
Sélection de variables dans les modèles d'équations structurelles : application à la transition écologique.

Mots-clés : Modèles d'équations structurelles, Sélection de variables, Structure latente, Causalité.

Contexte :

Le succès de la transition écologique passe par une modification profonde des actions et pratiques de l'Homme en lien avec son environnement. Ces dernières années ont exacerbé le besoin de ces modifications de pratiques pour assurer la préservation de la biodiversité et des écosystèmes. Ces changements doivent agir sur de nombreux domaines (économie, transports, alimentation ou encore agriculture) qui, individuellement, ont des comportements qui peuvent se modéliser par des systèmes complexes. Ainsi, pour garantir une efficacité à long terme des actions menées, il est essentiel d'étudier ces actions par une **approche holistique** qui prend en compte le système d'étude dans sa globalité.

Pour étudier un système complexe il semble nécessaire d'augmenter le nombre de variables observées pour caractériser au mieux le système. Toutefois, cette augmentation pose un certain nombre de difficultés tant d'un point de vue expérimental, pour des questions de coût notamment, que d'un point de vue statistique où l'on peut être confronté à une forte multicollinéarité et au fléau de la dimension. Ces difficultés sont déjà bien identifiées dans le contexte de la transition agroécologique où l'étude des agrosystèmes doit permettre d'identifier les leviers (c'est à dire les variables) que l'agriculture peut actionner, notamment pour réduire son impact sur l'environnement. L'enjeu principal consiste donc à proposer une approche de modélisation à la fois globale et parcimonieuse.

Problématique :

L'objectif du stage est de proposer une méthode statistique permettant la sélection de variables dans la modélisation de systèmes complexes.

Descriptif du stage :

La modélisation holistique de systèmes complexes et en particulier d'agrosystèmes doit pouvoir tenir compte du caractère multi-échelle et multi-compartiment du système ainsi que des relations causales qui pilotent l'évolution du système. Les modèles d'équations structurelles offrent un cadre mathématique et théorique particulièrement adapté à ces caractéristiques, notamment en s'appuyant sur la définition d'un graphe orienté acyclique (DAG) et sur l'utilisation d'une structure latente permettant une hiérarchisation du système [Bollen, 1989]. Plus précisément les modèles d'équations structurelles, par un principe de systèmes d'équations (linéaires ou non-linéaires), permettent de définir les relations causales directes et indirectes entre les variables caractérisant un système d'étude [Pearl et al., 2016]. Les approches parcimonieuses dans les modèles d'équations structurelles se sont pour l'instant limitées à un ensemble de variables observées choisi et/ou des DAGs fixe dans des approches confirmatoires [Shiple, 2016].

Le stage s'inscrit dans une approche exploratoire où plusieurs DAGs peuvent être confrontés aux observations et pour laquelle il est attendu que plusieurs modèles ne soit pas rejetés par les données. Dans ce contexte, il est nécessaire de pouvoir comparer les modèles entre eux afin de déterminer, de façon parcimonieuse, un ensemble de variables "suffisant" pour caractériser la structure causale globale du système.

Le/la candidat/e sera amené/e à proposer et étudier une procédure de sélection de variables dans les modèles d'équations structurelles à variables latentes. Une attention toute particulière sera portée à la comparaison de cette nouvelle avec des pratiques existantes pour identifier les situations expérimentales favorables. Dans e cadre d'un partenariat avec la chaire AEI (Agriculture Ecologiquement Intensive), la méthodes développée sera également testé sur l'analyse des interactions entre le sol et les plantes dans les agrosystèmes.

Durée et rémunération : Le stage est prévu pour une durée de 4 à 6 mois et sera rémunéré selon la réglementation des gratifications de stage (entre 550 et 600 euros par mois).

Lieu du stage : Le stage se déroulera à l'institut Agro, Agrocampus Ouest à Rennes, au sein de l'équipe de Statistique de l'Institut de Recherche Mathématique de Rennes (UMR-CNRS 6625).

Connaissances requises : Ce stage demande des connaissances générales en statistique et statistique appliquée ainsi qu'une bonne maîtrise du logiciel R. Des connaissances en biologie, écologie et agroécologie ne sont pas nécessaires, cependant un intérêt particulier pour le traitement de données biologiques sera apprécié.

Contact : Mathieu Emily - Institut Agro Agrocampus Ouest/IRMAR

▷ Email : mathieu.emily@agrocampus-ouest.fr

▷ <http://emily.perso.math.cnrs.fr/>

Bibliographie :

▷ K. A. Bollen *Structural Equations with Latent Variables*, Wiley, 1989.

▷ J. Pearl, M. Glymour and N. P. Jewell *Causal Inference in Statistics : A primer*, Wiley, 2016.

▷ B. Shipley *Cause and correlation in biology : A user's guide to path analysis, structural equations and causal inference*, Cambridge university, 2016.