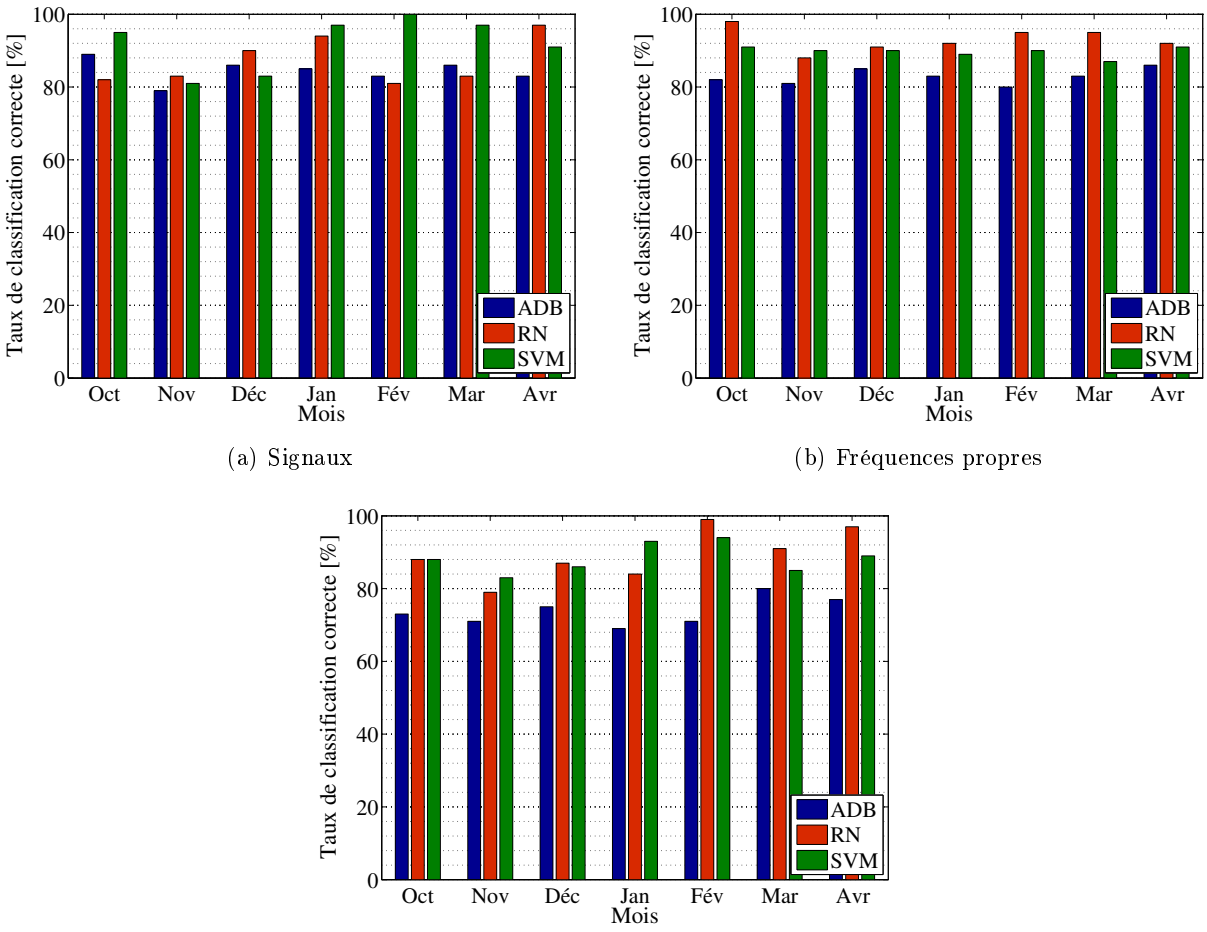
C.4.2.1 Application aux paramètres modaux corrigés

Les Figures C.23 et C.24 regroupent les pourcentages moyens évalués en considérant 40% et 50% des essais pour l'entraînement, respectivement, en considérant la transformation en histogrammes et l'utilisation du filtre linéaire.

Les Figures C.25 et C.26 regroupent les pourcentages moyens évalués en considérant 40% et 50% des essais pour l'entraînement, respectivement, en considérant la transformation en histogrammes et l'utilisation du filtre non linéaire.

Les Figures C.27, C.23 et C.24 regroupent les pourcentages moyens évalués en considérant 30%, 40% et 50% des essais pour l'entraînement, respectivement, en considérant la transformation en intervalles interquartiles et l'utilisation du filtre linéaire.

Les Figures C.27, C.23 et C.24 regroupent les pourcentages moyens évalués en considérant 30%, 40% et 50% des essais pour l'entraînement, respectivement, en considérant la transformation en intervalles interquartiles et l'utilisation du filtre non linéaire.

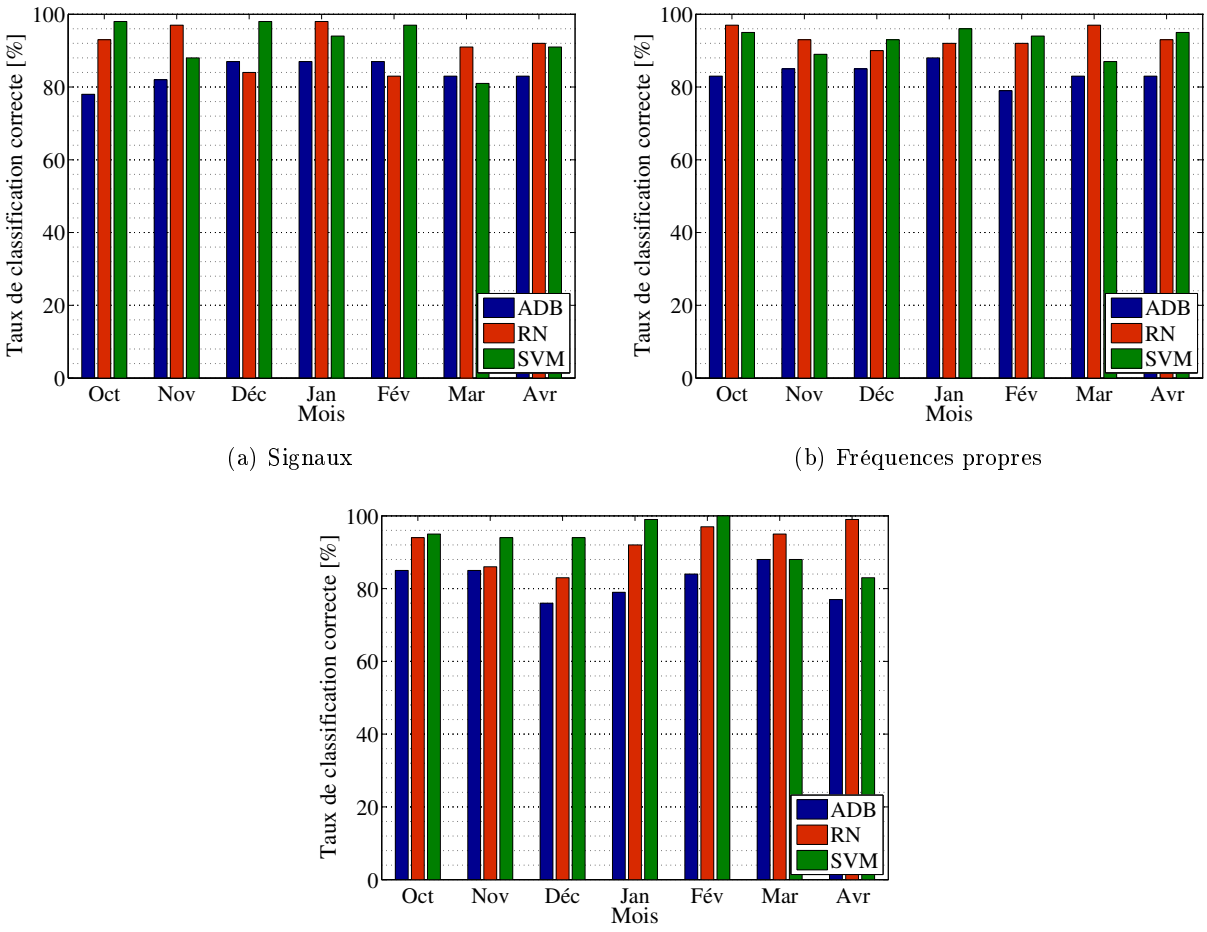


(a) Signaux

(b) Fréquences propres

(c) Modes propres

FIGURE C.18 – Méthodes de classification supervisée appliquées aux signaux et paramètres modaux transformés en histogrammes (analyse mois par mois ; 40% entr., 10% val., 50% test ; valeurs en %).

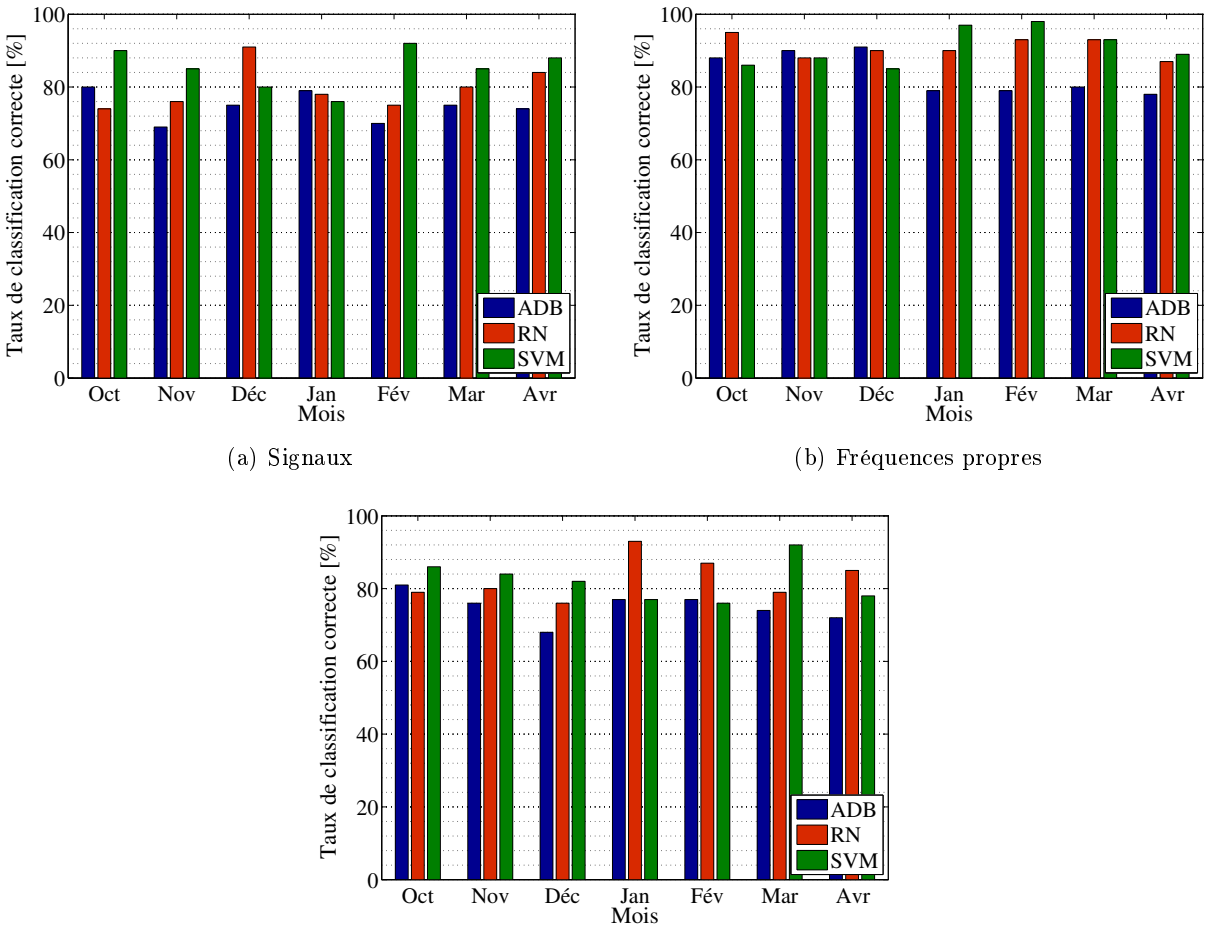


(a) Signaux

(b) Fréquences propres

(c) Modes propres

FIGURE C.19 – Méthodes de classification supervisée appliquées aux signaux et paramètres modaux transformés en histogrammes (analyse mois par mois ; 50% entr., 10% val., 40% test ; valeurs en %).

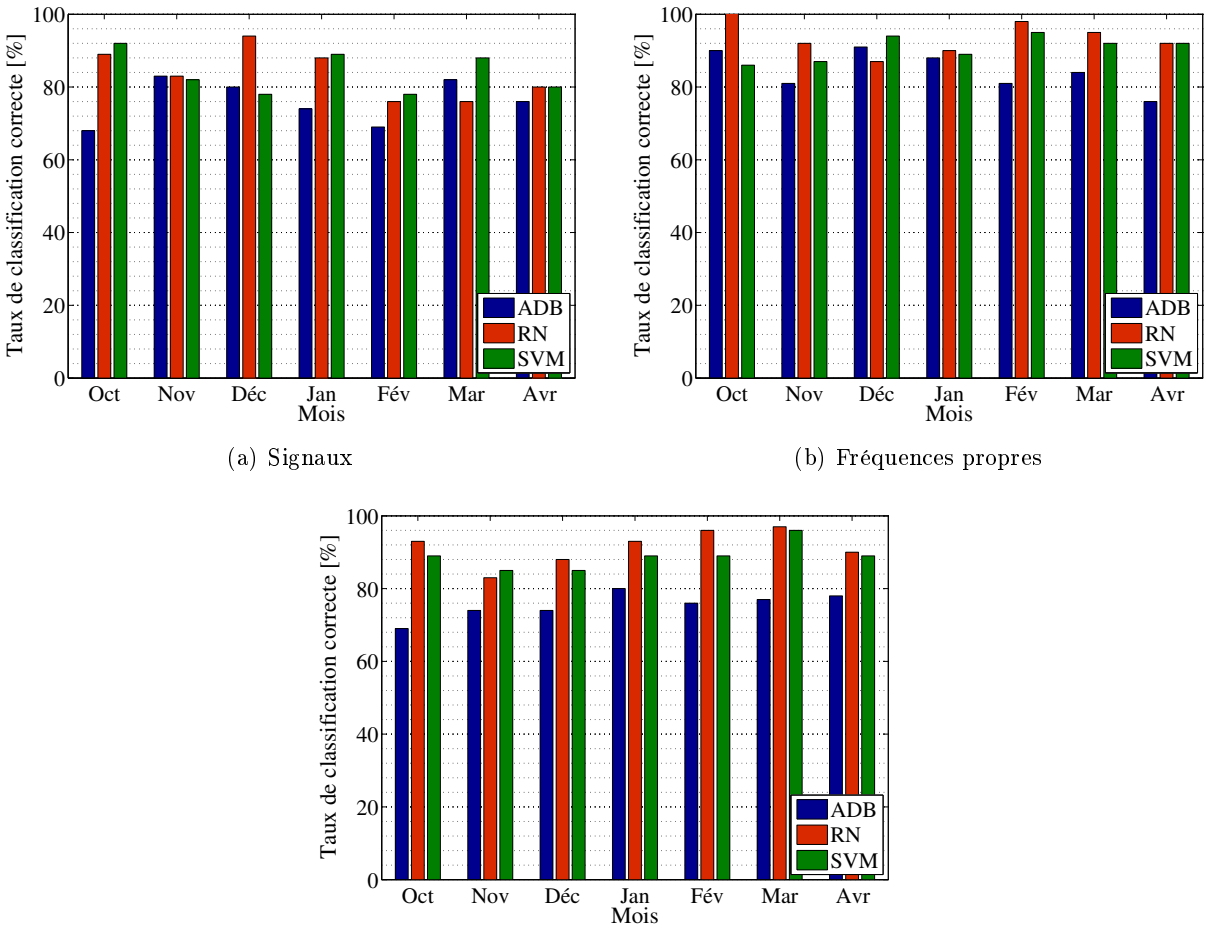


(a) Signaux

(b) Fréquences propres

(c) Modes propres

FIGURE C.20 – Méthodes de classification supervisée appliquées aux signaux et paramètres modaux transformés en intervalles (analyse mois par mois ; 30% entr., 10% val., 60% test ; valeurs en %).

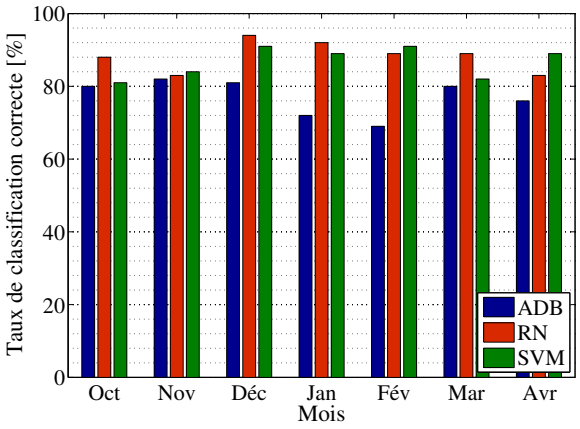


(a) Signaux

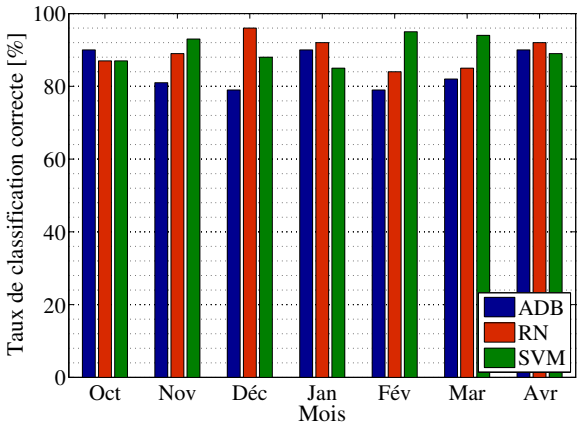
(b) Fréquences propres

(c) Modes propres

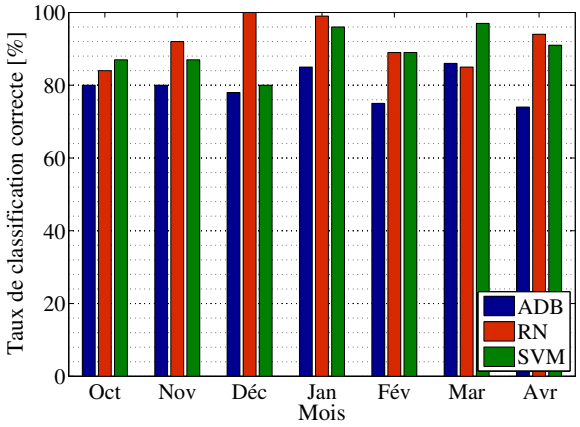
FIGURE C.21 – Méthodes de classification supervisée appliquées aux signaux et paramètres modaux transformés en intervalles (analyse mois par mois ; 40% entr., 10% val., 50% test ; valeurs en %).



(a) Signaux

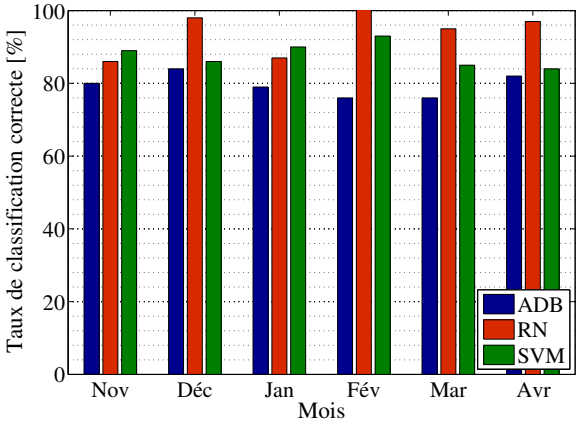


(b) Fréquences propres

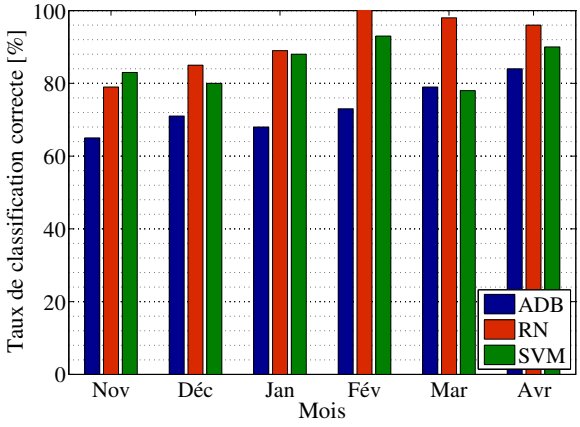


(c) Modes propres

FIGURE C.22 – Méthodes de classification supervisée appliquées aux signaux et paramètres modaux transformés en intervalles (analyse mois par mois ; 50% entr., 10% val., 40% test ; valeurs en %).

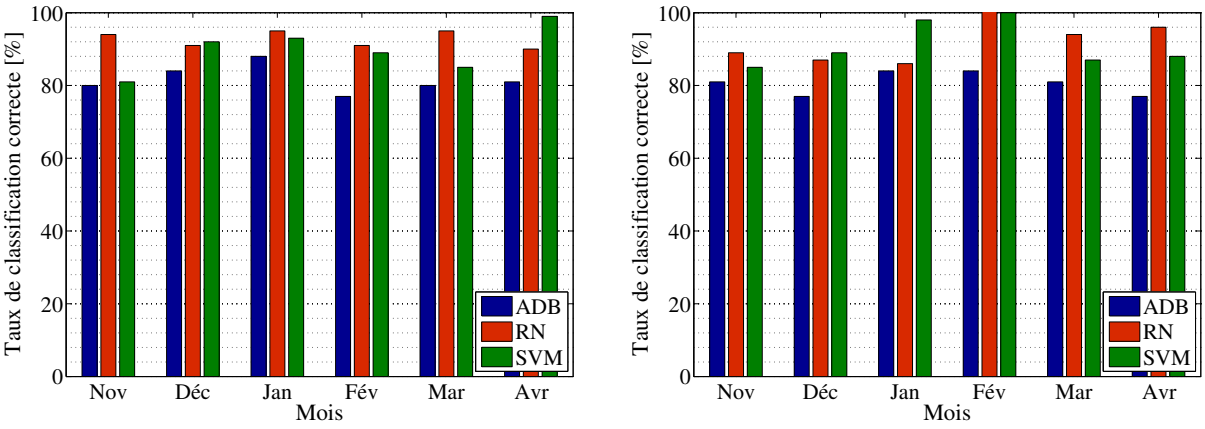


(a) Fréquences propres



(b) Modes propres

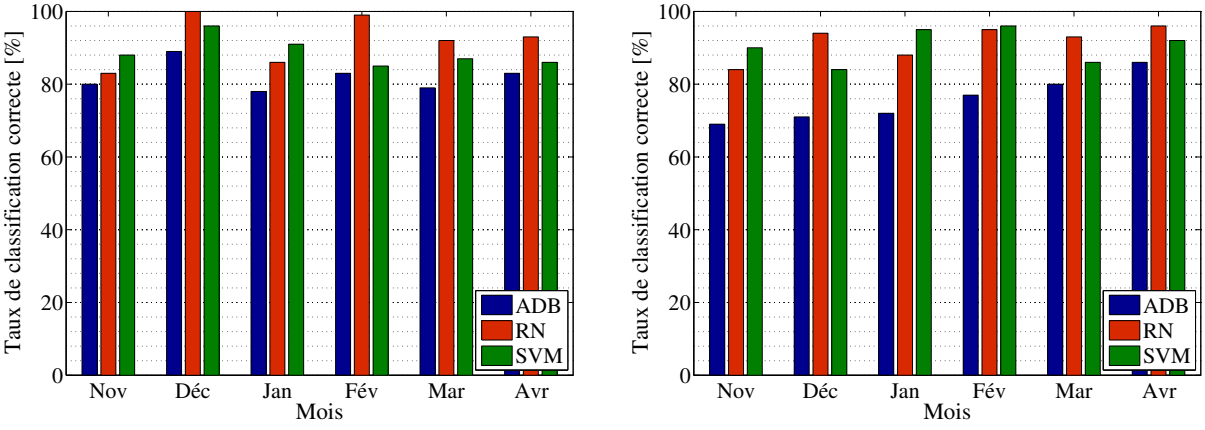
FIGURE C.23 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en histogrammes (filtre linéaire ; analyse mois par mois ; 40% entr., 10% val., 50% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

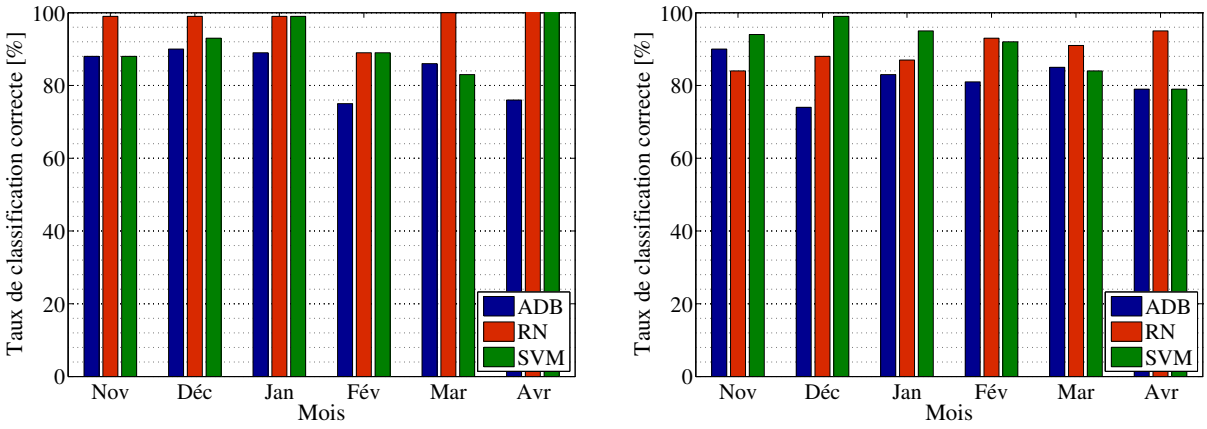
FIGURE C.24 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en histogrammes (filtre linéaire ; analyse mois par mois ; 50% entr., 10% val., 40% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

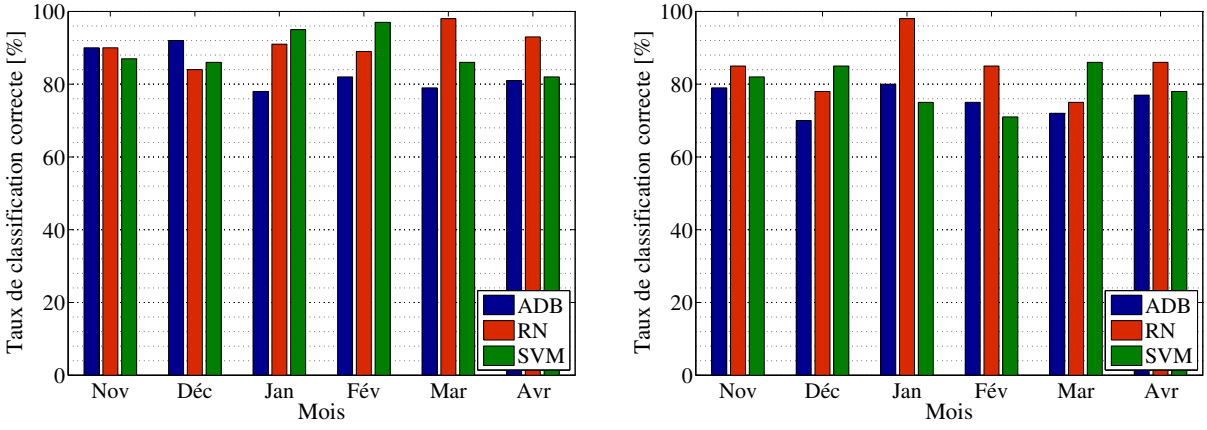
FIGURE C.25 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en histogrammes (filtre non linéaire ; analyse mois par mois ; 40% entr., 10% val., 50% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

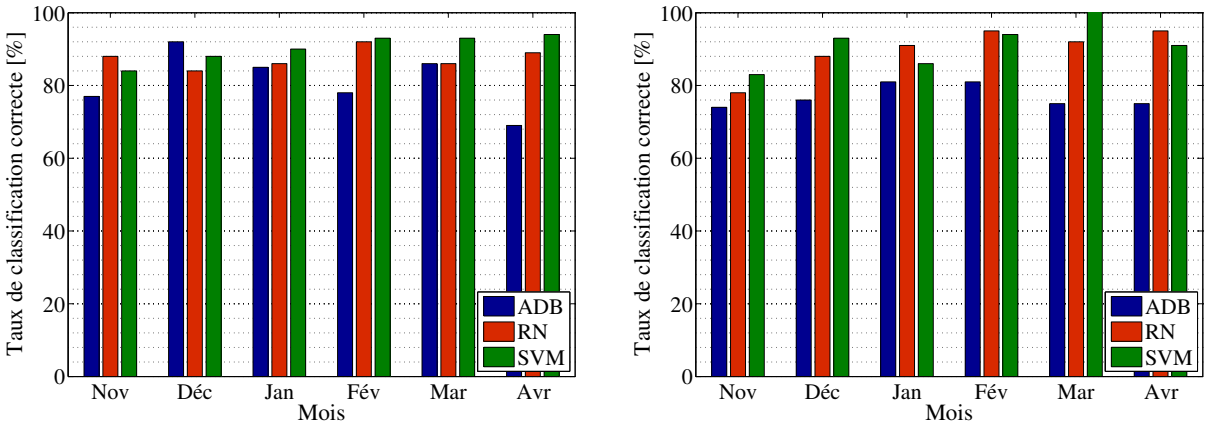
FIGURE C.26 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en histogrammes (filtre non linéaire ; analyse mois par mois ; 50% entr., 10% val., 40% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

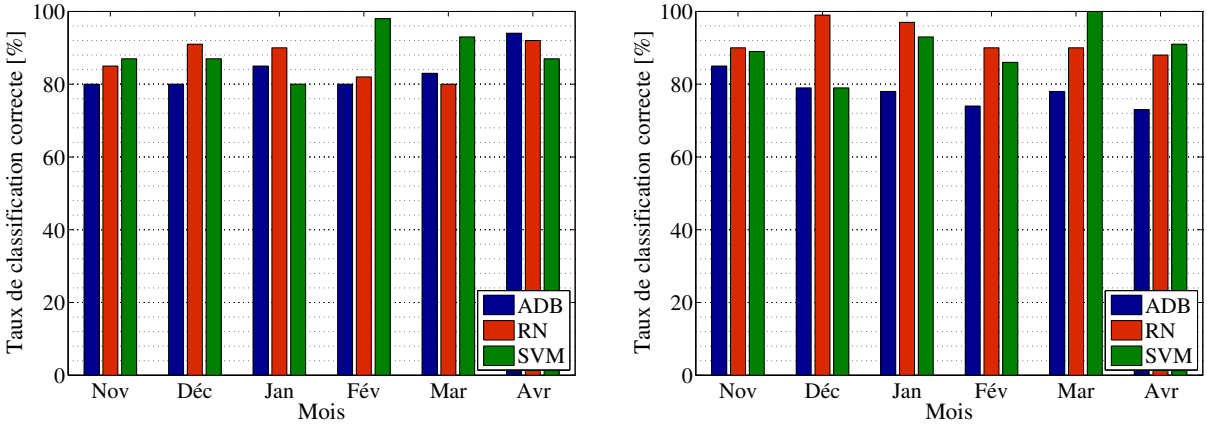
FIGURE C.27 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en intervalles (filtre linéaire ; analyse mois par mois ; 30% entr., 10% val., 60% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

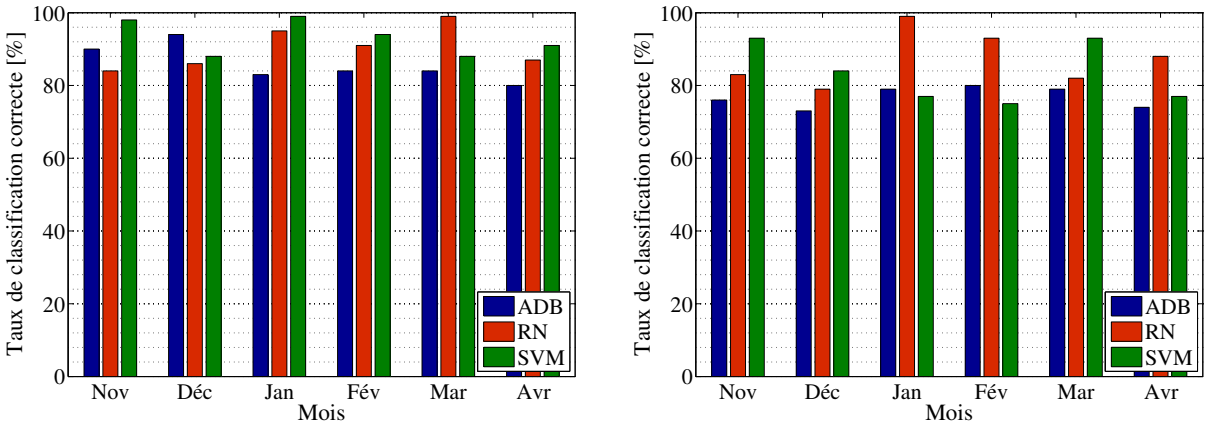
FIGURE C.28 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en intervalles (filtre linéaire ; analyse mois par mois ; 40% entr., 10% val., 50% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

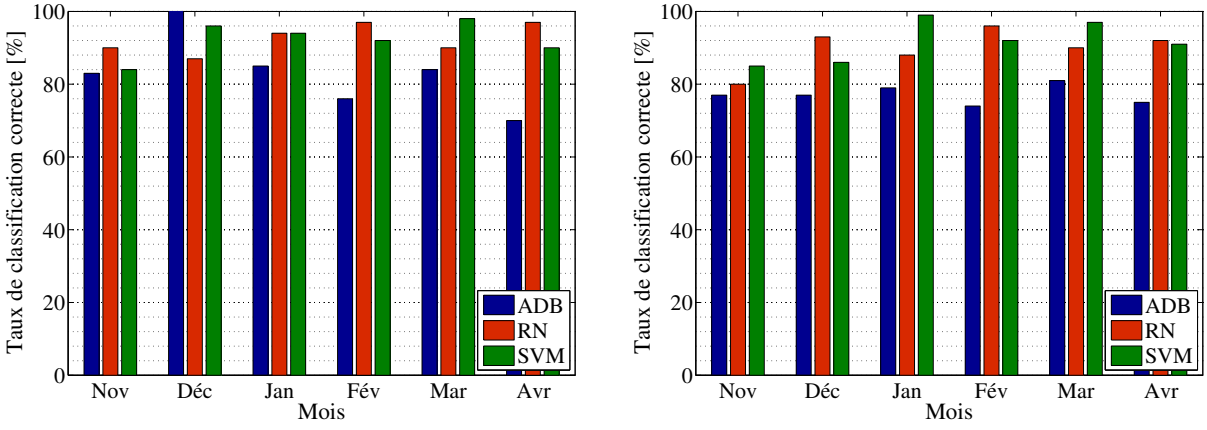
FIGURE C.29 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en intervalles (filtre linéaire ; analyse mois par mois ; 50% entr., 10% val., 40% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

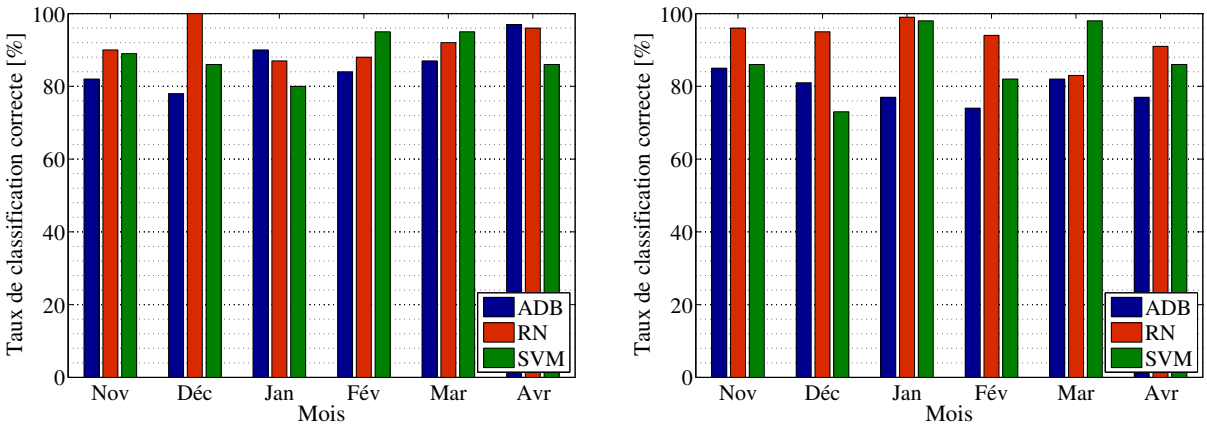
FIGURE C.30 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en intervalles (filtre non linéaire ; analyse mois par mois ; 30% entr., 10% val., 60% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

FIGURE C.31 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en intervalles (filtre non linéaire ; analyse mois par mois ; 40% entr., 10% val., 50% test ; valeurs en %).



(a) Fréquences propres

(b) Modes propres

FIGURE C.32 – Méthodes de classification supervisée appliquées aux paramètres modaux transformés en intervalles (filtre non linéaire ; analyse mois par mois ; 50% entr., 10% val., 40% test ; valeurs en %).

Annexe D

Méthodes d'identification modale

Cette annexe a pour objectif de donner un bref aperçu des méthodes d'identification modale utilisées dans le cadre de cette thèse. L'idée est de montrer comment ces méthodes ont été mises en œuvre dans chacune des applications présentées dans les chapitres 4, 5 et 6.

Dans un premier temps, la méthode des Réalisations Stochastiques (MRS) est présentée. Cette méthode a été utilisée lors des campagnes d'essais dynamiques réalisées sur le pont PI-57, sur l'Oise. Cette technique est couplée à la méthode du Décrément Aléatoire (MDA), qui est présentée dans un deuxième temps. La MDA a également été utilisée pour l'identification modale du pont-rail PK 057+317. Dans un troisième temps, un bref aperçu de la méthode d'Ibrahim est donné. Cette méthode a été couplée à la MDA lors des simulations numériques présentées dans le chapitre 4.

D.1 Méthode des Réalisations Stochastiques (MRS)

La méthode des Réalisations Stochastiques est une généralisation de la méthode des Réalisations Déterministes (ERA, en anglais). L'ERA a été développée par (Juang & Pappa 1984, 1985) pour des applications dans le domaine aérospatial. Cette technique suppose que la fonction des réponses libres peut être obtenue par des essais dynamiques (vibration impulsionnelle ou vibration libre) et, qu'à partir des ces fonctions, les paramètres modaux peuvent être estimés. La MRS remplace les fonctions des réponses libres par des fonctions de corrélation dans les cas où la structure est excitée par un chargement stationnaire (Juang et al. 1988). L'ERA et la MRS sont basées sur une fonction représentée par des décréments exponentiels à partir desquels les caractéristiques modales sont estimées.

Considérons le cas d'un système dynamique multidimensionnel décrit par une équation linéaire de seconde ordre n-dimensionnelle :

$$\begin{cases} [M]\{\ddot{y}(t)\} + [D]\{y(t)\} + [K]\{y(t)\} = \{f(t)\} \\ \{y(0)\} = \{y_0\}, \{\dot{y}(0)\} = \{\dot{y}_0\} \end{cases} \tag{D.1}$$

où $[M], [D], [K]$ $(\in \mathbb{R}^{n\times n})$ sont respectivement les matrices de masse, d'amortissement et de rigidité, $\{y(t)\}$ le vecteur de déplacement et $\{f(t)\}$ l'excitation. Cette équation est équivalente à l'équation d'états :

$$\begin{cases} \left\{\dot{X}(t)\right\} = [A_c]\{X(t)\} + [b_c]\{f(t)\} \\ \{y(t)\} = [H_c]\{X(t)\} \end{cases} \tag{D.2}$$

où $\{X\} = \left(\{y\}^T \ \{\dot{y}\}^T\right)^T \in \mathbb{R}^{2n}$ est le vecteur d'état et les matrices $[A_c] \in \mathbb{R}^{2n\times 2n}$, $[H_c] \in \mathbb{R}^{n\times 2n}$, $[b_c] \in \mathbb{R}^{2n\times n}$ sont définies par

$$[A_c] = \begin{bmatrix} [0]_{n\times n} & [I]_{n\times n} \\ -[M]^{-1}[K] & -[M]^{-1}[D] \end{bmatrix}; [b_c] = \begin{bmatrix} [0]_{n\times n} \\ [M]^{-1} \end{bmatrix}; [H_c] = \begin{bmatrix} [I]_{n\times n} & [0]_{n\times n} \end{bmatrix} \tag{D.3}$$

Étant donné que la réponse n'est pas discrète dans le domaine temporel, l'expression d'état (Équation (D.2)) doit être remplacée par une formulation discrète :

$$\begin{cases} \left\{\dot{X}_k\right\} = [A]\{X_k\} + [B]\{f_k\} \\ \{y_k\} = [D]\{X_k\} \end{cases} \tag{D.4}$$

Il est démontrable que la matrice $[A]$ est liée à la matrice $[A_c]$ par l'expression suivante (Crémona 1990) :

$$[A] = \exp\left(\Delta t\ [A_c]\right) \tag{D.5}$$

où Δt est le temps d'échantillonnage. Les fréquences propres et les taux d'amortissement sont évalués à partir des valeurs propres λ_r de la matrice $[A]$ comme suit :

$$\begin{cases} f_r = \frac{1}{2\pi\,\Delta t}\sqrt{\left[\ln\sqrt{\Re(\lambda_r)^2+\Im(\lambda_r)^2}\right]^2 + \left[\tan^{-1}\left(\Im(\lambda_r)/\Re(\lambda_r)\right)\right]^2} \\ \varsigma_r = \frac{-\tan^{-1}(\Im(\lambda_r)/\Re(\lambda_r))}{\sqrt{\left[\ln\sqrt{\Re(\lambda_r)^2+\Im(\lambda_r)^2}\right]^2+\left[\tan^{-1}(\Im(\lambda_r)/\Re(\lambda_r))\right]^2}} \end{cases} \tag{D.6}$$

où $\Im(\lambda_r), \Re(\lambda_r)$ sont respectivement les parties imaginaire et réelle de la valeur propre λ_r de la matrice $[A]$. Les fonctions des réponses impulsionnelles sont décrites par

$$\{h_k\} = [C]\,[A]^{k-1}\,[B] \tag{D.7}$$

tandis que, sous excitation stationnaire, les fonctions de corrélation sont données par :

$$[R_k] = [C]\,[A]^{k-1}\,\mathbb{E}\left(\left(\{X_{k+1}\}\ \{y_k\}^T\right)\right) = [C]\,[A]^{k-1}\,[G] \tag{D.8}$$

Les Équations (D.7) et (D.8) montrent que les décréments exponentiels et les corrélations sont similaires et que les techniques utilisées pour l'estimation de la matrice $[A]$ peuvent être employées de manière indifférente pour les deux fonctions. Ceci est la raison pour laquelle la MRS est considérée comme une généralisation stochastique de l'ERA.

Dans le cadre de cette thèse, la MRS est appliquée au capteur de référence afin de pouvoir estimer les principales caractéristiques modales d'une réponse dynamique. Dénotant y_k la réponse discrète mesurée par ce capteur, les fonctions de corrélation sont estimées par :

$$\hat{R}_k = \frac{1}{M-k}\sum_{m=0}^{M-k-1} y_{k+m}\ y_k \tag{D.9}$$

où M est la longueur du signal. La MRS consiste à appliquer les étapes suivantes : soit $[H_p]$ la matrice de Hankel par blocs composée de p blocs lignes (et colonnes) de la matrice de corrélation R_k (Hermans & Van der Auwaraer 1999) :

$$[H_p] = \begin{bmatrix} R_1 & R_2 & \cdots & R_p \\ R_2 & R_3 & \cdots & R_{p+1} \\ \vdots & \cdots & \ddots & \vdots \\ R_p & R_{p+1} & \cdots & R_{p+p-1} \end{bmatrix} \tag{D.10}$$

Il est possible de démontrer qu'à partir de calculs directs des matrices de corrélations, la factorisation suivante est obtenue :

$$[H_p] = [O_p]\,[C_p] \tag{D.11}$$

avec

$$[O_p] = \begin{bmatrix} [C] \\ [C]\,[A] \\ \vdots \\ [C]\ \ [A]^{p-1} \end{bmatrix}; [C_p] = \begin{bmatrix} [G] & [A]\ [G] & \cdots & [A]^{p-1}\ [G] \end{bmatrix} \tag{D.12}$$

Les matrices $[O_p], [C_p]$ sont appelées « matrices d'observabilité » et de « contrôlabilité ». Il est supposé que le rang de ces matrices est égal à n. En appliquant une décomposition en valeurs singulières à la matrice de Hankel, il vient :

$$[H_p] = [U]\,[S]\,[V] \tag{D.13}$$

où $[S]$ contient n valeurs singulières non nulles en ordre décroissant, les n colonnes de $[U]$ sont les vecteurs singuliers correspondant gauches et les n colonnes de $[V]$ les vecteurs singuliers droits. En replaçant l'Équation (D.7) dans l'Équation (D.11), il vient :

$$[O_p] \ = \ [U]\,[S]^{1/2} \tag{D.14}$$

Les matrices $[A], [C]$ sont alors identifiées à partir de la structure décalée de $[O_p]$:

$$[C] = \text{ premier bloc colonne de } [O_p] \tag{D.15}$$

$$\left[O^{\uparrow}_{p-1}\right] \ = [O_{p-1}]\ [A] \tag{D.16}$$

où $\left[O^{\uparrow}_{p-1}\right]$ est la matrice obtenue en supprimant le dernier bloc de $[O_p]$ et $\left[O^{\uparrow}_{p-1}\right]$ est la matrice décalée d'un bloc vers le haut (en supprimant le premier bloc).

Il est noté que les Équations (D.15) et (D.16) ne sont valables que pour des « vraies » matrices de Hankel par blocs et que pour des systèmes avec un nombre de degrés de liberté fini. Dans la pratique, les structures admettent un nombre de modes infini dénombrable et la matrice de Hankel est construite à partir de fonctions d'auto-corrélation « empiriques ». En conséquence, la décomposition en valeurs singulières ne sera pas exactement de rang n. Les valeurs singulières doivent être séparées en deux familles, une liée aux modes réels et l'autre aux modes « fantômes ». Cette troncature est un point important et de nombreux critères existent pour distinguer ces différents modes (Crémona & Brandon 1992). Les paramètres n et p sont inconnus au préalable. Le principe est alors de faire varier le nombre de modes n et de définir la matrice de Hankel comme de dimension $p = 2n$ (Abdelghani et al. 1999). Afin d'éviter des erreurs importantes dans les premiers modes, il est souvent nécessaire de sur-parametrer la matrice de Hankel, même si un nombre inférieur de modes est souhaité. Le problème de la sur-parametrisation est bien connu dans le domaine d'identification des systèmes (Crémona 1990). Étant donné que le nombre de modes varie, un jeu de fréquences propres est identifié pour chaque ordre du modèle $p = 2n$. Cela produit des diagrammes de stabilisation à partir desquels les fréquences propres les plus stables sont considérées comme caractéristiques de la réponse dynamique. La Figure D.1 illustre un exemple d'un diagramme de stabilisation pour un essai du pont sur l'Oise. Dans cette figure, des bandes de filtrage ($\pm$ 10%) autour de chaque fréquence propre sont également montrées. Ces bandes de filtrage sont utilisées comme données d'entrée pour la MDA.

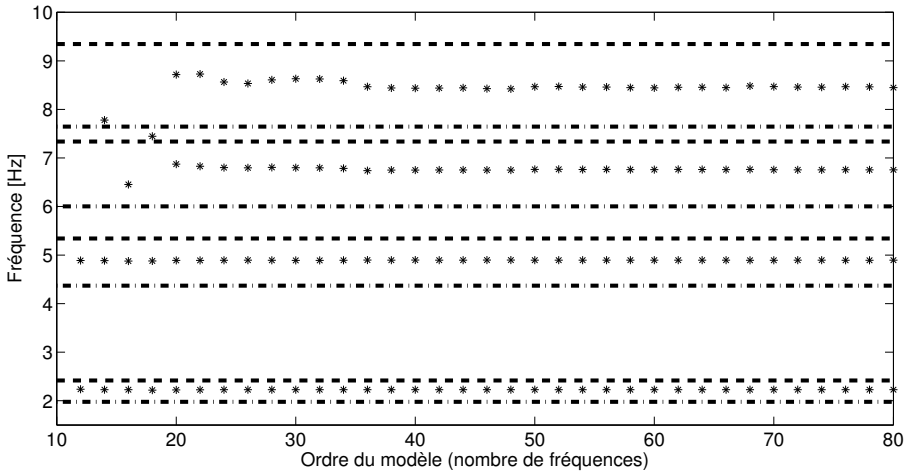


FIGURE D.1 – Exemple d'un diagramme de stabilisation des fréquences propres et des bandes de filtrage.

D.2 Méthode du Décrément Aléatoire (MDA)

La Méthode du Décrément Aléatoire a été introduite dans la fin des années 1960 par Henry Cole (Cole 1968, 1973). Cette méthode est attractive car sa mise en œuvre est relativement simple et le coût de calcul (au niveau du temps) est plutôt faible. Dans les premières recherches, les taux d'amortissement étaient estimés à partir des densités spectrales de puissance (évaluées en utilisant l'algorithme FFT). Toutefois, il a été observé que ces paramètres avaient un écart assez important. Par ailleurs, aucune approche n'était connue pour la détection des non-linéarités de la densité spectrale des réponses dynamiques sous sollicitation ambiante. Ainsi, Cole a utilisé des échantillons prélevés des fonctions d'auto-corrélation des mesures temporelles pour l'identification des paramètres modaux. En cela, les taux d'amortissement et les fréquences propres de la structure sont évalués à partir de l'enveloppe de la fonction d'auto-corrélation. Quelques limitations ont été cependant trouvées, étant donné que les fonctions d'auto-corrélation changent avec la variations des sollicitations ambiantes. Ces problèmes ont été à l'origine de la création de la MDA. De manière générale, Cole a transformé une série de mesures temporelles aléatoires dans des fonctions de décrément exponentiel. L'avantage est que ces décréments ne contiennent que des informations sur la structure et non sur les sollicitations aléatoires.

Afin d'expliquer le concept de la MDA, l'illustration suivante est donnée. La réponse aléatoire d'une structure à l'instant $t_0 + t$ est composée de trois parties : 1) la réponse produite par le déplacement initial au temps t_0, 2) la réponse produite par la vitesse initiale au temps t_0, 3) la réponse aléatoire qui se produit par une charge appliquée à la structure entre les instants t_0 et $t_0 + t$. Si un segment temporel est défini à chaque fois que la réponse aléatoire a un déplacement initial égal à a et si tous ces segments sont moyennés, la partie aléatoire due à la charge appliquée se réduira pratiquement à zéro. Par ailleurs, avec l'alternance des vitesses positives et négatives, la somme des réponses liées à la vitesse s'élimine (ou devient nulle) tandis que la somme des réponses liées au déplacement initial a est la seule partie conservée. Le principe de cette méthode est illustré à la Figure D.2.

Le déplacement initial a est souvent appelé *mesure de déclenchement* et l'instant t_0, le *point de déclenchement*. Le processus d'estimation des fonctions du décrément aléatoire peut être formulé comme la somme des segments temporels pris à partir de la réponse dynamique de la structure, pour les instants correspondant aux points de déclenchement avec un déplacement initial égal à

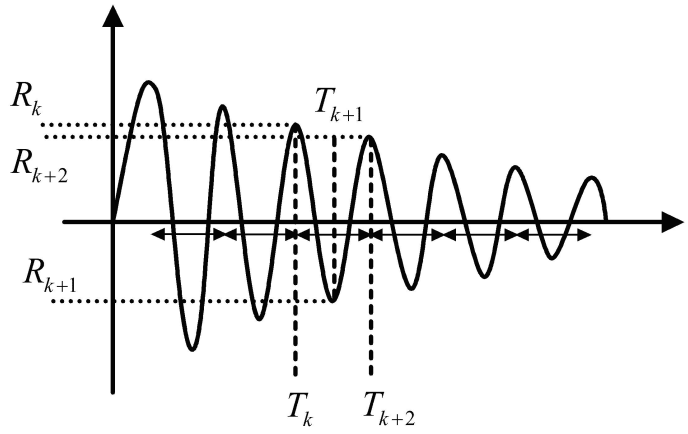


FIGURE D.2 – Principe de la méthode du décrément aléatoire.

a :

$$D_y(\tau) = \mathbb{E}\left[y(t+\tau)\,\middle|\,T_{y(t)}\right] \tag{D.17}$$

où $T_{y(t)}$ est la *condition de déclenchement*. Dans le cas de la condition « dépassement de zéro », l'expression est simplifiée :

$$T_{y(t)} = \{y(t) = a\} \tag{D.18}$$

Toutefois, cette formulation n'est pas correcte d'un point de vue probabiliste. En effet, plusieurs discussions ont été mises au point par Bernard (1988, 1990) qui a proposé une formulation mathématique plus élégante et correcte de la MDA. Chang (1975) a étudié d'autres conditions de déclenchement. Il en existe différents types qui peuvent être résumés de la manière suivante (Asmussen 1997) :

$$T_{y(t)} = \{0 \leq a_1 \leq y(t) < a_2, v_1 \leq \dot{y}(t) < v_2\} \tag{D.19}$$

Asmussen & Brincker (1996) a démontré que la fonction du décrément aléatoire d'un signal stationnaire de moyenne nulle peut être liée aux fonctions d'auto-corrélation :

$$D_y(\tau) = \frac{R_y(\tau)}{\sigma_y^2}\tilde{a} - \frac{\dot{R}_y(\tau)}{\sigma_{\dot{y}}^2}\tilde{b} \tag{D.20}$$

où $\dot{R}_y(\tau)$ est la dérivée de la fonction d'auto-corrélation $R_y(\tau)$. Les paramètres $\tilde{a}$ et $\tilde{b}$ sont déterminés à partir des fonctions de densité de probabilité de y et $\dot{y}$:

$$\tilde{a} = \frac{\int_{a_1}^{a_2} u\, f_y(u)\, du}{\int_{a_1}^{a_2} f_y(u)\, du}; \quad \tilde{b} = \frac{\int_{v_1}^{v_2} \dot{u}\, f_{\dot{y}}(\dot{u})\, d\dot{u}}{\int_{v_1}^{v_2} f_{\dot{y}}(\dot{u})\, d\dot{u}} \tag{D.21}$$

Par exemple, pour la condition de déclenchement « dépassement de niveau » (utilisée dans l'application du pont-rail PK 075+317), l'Équation (D.18) devient :

$$D_y(\tau) = \frac{R_y(\tau)}{\sigma_y^2}a \tag{D.22}$$

tandis que pour la condition $T_{y(t)} = \{y(t) = 0, \dot{y}(t) > 0\}$ (utilisée dans l'application du pont sur l'Oise PI57), l'Équation (D.20) est simplifiée :

$$D_y(\tau) = -\frac{\dot{R}_y(\tau)}{\sigma_{\dot{y}}^2}\tilde{b} \tag{D.23}$$

Ces résultats confirment que les paramètres modaux peuvent être estimés à partir des fonctions du décrément aléatoire ou à partir des réponses impulsionnelles ou encore des fonctions d'autocorrelation. Ainsi, n'importe quelle technique peut être utilisée (Asmussen 1997). Étant donné que chaque voie de mesure (capteur) est post-traitée de manière indépendante (pour rappel, chaque réponse est filtrée autour d'un mode), la technique du décrément logarithmique est utilisée afin d'améliorer l'estimation des fréquences propres (et des taux d'amortissement associés) et des amplitudes modales (normalisées par rapport aux point de référence).

D.2.1 Couplage MRS-MDA

Dans le chapitre 6, la MRS est couplée avec la MDA pour l'extraction des paramètres modaux. Dans ce cas, la MRS est employée pour la détermination des n bandes de filtrage autour des modes (n est le nombre de modes souhaité). Ces bandes sont évaluées à partir des valeurs moyennes des fréquences propres identifiées ($\pm$ 10%). Toutes les voies de mesures (capteurs) sont filtrées à partir de ces bandes et la MDA est appliquée. À partir des fonctions du décrément, les fréquences propres, les taux d'amortissement est les modes propres sont déterminés en utilisant la technique du décrément logarithmique. Chaque pas du décrément est considéré comme une réalisation de la procédure d'identification. Les réalisations sont évaluées afin de fournir des estimations (moyennes et écarts-types) pour les paramètres modaux identifiés. Les fréquences propres et taux d'amortissement sont alors évalués par

$$f_k = \frac{1}{T_{k+2} - T_k}; \varsigma_k \approx \frac{1}{2\pi} \ln\left(\frac{R_k}{R_{k+2}}\right) \tag{D.24}$$

D.3 Méthode d'Ibrahim

La méthode d'Ibrahim est couplée avec la MDA pour la détermination des paramètres modaux dans le cas des simulations numériques présentées dans le chapitre 4. La méthode d'Ibrahim est une technique temporelle d'identification modale basée sur une réponse transitoire. En utilisant deux matrices déduites de la réponse transitoire, Ibrahim a présenté une technique d'identification des paramètres modaux structuraux (Ibrahim & Milkucik 1977). L'une des matrices est la matrice de réponse et l'autre est celle décalée par rapport au temps. Par l'étude d'un problème aux valeurs propres, les paramètres modaux de la structure peuvent être estimés.

D.3.1 Lien avec la MDA

Comme décrit au paragraphe D.2, les fonctions de décrément aléatoire sont proportionnelles aux fonctions d'auto-corrélation et ces fonctions sont aussi équivalentes aux réponses libres de la structure. Ainsi, les fonctions de décrément aléatoire peuvent être considérées comme une réponse libre de la structure à partir de laquelle les caractéristiques modales de la structure peuvent être estimées. Dans ce paragraphe, les étapes d'extraction des paramètres modaux sont expliquées.

La réponse transitoire d'une structure à N degrés de liberté peut être obtenue en considérant l'expression :

$$[M]\{\ddot{y}(t)\} + [D]\{\dot{y}(t)\} + [K]\{y(t)\} = \{f(t)\} \tag{D.25}$$

où $[M], [D], [K]$ ($\in \mathbb{R}^{N\times N}$) sont respectivement les matrices de masse, d'amortissement et de rigidité, $\{y(t)\}$ le vecteur de déplacement et $\{f(t)\}$ l'excitation.

La solution de l'Équation (D.25) mesurée à la position n et à l'instant t peut être écrite par

(Ewins 2000)

$$\{y\}_n(t_j) = \sum_{r=1}^{2N} \phi_{nr} e^{s_r t_j}, \tag{D.26}$$

où ϕ_{nr} est la $n^{\text{ème}}$ composante du $r^{\text{ème}}$ vecteur propre et s_r la $r^{\text{ème}}$ paire de racines conjuguées du système, donnée par

$$s_r = a_r \pm I b_r = -\xi_r \omega_r \pm I \omega_r \sqrt{1-\xi_r^2}, \tag{D.27}$$

où $I = \sqrt{-1}$, ξ_r est le $r^{\text{ème}}$ taux d'amortissement, ω_r est la $r^{\text{ème}}$ fréquence propre correspondant au $r^{\text{ème}}$ vecteur propre.

En considérant que la réponse du système est prélevée sur p voies de mesure avec q points temporels, l'Équation (D.26) pour un système de dimension $p \times q$ est donnée par

$$\begin{bmatrix} y_1(t_1) & y_1(t_2) & \dots & y_1(t_q) \\ y_2(t_1) & y_2(t_2) & \dots & y_2(t_q) \\ \vdots & \vdots & \ddots & \vdots \\ y_p(t_1) & y_p(t_2) & \dots & y_p(t_q) \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1\ 2N} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2\ 2N} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{p1} & \phi_{p2} & \dots & \phi_{p\ 2N} \end{bmatrix} \begin{bmatrix} e^{s_1 t_1} & e^{s_1 t_2} & \dots & e^{s_1 t_q} \\ e^{s_2 t_1} & e^{s_2 t_2} & \dots & e^{s_2 t_q} \\ \vdots & \vdots & \ddots & \vdots \\ e^{s_{2N} t_1} & e^{s_{2N} t_2} & \dots & e^{s_{2N} t_q} \end{bmatrix}$$

ou, en notation matricielle,

$$[Y]_{(p\times q)} = [\Phi]_{(p\times 2N)} [\Lambda]_{(2N\times q)}, \tag{D.28}$$

avec $[Y]$ la matrice de réponses (connue), $[\Phi]$ la matrice des vecteurs propres (inconnue) et $[\Lambda]$ la matrice des valeurs propres (inconnue) aux instants t_j (connus).

Un deuxième système d'équations est crée à partir de l'Équation (D.26), en considérant un décalage temporel Δt :

$$\{y\}_n(t_j + \Delta t) = \sum_{r=1}^{2N} \phi_{nr} e^{s_r (t_j + \Delta t)} \tag{D.29}$$

ou

$$\{\hat{y}\}_n(t_j) = \sum_{r=1}^{2N} \hat{\phi}_{nr} e^{s_r t_j}, \tag{D.30}$$

avec

$$\{\hat{y}\}_n(t_j) = \{y\}_n(t_j + \Delta t) \quad \text{et} \tag{D.31}$$

$$\hat{\phi}_{nr} = \phi_{nr} e^{s_r \Delta t}, \tag{D.32}$$

en d'autres termes

$$[\hat{Y}]_{(p\times q)} = [\hat{\Phi}]_{(p\times 2N)} [\Lambda]_{(2N\times q)}, \tag{D.33}$$

où $[\hat{Y}]$ et $[\hat{\Phi}]$ sont les matrices des réponses et des vecteurs propres décalés par rapport au temps, respectivement.

En considérant $p = 2N$, il est possible d'inverser les matrices $[\Phi]$ et $[\hat{\Phi}]$. En cela,

$$[\Lambda]_{(2N\times q)} = [\Phi]^{-1}_{(2N\times 2N)} [Y]_{(2m\times q)}, \quad \text{et} \tag{D.34}$$

$$[\Lambda]_{(2N\times q)} = [\hat{\Phi}]^{-1}_{(2N\times 2N)} [\hat{Y}]_{(2N\times q)}. \tag{D.35}$$

En égalant l'Équation (D.35) à l'Équation (D.34), il vient

$$[\Phi]^{-1}[Y] = [\hat{\Phi}]^{-1}[\hat{Y}] \Rightarrow [A][Y] = [\hat{Y}], \tag{D.36}$$

où

$$[A] = [\hat{\Phi}][\Phi]^{-1}. \tag{D.37}$$

La matrice $[A]$ est nommée matrice d'état du système, étant donné qu'elle emmène la matrice $[Y]$ de l'instant t à l'instant $t + \Delta t$. En post-multipliant l'Équation (D.36) par $[Y]^T$, il vient :

$$[A][Y][Y]^T = [\hat{Y}][Y]^T. \tag{D.38}$$

En post-multipliant maintenant l'Équation (D.38) par $([Y][Y]^T)^{-1}$

$$[A] = ([\hat{Y}][Y]^T)([Y][Y]^T)^{-1}. \tag{D.39}$$

Ainsi, la matrice $[A]$ est définie uniquement en fonction des mesures (les réponses de la MDA).

En remplaçant les Équations (D.28) et (D.33) dans l'Équation (D.36), il vient :

$$[A]([\Phi][\Lambda]) = [\hat{\Phi}][\Lambda] \tag{D.40}$$

et en considérant l'Équation (D.32) :

$$[A]([\Phi][\Lambda]) = [\Phi]e^{s_r \Delta t}[\Lambda] \Rightarrow (\mathbf{A} - e^{s_r \Delta t}\mathbf{I})[\Phi] = \mathbf{0}. \tag{D.41}$$

où $[I]$ est la matrice identité.

L'Équation (D.41) est un problème aux valeurs propres. Une fois les valeurs et vecteurs propres de la matrice $[A]$ déterminés, il est alors possible d'évaluer les fréquences propres, les taux d'amortissement et les déformées modales du système.

D.3.2 Extraction des paramètres modaux

À partir des Équations (D.27) et (D.41), les fréquences propres et les taux d'amortissement peuvent être évalués. En écrivant les valeurs propres extraites de la matrice $[A]$ comme

$$e^{s_r \Delta t} = \beta_r + I\gamma_r, \tag{D.42}$$

où I est l'unité imaginaire et en utilisant la notation $s_r = a_r + Ib_r$, il vient

$$e^{(a_r + Ib_r)\Delta t} = \beta_r + I\gamma_r. \tag{D.43}$$

En considérant la relation d'Euler,

$$e^{I\alpha} = \cos\alpha + I\sin\alpha \tag{D.44}$$

il est possible d'écrire

$$\beta_r + I\gamma_r = \rho(\cos\alpha + I\sin\alpha) = \rho e^{I\alpha}, \tag{D.45}$$

où $\rho = \sqrt{\beta_r^2 + \gamma_r^2}$ et $\alpha = \arctan\left(\dfrac{\gamma_r}{\beta_r}\right)$ sont, respectivement, le module et la phase entre la composante réelle et imaginaire d'un complexe

En remplaçant l'Équation (D.45) dans l'Équation (D.43) avec les valeurs ρ et α, il vient

$$e^{(a_r + Ib_r)\Delta t} = \left(\sqrt{\beta_r^2 + \gamma_r^2}\right)\ e^{\left[I \arctan\left(\frac{\gamma_r}{\beta_r}\right)\right]}. \tag{D.46}$$

En appliquant le logarithme naturel aux deux membres de l'Équation (D.46) :

$$(a_r + Ib_r)\Delta t = \ln\left(\sqrt{\beta_r^2 + \gamma_r^2}\right) + I \arctan\left(\frac{\gamma_r}{\beta_r}\right). \tag{D.47}$$

En égalant les parties réelles et imaginaires, il vient :

$$a_r = \frac{1}{2\Delta t} \ln\left(\beta_r^2 + \gamma_r^2\right), \quad \text{et} \tag{D.48}$$

$$b_r = +\frac{1}{\Delta t} \arctan\left(\frac{\gamma_r}{\beta_r}\right). \tag{D.49}$$

Une fois les coefficients a_r e b_r déterminés, les fréquences propres et taux d'amortissement sont obtenus à partir de l'Équation (D.27) :

$$\omega_r = \sqrt{a_r^2 + b_r^2}, \quad \text{et} \tag{D.50}$$

$$\xi_r = -\frac{a_r}{\omega_r}. \tag{D.51}$$

Les déformées modales sont extraites à partir des vecteurs propres de la matrice $[A]$. Les modes propres complexes sont écrits par

$$\{\phi\}_r = \{c\}_r + I\{d\}_r, \tag{D.52}$$

où $\{c\}_r$ et $\{d\}_r$ sont les vecteurs contenant les parties réelles et imaginaires du mode propre $\{\phi\}_r$, respectivement.

Le module de la $n^{\text{ème}}$ composante du $r^{\text{ème}}$ mode propre réel $\{\phi\}_{nr}^*$ est obtenu par

$$\phi_{nr}^* = \sqrt{c_{nr}^2 + d_{nr}^2}, \tag{D.53}$$

où c_{nr} et d_{nr} représentent la $n^{\text{ème}}$ composante du $r^{\text{ème}}$ vecteur contenant les paires réelles et imaginaires du mode propre ϕ_{nr}^*, respectivement.

Toutefois, la valeur obtenue par l'Équation (D.53) doit être multipliée par une constante η, donnée par

$$\eta = \begin{cases} -1, \text{ si } \frac{\pi}{2} < \alpha < \frac{3\pi}{2}, \\ 1, \text{ au cas contraire.} \end{cases} \tag{D.54}$$

où $\alpha = \arctan\left(\frac{d_{nr}}{c_{nr}}\right)$.

Une difficulté de la méthode d'Ibrahim est liée au besoin de disposer de deux voies de mesures pour la détermination d'une fréquence propre (et du taux d'amortissement et du mode propre associés). Cette limitation peut être contournée à partir de la manipulation des signaux, en les découpant en plusieurs *pseudo mesures*.

Un signal constitué de 100 valeurs temporelles peut, par exemple, être divisé en 4 parties (1 signal et 3 pseudo mesures) avec 25 valeurs chacune. Une autre manière de créer des mesures artificielles est de décaler les signaux. Dans ce cas, le même signal avec 100 valeurs temporelles est découpé de telle sorte que

- Signal 1 : valeurs temporelles de 1 a 70 ;
- Pseudo mesure 2 : valeurs temporelles de 11 a 80 ;
- Pseudo mesure 3 : valeurs temporelles de 21 a 90 ;
- Pseudo mesure 4 : valeurs temporelles de 31 a 100 ;

Dans cette méthode, les réalisations (estimations) des paramètres modaux sont déterminées à partir du nombre de pseudo mesures artificiellement créées.