

MODÉLISATION STATISTIQUE DES CHANGEMENTS CLIMATIQUES, DÉTECTION, ET ATTRIBUTION

Aurélien Ribes

CNRM-GAME, Météo France - CNRS

*CNRM/GMGEC/VDR
42, Av. Gaspard Coriolis
31057 Toulouse*

Mots-clés : Modélisation statistique, climat, environnement, détection, attribution.

Résumé : Dans le cadre des études de détection et d’attribution, différents outils statistiques sont utilisés afin d’étudier les changements climatiques en cours. La détection d’un changement, tout d’abord, consiste à montrer qu’un phénomène est effectivement un changement ; on montre, statistiquement, que ce phénomène est incohérent avec la seule variabilité interne du système climatique, considérée comme aléatoire. L’attribution d’un changement à une ou plusieurs causes consiste à établir un lien de causalité entre différents facteurs explicatifs physiquement plausibles et le changement étudié. Nous présentons ici un bref descriptif des modèles et de la démarche statistiques mis en œuvre dans ce cadre, qui accordent une place importante aux tests d’hypothèses. Nous introduisons ensuite quelques unes des problématiques statistiques pouvant être rencontrées pour mener à bien ces études.

Abstract : Statistical tools are widely used in order to study climate changes, in the context of detection and attribution studies. First, the detection of a change involves proving the existence of a change. It is a statistical demonstration that an observed phenomenon is not consistent with the internal climate variability alone, considered as random. Second, the attribution of a change to one or several causes involves showing a causal link between various physically plausible predictors and the investigated change. We here give a short presentation of the main statistical models and approaches implemented for these issues. Hypothesis testing are particularly used in this context. We also introduce some statistical problems that can be of interest for detecting or attributing climate changes.

Résumé long :

Depuis sa création en 1988, l’une des principales questions assignées au Groupe d’experts Intergouvernemental sur l’Évolution du Climat (GIEC, ou IPCC en anglais) concerne

la quantification de l'impact des activités humaines dans la tendance au réchauffement observée en moyenne sur la planète. 20 ans plus tard, les notions de détection des changements climatiques, et d'attribution de ces changements à certaines causes, ont été avancées pour apporter des éléments de réponse ; ces techniques sont avant tout basées sur des arguments statistiques (Hegerl et al., 2007).

Avant de présenter le formalisme mathématique propre aux études de détection et d'attribution, il convient de délimiter les contours de l'objet d'étude principal : le système climatique. Le climat (à l'échelle planétaire comme localement) résulte de l'interaction de différents composants : l'atmosphère, l'hydrosphère (en particulier les océans), la cryosphère (glaciers et eau solide en général), la biosphère, les surfaces continentales. Le système ainsi constitué possède une certaine dynamique, dont résulte le *temps*, lequel varie, d'un jour à l'autre, d'une année à l'autre, ou encore d'une décennie à l'autre.

Par ailleurs, ce système est soumis à plusieurs facteurs, appelés forçages, qui lui sont extérieurs mais peuvent influencer sa dynamique. Parmi ces forçages on distingue généralement les influences naturelles des influences anthropiques. On peut citer, en particulier, les variations de l'activité solaire, les variations astronomiques de l'orbite terrestre, ou encore, l'effet de serre d'origine anthropique ; mais la liste est en fait bien plus longue (IPCC, 2007).

Cette description succincte de l'objet d'étude est suffisante pour introduire le formalisme statistique des études de détection et d'attribution. Considérons que l'on dispose d'un échantillon d'observations climatiques $(y_{s,t})_{1 \leq s \leq p, 1 \leq t \leq n}$, faites en différents lieux s , et à différents instants t (t désigne typiquement une année). Ces observations peuvent être de différentes natures : températures, cumuls de précipitations, etc. En l'absence de toute influence extérieure, considérons le modèle de référence suivant :

$$y_{s,t} = m_s + \varepsilon_{s,t}, \tag{1}$$

dans lequel on sépare simplement l'espérance m_s et un terme aléatoire centré traduisant la variabilité interne du système, $\varepsilon_{s,t}$. Sans chercher à le justifier ici (voir Ribes, 2009 pour une discussion plus approfondie), nous considérons que ce terme aléatoire ε , est assimilable à une variable aléatoire gaussienne centrée, vérifiant :

$$\text{Cov}(\varepsilon_{s,t}, \varepsilon_{s',t'}) = C_{s,s'} \delta_{t,t'}, \tag{2}$$

où C est une matrice de taille $p \times p$ définie positive inconnue, et $\delta_{t,t'} = \mathbb{1}_{t=t'}$. On obtient de cette façon une modélisation statistique de la variable climatique étudiée, toujours en l'absence de tout forçage externe.

D'autres hypothèses peuvent être faite concernant, cette fois-ci, la forme que peut prendre la réponse du système à une influence extérieure, c'est-à-dire le choix du modèle statistique en présence d'un ou plusieurs forçages. Classiquement, nous nous concentrerons ici sur l'étude du modèle additif, dans lequel la perturbation induite par les forçages porte sur l'espérance de y , et dans lequel les effets de plusieurs forçages sont la somme des effets

de chacun des forçages à l'œuvre. En considérant K forçages distincts, et en notant $\psi_{s,t}^{(k)}$ la réponse du système, au lieu s et à l'instant t , au k -ième forçage, on obtient le modèle

$$y_{s,t} = m_s + \sum_{k=1}^K \psi_{s,t}^{(k)} + \varepsilon_{s,t}. \quad (3)$$

On obtient ainsi une modélisation statistique du système sous l'influence de forçages.

Dans un premier temps, l'étude d'un éventuel *changement* du climat se porte sur sa mise en évidence statistique, ou son existence. C'est la question de la détection. On cherche à savoir si les observations étudiées sont cohérentes avec la variabilité interne du système, ou si au contraire, elles témoignent de la présence d'une influence extérieure. Cette démonstration prend généralement la forme d'un test d'hypothèse, dans lequel, pour détecter, on cherche à rejeter l'hypothèse nulle selon laquelle le climat est stationnaire. De façon équivalente, il s'agit de construire un test de (1), que l'on assimile ici à l'hypothèse nulle H_0 . Dans l'absolu, un test de détection ne requiert pas formellement de spécifier une hypothèse alternative. La prise en compte d'une alternative adaptée permet néanmoins de construire un test plus puissant ; différentes alternatives plausibles peuvent alors être considérées.

En particulier, il est courant, mais pas indispensable, d'avoir recours à des modèles physiques simulant le système climatique pour préciser cette hypothèse alternative. On considère alors que $\psi^{(k)} = \beta_k \phi^{(k)}$, où $\phi^{(k)}$ est la réponse du système au k -ième forçage, telle que simulée par le modèle, et β_k un paramètre réel inconnu. Cette hypothèse revient à considérer que la réponse du système est connue, à un facteur d'amplitude près. Le modèle statistique devient alors

$$y_{s,t} = m_s + \sum_{k=1}^K \beta_k \phi_{s,t}^{(k)} + \varepsilon_{s,t}, \quad (4)$$

dans lequel les paramètres inconnus sont $\beta = (\beta_1, \dots, \beta_K)$ et C . Dans un tel cadre, un test de détection d'un changement du climat met en jeu les hypothèses :

$$H_0 : \langle \beta = 0_K \rangle, \quad \text{vs} \quad H_1 : \langle \beta \neq 0_K \rangle. \quad (5)$$

Dans un second temps, l'étude d'un changement identifié se porte sur les causes pouvant l'expliquer. L'objectif est alors de montrer le lien de cause à effet entre un ou plusieurs forçages, et le changement identifié (ou détecté). Historiquement, ce problème s'est d'abord posé pour évaluer le rôle joué par les activités humaines dans l'élévation de la température moyenne à la surface de la planète. Plusieurs questions peuvent alors être

posées, parmi lesquelles : peut-on considérer que les activités humaines ont contribué significativement au changement ? Peut-on quantifier les contributions des différentes causes envisagées au réchauffement observé ?

Ce problème est plus délicat que le précédent, principalement parce qu'il demande d'avoir une connaissance des effets des différents facteurs en présence (Hergel, 2007). Dans de nombreuses sciences, cette connaissance est acquise par le biais d'expériences reproductibles. Dans le cas du système climatique, il n'y a qu'une seule réalisation, et donc il est difficile d'observer ou de mesurer directement les effets d'un forçage pris isolément. Le recours à la modélisation numérique est donc plus systématique afin de préciser les effets de chacune des causes. La compréhension des données historiques est également primordiale, mais ne sera pas abordée ici.

Dans le cadre du modèle statistique précédent, le problème de l'attribution consiste à proposer une estimation du vecteur β , afin de préciser les contributions de chacun des termes. Ensuite, afin de montrer le caractère incontestable du rôle joué par un forçage (par exemple celui correspondant à $k = 1$), plusieurs tests peuvent être proposés. Tout d'abord, on va chercher à détecter l'empreinte de ce forçage, soit à rejeter l'hypothèse nulle

$$H_0 : \langle \beta = 0_K \rangle, \quad \text{vs} \quad H_1 : \langle \beta_1 \neq 0, (\beta_2, \dots, \beta_K) = 0_{K-1} \rangle. \quad (6)$$

Ensuite, on cherchera à montrer que les observations étudiées ne sont pas cohérentes avec les effets des autres causes envisageables (ici les forçages $k = 2, \dots, K$). Cela peut être fait en rejetant l'hypothèse nulle

$$H_0 : \langle \beta_1 = 0 \rangle, \quad \text{vs} \quad H_1 : \langle \beta_1 \neq 0 \rangle. \quad (7)$$

Enfin, afin de contribuer à montrer la cohérence des résultats, on vérifie que les observations sont cohérentes avec les différentes causes prises en compte, ce qui peut être fait en montrant qu'un test des hypothèses

$$H_0 : \langle \beta = \mathbb{1}_K \rangle, \quad \text{vs} \quad H_1 : \langle \beta \neq \mathbb{1}_K \rangle, \quad (8)$$

ne peut pas être rejeté.

L'ensemble de ces techniques a fait l'objet de nombreux travaux au sein de la communauté des sciences du climat, travaux ayant contribué aux conclusions présentées par le GIEC (2007). Parmi ces travaux, plusieurs abordent des questions de nature méthodologique, et discutent les techniques d'inférence dans le modèle (4). Parmi ces travaux, on peut citer l'étude de Allen et Stott (2003), proposant de prendre en compte une certaine incertitude sur les paramètres $\phi^{(k)}$ (réputés connus). Un autre exemple concerne l'estimation de la matrice de covariance C , qui peut être rendue délicate en raison de la grande dimension des données. Ce problème est récurrent dès lors que l'on traite des données climatiques spatialisées, et Ribes et al. (2009) proposent une première piste pour rendre la procédure d'estimation plus précise dans un tel cadre.

Bibliographie

- [1] IPCC (2007) Climate change 2007 : The physical science basis. Contribution of the working group I to the fourth assessment report of the international panel on climate change. *Cambridge University Press*, Cambridge, United Kindom, and New York, USA.
- [2] Hegerl, G., Zwiers, F., Braconnot, P., Gillet, N., Luo, Y., Marengo Orsini, J., Nicholls, N., Penner, J., et Stott, P. (2007) Understanding and attributing climate change. In : *Climate change 2007 : The physical science basis. Cambridge University Press*, Cambridge, United Kindom, and New York, USA.
- [3] Ribes, A. (2009) Détection statistique des changements climatiques, *Thèse de l'université Toulouse III - Paul Sabatier*, Toulouse.
- [4] Allen, M. R. et Stott, P. A. (2003) Estimating signal amplitudes in optimal fingerprinting, part I : theory. *Climate Dynamics*, 21, 477–491.
- [5] Ribes, A., Azaïs, J.-M. et Planton, S. (2009) Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance matrix estimate. *Climate Dynamics*, 33 (5), 707–722.