

# Application de l'Analyse des Correspondances Ordinales au suivi d'espèces végétales aquatiques.

Claude Manté, Guillaume Bernard, Patrick Bonhomme

► **To cite this version:**

Claude Manté, Guillaume Bernard, Patrick Bonhomme. Application de l'Analyse des Correspondances Ordinales au suivi d'espèces végétales aquatiques.. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494699>

**HAL Id: inria-00494699**

**<https://hal.inria.fr/inria-00494699>**

Submitted on 24 Jun 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# APPLICATION DE L'ANALYSE DES CORRESPONDANCES ORDINALES AU SUIVI D'ESPÈCES VÉGÉTALES AQUATIQUES.

Claude Manté<sup>1</sup>, Guillaume Bernard<sup>2</sup> et Patrick Bonhomme<sup>3</sup>

<sup>1</sup>LMGEM, UMR CNRS 6117, Campus de Luminy - 13009 Marseille

<sup>2</sup>GIPREB, Cours Mirabeau - 13130 Berre-l'étang

<sup>3</sup>GIS Posidonie, Campus de Luminy - 13009 Marseille

## Résumé

Les communautés de macrophytes (algues et phanérogames), dont certaines espèces sont protégées, constituent un des indicateurs clés de l'état écologique des lagunes, tel que défini par la Directive Européenne Cadre sur l'Eau (DCE). Par ailleurs, la restauration de ces communautés de macrophytes, et des herbiers de *Zostera* en particulier, est l'un des objectifs principaux du plan de réhabilitation de l'Etang de Berre (Bouches-du-Rhône) porté par le Gipreb (Groupement d'intérêt public pour la réhabilitation de l'Etang de Berre). C'est pourquoi celui-ci met en œuvre, depuis 1996, le suivi annuel des principales espèces de macrophytes encore présentes dans l'étang, afin d'évaluer l'impact sur ce compartiment de mesures de régulation des rejets d'eau douce d'une centrale hydroélectrique, principale source de perturbation du système.

Une suite de onze tableaux annuels relatifs à la densité de l'espèce *Zostera noltii* est ici étudiée via l'Analyse des Correspondances Ordinales de Beh (1997, 2004), dont les principes sont rappelés. Les résultats obtenus seront comparés avec ceux de l'AFC classique.

*Mots clés : Analyse des données, Environnement*

## Abstract

Macrophytes communities, (seaweeds and seagrass), some of them are protected, consist in a key indicator of the ecological health of lagoons, as stated by the European Water Framework Directive. Furthermore, the restoration of these communities, especially the *Zostera* meadows is one of the main objectives of the Berre lagoon restoration plan, led by the Gipreb (Groupement d'intérêt public pour la réhabilitation de l'Etang de Berre). Consequently, this organism is monitoring since 1996 the main macrophytes species still present in the lagoon, in order to evaluate the impact of freshwater releases from an electric power plant on this ecosystem.

A sequence of eleven annual tables of density of *Zostera noltii* is studied through the Ordinal Correspondence Analysis of Beh (1997, 2004), whose principles are first recalled. The obtained results are compared with those of classical Correspondence Analysis.

*Keywords : Data analysis, Environment*

## Introduction

L'étang de Berre, une des plus grandes lagunes méditerranéennes, est soumise depuis 1966 à des apports massifs d'eau douce et de limons par la centrale hydroélectrique de St-Chamas. Il en a résulté des modifications brutales de la salinité et de la transparence des eaux de surface, ce qui a conduit à la régression des communautés de macrophytes de l'étang. Ces peuplements sont suivis depuis 1996 par le Gipreb: la localisation et l'abondance de 8 espèces ou groupes d'espèces de macrophytes sont ainsi annuellement relevés (en période estivale, période de plus forte végétation) le long du littoral de l'étang. La densité de chaque espèce ou groupe d'espèces a été évaluée visuellement par des plongeurs le long de 31 transects (composés chacun de 20 **segments** de 5 m) répartis régulièrement sur l'ensemble du pourtour de l'étang, et ce pendant 11 années (1996 à 2008). L'évaluation de la densité consiste en un **codage en 6 modalités** (cf. figure 4) du taux de couverture du fonds par l'espèce. L'état annuel de la population de chaque espèce est donc décrit par un tableau de type 6x20, croisant la densité avec la position : deux variables ordinales. Il s'agit pour nous de

tester aussi rigoureusement que possible l'existence d'effets spatio-temporels pour chaque macrophyte. Nous tenterons donc, pour l'espèce *Zostera noltii* (espèce reconnue d'importance patrimoniale, protégée et qui par sa présence indique un bon fonctionnement écologique du système lagunaire), d'apporter une réponse aux questions suivantes :

1. la densité de l'espèce dépend-elle de la distance à la côte ?
2. si oui, quelles sont les zones du transect importantes pour sa dynamique?
3. y a-t-il eu un changement à partir de 2005 (année n°8 et suivantes), année à compter de laquelle des nouvelles régulations des apports d'eau douce par la centrale hydroélectrique de Saint-Chamas ont été appliquées ?

Une manière classique d'analyser de ce genre de données est de procéder à l'AFC du tableau  $T$  de type  $6 \times 20$  associé à chaque espèce, obtenu en sommant les onze tableaux annuels. Cette méthode (Benzécri, 1967 ; Beh, 2004) permet d'analyser un tableau de contingence  $T$  à  $I$  lignes et  $J$  colonnes, d'effectif total  $n$  (ici,  $I=6$ ,  $J=20$ , et  $n=11 \times 31 \times 20=6820$ ). On note  $P := T/n$  la probabilité associée, et  $P_i := \sum_{j=1}^J P_{i,j}$  (resp.  $P_j := \sum_{i=1}^I P_{i,j}$ ) le poids de la  $i^{\text{ème}}$  ligne (resp. de la  $j^{\text{ème}}$  colonne). L'AFC met en évidence la manière dont la relation entre lignes et colonnes de  $T$  s'écarte de

l'indépendance, en décomposant le tableau  $\Theta$  de terme général  $\theta_{i,j} := \frac{P_{i,j}}{P_i \times P_j}$ . Dans la pratique,

$X^2 := n \sum_{i,j} P_i \times P_j (\theta_{i,j} - 1)^2$  suit (sous hypothèse d'indépendance) une loi  $\chi^2((I-1)(J-1))$ , et l'AFC consiste essentiellement à déterminer des couples d'axes factoriels  $(A_m, B_m) : 1 \leq m \leq M$ ,

avec  $M := \text{Min}[I-1, J-1]$ , tels que l'on ait :  $\theta_{i,j} = \sum_{m=0}^M A_{m,i} \lambda_m B_{m,j}$ . On a alors les relations

fondamentales :  $X^2 = n \sum_{m=1}^M \lambda_m^2$ , et :  $\lambda_m \delta_m^p = \sum_{(i,j) \in I \times J} A_{m,i} P_{i,j} B_{p,j}$ .

## La méthode de Beh pour les tableaux ordonnés

Mais les facteurs d'une AFC ne sont pas toujours faciles à interpréter, et surtout leur construction ne dépend pas de l'ordre des modalités. C'est pourquoi Beh (1997, 2004) a proposé une méthode alternative de décomposition de  $X^2$ , adaptée au cas où les lignes et/ou les colonnes du tableau traité sont **ordonnées**.  $X^2$  n'est plus alors décomposée en fonction d'axes factoriels dépendant de l'ensemble du tableau analysé, mais en fonction de **polynômes** dépendant des variables ordinales structurant le tableau au travers de **scores**, et du tableau au travers de ses **marges**  $\{P_i : 1 \leq i \leq I\}$  et  $\{P_j : 1 \leq j \leq J\}$  uniquement. Ces polynômes, ici notés  $A_m$  pour les lignes (densité) et  $B_p$  pour les colonnes (segments), jouent le rôle de facteurs. Comme des axes factoriels, ils sont centrés et orthonormés. Par contre, ce ne sont pas des vecteurs singuliers de  $\Theta$ , ce qui fait qu'il est impossible de parler de **paires d'axes** associés à une part de variance commune, comme c'est le cas en AFC. Pour la même raison, des polynômes de degrés différents peuvent être corrélés. Si l'on note  $\mu$  (resp.  $\eta$ ) les parts de variance associés aux polynômes sur la position (resp. la densité), nous

savons seulement que :  $\frac{X^2}{n} = \sum_{m=1}^{I-1} \eta_m = \sum_{p=1}^{J-1} \mu_p$ . On a donc ici deux autres manières de décomposer  $X^2$ , distinctes de celle de l'AFC.

## Les tests

Il est bien connu qu'il n'existe pas de test rigoureux de significativité des axes factoriels en AFC, en raison de la complexité de la loi des valeurs propres associées. A notre connaissance, il n'existe guère que la procédure classique basée sur le cumul des valeurs propres proposée par Benzécri (1976). En revanche, dans le cas de la méthode de Beh, il est possible de

tester l'effet des polynômes et de leurs interactions, quantifiées par des **moments bivariés**. Le moment bivarié d'ordre  $(m,p)$  est : (Beh, 2004)  $\lambda_{m,p} := \sum_{(i,j) \in I \times J} A_{m,i} P_{i,j} B_{p,j}$ . Remarquons que cette formule est formellement identique à celle donnant la corrélation entre les axes factoriels de l'AFC. Mais ici les axes (polynômes sur les deux ensembles) ne sont plus décorrélés : toutes les interactions sont possibles. Rayner et Best (1996) ont proposé des tests pour ces moments, en démontrant qu'ils sont asymptotiquement indépendants et de loi  $N(0,1)$ , sous hypothèse d'indépendance des lignes et des colonnes. On en déduit une statistique permettant sous cette hypothèse de tester la significativité du polynôme  $A_m, \forall 1 \leq m \leq I-1$  :  $\eta_m = \sum_{p=1}^{J-1} \lambda_{m,p}^2 \approx \chi_{(J-1)}^2$ . On a de même pour les colonnes la statistique :  $\forall 1 \leq p \leq J-1, \mu_p = \sum_{m=1}^{I-1} \lambda_{m,p}^2 \approx \chi_{(I-1)}^2$ .

### Un exemple : l'analyse des transects de Zostères

*Zostera noltii* est une espèce qualifiée dans l'étang de Berre de "pseudo-annuelle", les herbiers qu'elle constitue présentent dont un caractère pérenne d'une année à l'autre, en terme au moins de localisation géographique et d'extension. Le recouvrement sur le fond des faisceaux de feuilles, en revanche, connaît des fluctuations saisonnières et interannuelles, qui témoignent de la vitalité et de la dynamique de l'espèce d'une année à l'autre. Dans l'étang de Berre, son extension vers le large semble contrainte par la profondeur, directement en lien avec la lumière disponible sur le fond (Bernard et al., 2007). Les zostères étant totalement absentes des 10 derniers segments des transects, nous les avons supprimés de cette analyse, et donc exceptionnellement  $J=10$ .

Le test du Chi-deux global ( $P(\chi_{(45)}^2 > X^2 = 84.12) \approx 3.10^{-4}$ ) montre que la densité et la distance au rivage sont en relation forte, et le test sur la trace ne donne pour significatif que le 1<sup>er</sup> facteur de l'AFC (72.6% de l'inertie), le 2<sup>ème</sup> correspondant à un taux d'inertie de 12.%.

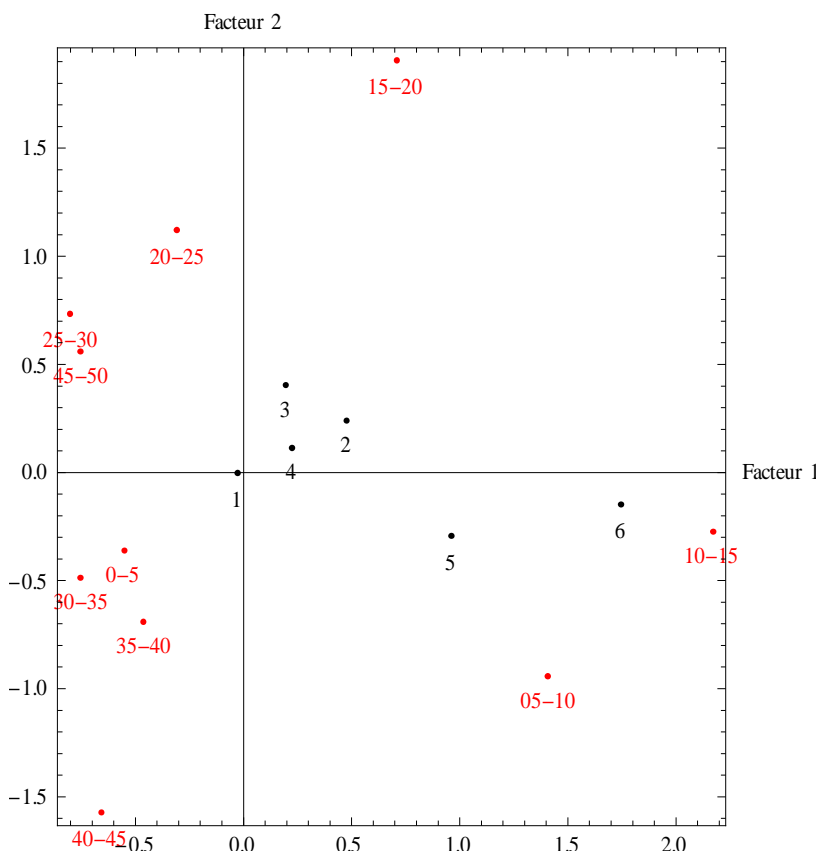


Figure 1 : 1er plan de l'Analyse des Correspondances classique

Il apparaît clairement sur ce plan que la densité « 1 » (nulle) est très fréquente (proche de la

moyenne), alors que les fortes densités de Zostères sont associées aux segment 5-10m et surtout 10-15m.

### Les scores et les polynômes orthonormés

On remarque sur la figure 2 ci-dessous que le poids des segments, contrairement à celui des niveaux, est constant. En effet, chaque colonne de chaque tableau à toujours pour effectif le nombre de stations, 31.

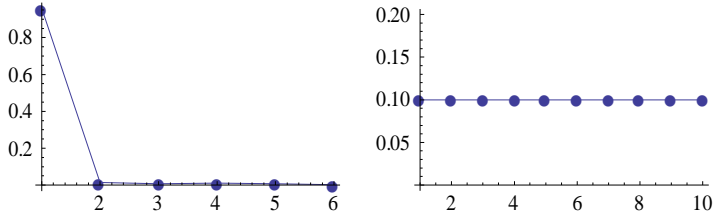


Figure 2 : poids des niveaux de densité et des segments

Beh (1998) quantifie la différence entre les modalités consécutives d'une des variable ordinales au moyen d'un score  $S$ . Le score « naturel » est  $S(j) := j$ , où la distance entre modalités est constante. C'est celui que nous avons choisi pour coder la position le long du transect, car les segments considérés sont de même largeur. Les polynômes orthonormés sur les transects,  $\{B_1, \dots, B_9\}$  ont donc une allure parfaitement symétrique : ce sont (en gros) des polynômes de Legendre discrétisés et changés d'échelles.

La situation est différente pour la variable "densité". D'abord, notons (voir ci-dessus) que le poids de la première modalité est proche de 1 (0.96) : très fréquemment, les Zostères sont absentes. Pour ce qui est du score, nous avons associé à chaque classe le **taux de recouvrement central** évalué par le plongeur (voir Bonhomme *et al.* (2008)), ce qui donne la figure 3 ci-dessous. La non-linéarité du score, et le poids considérable de la modalité « 1 », font que les polynôme relatifs à cette variable ne sont plus du tout symétriques.

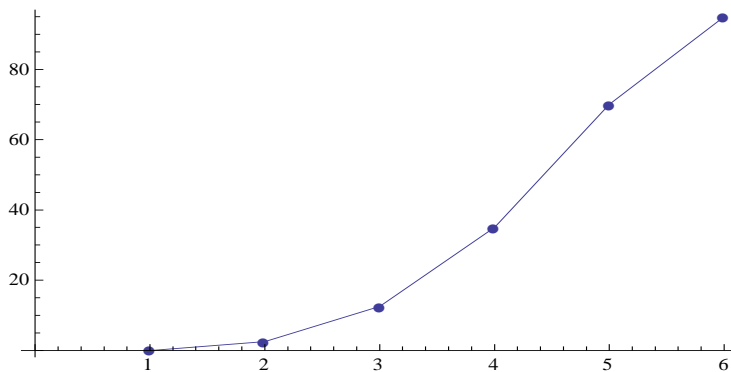


Figure 3 : score associé au taux de recouvrement

Remarquons maintenant que sur la figure 4 les taux d'inertie de cette analyse ne sont pas systématiquement décroissants, contrairement à ce qui se produit en AFC. Ces graphiques montrent que les variations de structure des transects sont expliquées le mieux possible par  $\{A_1, A_5\}$ , alors que les variations de densité sont expliquées par les polynômes  $\{B_1, B_3, B_4\}$ .

Tests du  $\chi^2$  sur les composantes de Beh (Rouge: 0.95, Vert: médiane)

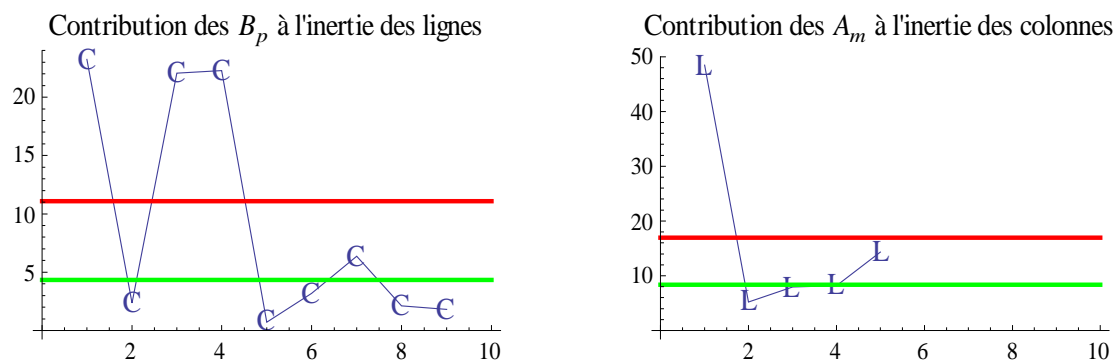


Figure 4 : épreuve de validité des composantes de Beh

Les tests portant sur les interactions permettent davantage de précision. Ils sont résumés par la figure 5 ci-dessous, représentant la matrice  $\Lambda$  des moments bivariés. La case de coordonnées  $(m,p)$  est blanche lorsque l'hypothèse :  $\lambda_{m,p} \approx N(0,1)$  n'est pas rejetée au seuil 0.05 (test bilatère) ; sinon elle est d'autant plus rouge que l'interaction est statistiquement significative.

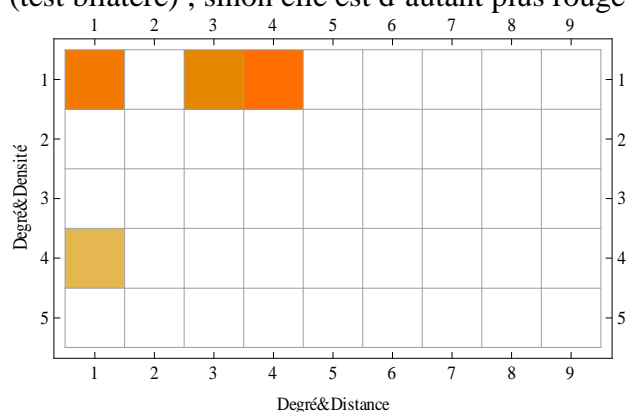


Figure 5 : test de significativité des éléments de la matrice  $\Lambda$ .

D'après la figure 4, le plan factoriel le plus pertinent (74% de la variance) pour projeter les segments est le plan  $(1,5)$ , représenté sur la figure 6. Sur ce plan, les segments sont représentés en abscisse (resp. en ordonnée) par la composante de Beh (1997)  $G_*(1)$  (resp.  $G_*(5)$ ) dont la  $j^{\text{ième}}$  coordonnée est la projection du  $j^{\text{ième}}$  segment sur  $A_1$  (resp.  $A_5$ ).

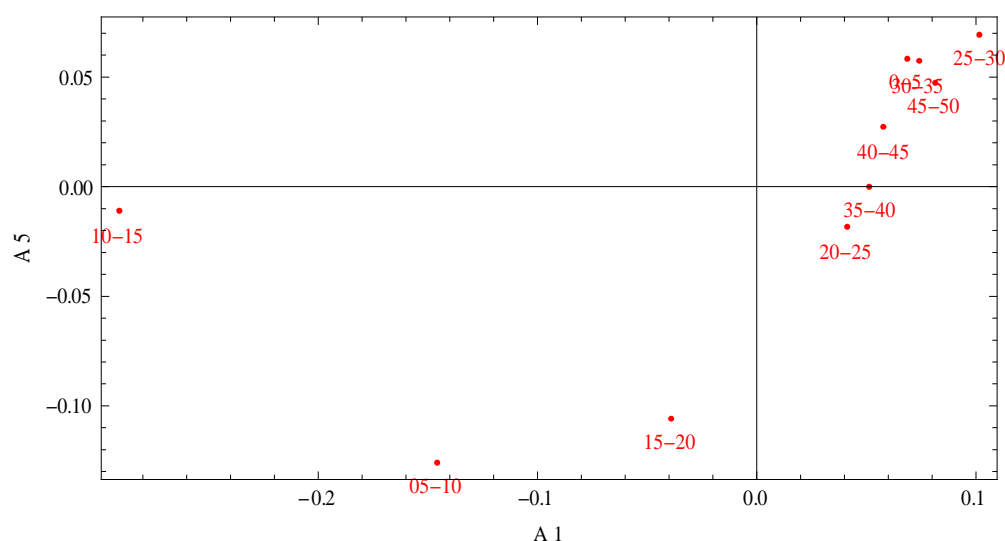


Figure 6 : plan  $(1,5)$  de l'analyse ordinale de Beh.

Enfin, il est aussi possible d'introduire le temps dans l'analyse, en remarquant que

$$P = \sum_{a=1}^{11} \alpha_a P^a, \text{ où } P^a \text{ désigne la probabilité associée au tableau partiel des mesures de l'année}$$

« a », et  $\alpha_a = 1/11$  (les tableaux partiels ont ici tous la même masse). Il s'ensuit que la position de chaque moyenne temporelle est au barycentre des points annuels correspondant à ce segment. Cela permet de tracer une ultime figure (non représentée, faute de place) mettant en évidence le(s) segment(s) où la dynamique temporelle est la plus agitée, et une éventuelle rupture en 2005 (année 8 et suivantes).

### +Conclusions :

L'analyse en correspondance ordinale s'avère être un outil intéressant pour le suivi à long terme des communautés littorales (ici les macrophytes). Appliquée aux herbiers de zostères (*Z. Noltii*), elle permet de mettre clairement en évidence la dépendance de l'abondance des Zostères en un point avec sa distance au rivage. Cela confirme la contrainte liée à la profondeur (et donc à l'éclairement) pour leur développement et leur extension vers le large. En première approximation, leur densité décroît le long du transect (car  $\lambda_{1,1} = -3.71$ ). Cependant, comme on peut le voir sur la figure 5, il existe aussi des interactions de degrés 3 et 4 significatives. Les segments où la variabilité temporelle des Zostères est maximale sont ceux correspondant à une distance au rivage de 0-25m, surtout le segment 10-15m (très proches de la côte). En terme d'évolution temporelle, d'après l'interprétation de la 1<sup>ère</sup> composante, la régression de l'espèce observée depuis 1996 se poursuit en 2005, et après. Toutefois, on ne peut constater de changement clair que sur le segment 10-15m.

### Bibliographie

- [1] Beh, E. J. (1997) Simple Correspondence Analysis of ordinal cross-classifications using orthogonal polynomials, *Biometrical Journal*, 39, 5, 589-613.
- [2] Beh, E. J. (1998) A comparative study of Scores for Correspondence Analysis with ordered categories, *Biometrical Journal*, 40, 4, 413-429.
- [3] Beh, E. J. (2004) Simple Correspondence Analysis : a bibliographic review, *International Statistical Review*, 72, 2, 257-284.
- [4] Benzécri, J. P. (1976) *L'Analyse des Données*, tome 2, 2<sup>ème</sup> ed., Dunod.
- [5] Bernard G., Boudouresque CF., Picon P. (2007) Long term changes in *Zostera* meadows in the Berre lagoon (Provence, Mediterranean Sea). *Estuarine, Coastal and Shelf Science*, 73:617-629.
- [5] Bonhomme, P., Gravel, R. Ruitton, S., Bonhomme, Boudouresque, C.-F. (2008) *Système de surveillance des principales espèces de macrophytes dans les étangs de Berre et de Vaïne : Suivi 2007*, Contrat GIS Posidonie/GIPREB. GIS Posidonie publ., Marseille, Fr. : 1- 47+ Annexes 1-3.
- [6] Rayner, J.C.W. and Best, D.J. (1996) Smooth extensions of Pearson's product moment correlation and Spearman's Rho, *Statistics & Probability Letters*, 30, 171-177.