

# Un nouveau Modèle Autorégressif Vectoriel basé sur les Noyaux à Valeur Opérateur et son application à l'inférence de réseaux

Néhémy Lim<sup>1,2</sup>, George Michailidis<sup>3</sup>, Cédric Auliac<sup>1</sup>, et Florence d'Alché-Buc<sup>4,2</sup>

<sup>1</sup>CEA, LIST, Laboratoire Analyse de Données et Intelligence des Systèmes, 91191 Gif-sur-Yvette CEDEX, France

<sup>2</sup>IBISC EA 4526, Université d'Évry-Val d'Essonne

<sup>3</sup>Department of Statistics, University of Michigan

<sup>4</sup>INRIA-Saclay, TAO, LRI umr CNRS 8623, Université Paris Sud

4 juin 2013

## Résumé

La rétro-ingénierie de systèmes dynamiques à partir de données de séries temporelles est un problème canonique difficile en fouille de données. Pour réaliser cette tâche d'apprentissage, un certain nombre d'approches principalement fondées sur des modèles linéaires parcimonieux ou sur la causalité de Granger ont été proposées dans la littérature. Cependant, lorsque les dynamiques sont non-linéaires, il n'existe pas de réponse systématique qui tienne compte de la nature du système sous-jacent. Nous introduisons une nouvelle famille de modèles autorégressifs vectoriels basés sur les noyaux à valeur opérateur pour identifier le système dynamique et extraire le réseau cible. De même que dans le cas linéaire, il est essentiel de pouvoir contrôler la parcimonie du modèle. Ce contrôle peut être réalisé à travers l'apprentissage conjoint de la structure du noyau et des vecteurs de base. Pour résoudre cette tâche d'apprentissage, nous proposons un algorithme d'optimisation alterné basé sur des procédures de gradient proximal qui permet d'apprendre à la fois la structure du noyau et les vecteurs de base. Les résultats obtenus sur les réseaux de régulation génique de taille 100 de la compétition DREAM montrent que notre modèle est plus performant que des méthodes existantes. Une autre application du modèle sur des données du climat a permis d'identifier des interactions pertinentes et interprétables entre des facteurs naturels et ceux liés à l'activité humaine, confirmant ainsi la capacité de notre procédure d'apprentissage à extraire des dépendances entre les variables d'état.

**Mots-clef** : Inférence de réseaux, noyau à valeur opérateur, modèle autorégressif vectoriel.