

## **Décrire les résultats de nombreuses simulations : quelques considérations de robustesse et comparaison de méthodes**

Isabelle Alvarez

Résumé :

La modélisation de certains systèmes sociaux ou environnementaux conduit à utiliser des approches individus-centrées et stochastiques. Les résultats de ces modèles sont complexes parce qu'ils sont constitués de grands ensembles de trajectoires. En effet, l'état du système en fonction du temps n'est généralement plus décrit par une variable scalaire, mais par un vecteur, une distribution. De plus, on ne s'intéresse plus seulement à l'état du système à un horizon donné, ou à son comportement asymptotique, mais à l'évolution au cours du temps. C'est le cas par exemple des modèles de populations forestiers. Des difficultés de compréhension des résultats se présentent donc déjà pour les chercheurs (modélisateurs et experts « écologues » concepteurs des modèles) pour connaître le fonctionnement du modèle. On ne parle même pas des difficultés de compréhension des gestionnaires, utilisateurs ultimes des résultats de ces modèles. D'une manière générale, ces modèles posent des problèmes de méthodologie pour leur conception et leur validation. Ici on s'intéressera à la description des résultats.

### 1. Caractériser des courbes d'évolution temporelle

\*\*\*\*\*

Une première simplification du problème de description des résultats peut se faire en examinant des grandeurs agrégées au niveau spatial (par exemple, le nombre d'arbre, un indice, etc.). Même ainsi, la description d'un ensemble de séries temporelles n'est pas simple. La courbe moyenne renseigne peu sur le comportement du modèle, les écarts-type à chaque pas de temps sont trompeurs (ils ne peuvent pas servir à établir des intervalles de confiance pour les courbes).

Exemples : dynamique comprenant plusieurs attracteurs ponctuels, comportements atypiques, etc.

On rencontre donc de manière très amplifiée les problèmes de description de distribution multi-modale.

Une piste de description est la recherche de comportements "typiques". Ces prototypes de comportement présentent des éléments de stabilité communs à un ensemble de trajectoire. L'espace des résultats de simulation peut ensuite être segmenté en zone de comportement semblable.

Cependant la définition de ces comportements n'est pas évidente. Exemple de la dynamique d'eutrophisation des lacs : pour un lac donné, la dynamique présente deux attracteurs très simples à définir. Une version stochastique rend plus complexe l'identification des deux comportements. En particulier, la classification réalisée à partir d'une ACP ne donne pas directement le résultat espéré. Ce n'est pas une bonne nouvelle, étant donné qu'il s'agit d'un cas d'école très simple.

## 2. Quelques considérations de robustesse

\*\*\*\*\*

Beaucoup de monde utilise l'ACP, mais qui s'est posé même une fois la question de la robustesse des espaces propres? Rassurez-vous, c'est une méthode robuste au sens où : une petite modification des données entraîne une petite modification des valeurs propres et vecteurs propres (continuité) ; l'ajout d'une simulation supplémentaire peu éloignée du centre de gravité aussi (Th. d'inversion locale). La modification de l'horizon temporel rajoute une dimension et ne modifie pas l'ordre des espaces propres (pour des petites évolutions).

On peut se poser les mêmes questions pour l'ACP fonctionnelle...

Quant aux méthodes de classification... il est difficile d'énoncer des propriétés globales. Pour la CAH, et pour la distance euclidienne, on peut proposer des perturbations maximales qui préservent une classification donnée. Pour les autres méthodes ?

## 3. Comparaison de méthodes ACP+clustering

\*\*\*\*\*

Pour des familles de courbes artificielles, et des niveaux de bruit variables, on a comparé les résultats de la classification produite par l'ACP classique ou fonctionnelle, suivie d'une classification CAH ou k-means. Les résultats expérimentaux suggèrent une plus grande efficacité de l'ACP fonctionnelle suivi de k-means, et ce d'autant que le niveau de bruit est élevé. Mais pour des données réelles...