

Proposition de stage niveau M2 : Analyse par simulation de l'interaction climat/rendement

Unité MIA-T, INRA de Toulouse

Problématique du stage

De nombreux travaux récents en agronomie portent sur la modélisation de systèmes de culture (à différentes échelles : paysage, parcelle, plante) tenant compte de conditions climatiques : ensoleillement, pluviométrie, etc., typiquement sous forme de séries temporelles. Ces conditions ayant une influence très forte sur le comportement de la majorité de ces systèmes, intégrer l'aléa climatique aux études d'analyse de risque ou d'optimisation de rendement est une question primordiale.

Cependant, la complexité des interactions plante/climat à l'œuvre ne permet pas d'étudier les systèmes de manière explicite. On a alors recours aux approches dites par simulation, où les modèles sont traités comme des « boîtes noires » et les relations entrées/sorties inférées à l'aide d'outils statistiques à partir d'un échantillon simulé. Dans le cadre de ce stage, on s'intéresse à l'étude de variétés de tournesol à l'aide d'un modèle de simulation dynamique (SUNFLO). Dans ce cadre-ci, il est important de comprendre la manière dont le climat influence le rendement.

Différentes stratégies pourront être explorées pour répondre à cette question au cours du stage : analyse de sensibilité, fouille de données (data mining), approche par méta-modélisation... L'objectif final de ce stage sera de mettre en valeur les motifs climatiques les plus influents pour le rendement, indépendamment de la plante ou en fonction de celle-ci et d'estimer à partir de ceux-ci la distribution du rendement conditionnellement aux caractéristiques de la plante.

D'un point de vue théorique, les entrées climatiques du modèle SUNFLO peuvent être comprises soit comme des séries temporelles multivariées, soit comme des variables aléatoires fonctionnelles (prenant leurs valeurs dans un espace de fonctions, c'est-à-dire, en général, un espace de Hilbert de dimension infinie, voir [3] pour une introduction compréhensible sur le sujet). Les différentes questions posées par le contexte applicatif pourront donc être abordées en utilisant ou en développant des modèles adaptés à ce type de données :

- modèles de régression fonctionnelle [2]. Dans ce contexte-là, il sera particulièrement crucial de s'intéresser à trouver une distance adaptée aux données climatiques, qui tiennent compte des décalages temporels et de certaines autres caractéristiques propres à ce type de données ;
- recherche de valeurs atypiques dans les données fonctionnelles [1].

Une des particularités et des difficultés des données à traiter sont leur nature multi-dimensionnelle, c'est-à-dire, le fait que non pas une mais plusieurs variables fonctionnelles doivent être utilisées pour expliquer rendement et sa distribution.

Travail à effectuer

L'étudiant disposera de données complexes issues d'un simulateur déjà éprouvé (SUNFLO). Les données climatiques utilisées en entrées du simulateur sont des données réelles provenant de cinq stations climatiques en France et peuvent être vues comme une série temporelle multivariée.

L'étudiant aura à expérimenter plusieurs approches pour rechercher les valeurs climatiques influentes et devra tenir compte de la grande dimensionalité et de l'aspect fonctionnel des données. L'objectif final sera de proposer des stratégies d'échantillonnage des climats les plus représentatifs de la variabilité du rendement obtenu (variables les plus explicatives dans un modèle de régression par exemple) ou, au contraire, les plus atypiques, à des fins de propagation d'incertitude, d'analyse de risque ou d'optimisation. L'étudiant aura à faire des propositions, notamment en terme de :

- distance adaptée aux caractéristiques propres des données ;
- prise en compte de l'aspect multi-fonctionnel des données.

Ce sujet de stage pourra déboucher sur une thèse (financement demandé).

Pour effectuer son travail l'étudiant utilisera les ressources de la plateforme RECORD développée à l'INRA, et en particulier, le simulateur SUNFLO. La travail d'analyse, en lui-même, sera effectué à l'aide du logiciel R.

Conditions du stage

durée 4 à 6 mois

localisation Unité MIA-T, INRA de Toulouse (Castanet-Tolosan)

rémunération taux légal

encadrement Victor Picheny, Nathalie Villa-Vialaneix

Contact : victor.picheny@toulouse.inra.fr

Profil recherché

- niveau équivalent Master 2 en mathématiques appliquées ;
- maîtrise d'un langage de programmation scientifique (R, Matlab) ;
- bonnes connaissances théoriques en probabilités et statistiques
- aucune connaissance préalable en agronomie n'est nécessaire mais un goût pour ce domaine appliqué serait un plus ;
- maîtrise de l'anglais scientifique nécessaire.

Ce sujet de stage peut déboucher sur une thèse (financement demandé). Un goût pour des développements plus méthodologiques en mathématiques appliquées est donc souhaité dans cette perspective.

Références

Références

- [1] M. Febrero, P. Galeano, and W. González-Manteiga. Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics*, 2008.
- [2] F. Ferraty and P. Vieu. *NonParametric Functional Data Analysis*. Springer, 2006.
- [3] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. New York, 2nd edition edition, 2005.