

# Sélection de variables dans les données climatiques pour les modèles agro-écologiques

## Proposition de stage niveau M2 (printemps 2018)

Unité MIAT, INRA de Toulouse

Pour postuler envoyer CV et lettre de motivation à [victor.picheny@inra.fr](mailto:victor.picheny@inra.fr); [remi.servien@inra.fr](mailto:remi.servien@inra.fr); [nathalie.villalaneix@inra.fr](mailto:nathalie.villalaneix@inra.fr)

### Problématique du stage

De nombreux travaux récents en agronomie portent sur la modélisation numérique de systèmes agro-écologiques à différentes échelles : paysage, parcelle, plante, etc., et sur l'utilisation de ces modèles pour aider à l'amélioration de ces systèmes vis-à-vis d'aspects variés tels que la productivité, l'impact écologique ou la résilience au changement climatique. Dans le cadre de ce stage, on s'intéressera en particulier au modèle de culture STICS [1], développé à l'INRA depuis 1996, utilisé ici pour évaluer le rendement (en tonnes / hectare) de cultures de blé et de tournesol dans le bassin méditerranéen.

Ces modèles tiennent compte des conditions climatiques, typiquement données sous la forme de séries temporelles (mesures journalières sur la période de culture d'ensoleillement, de pluviométrie, de température...). Ces conditions ont généralement une influence très forte sur le comportement de ces systèmes, et comprendre les mécanismes liés à cet impact constitue une question primordiale en *agriculture numérique*.

Cependant, la complexité des interactions culture/climat 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 sont inférées à partir d'un échantillon simulé à l'aide d'outils statistiques. D'un point de vue théorique, les données climatiques, vues comme entrées d'un **modèle agro-écologique**, peuvent être comprises soit comme des **séries temporelles multivariées**, soit comme des **variables aléatoires fonctionnelles** (prenant leurs valeurs dans un espace de fonctions, voir [4] pour une introduction compréhensible sur le sujet). On se placera donc dans un cadre de **régression fonctionnelle**, où on essaie d'inférer une valeur scalaire (un rendement, sortie de notre modèle) à partir de prédicteurs fonctionnels (les données climatiques).

L'objectif de ce stage est de développer des méthodes afin de déterminer des **intervalles au sein des données climatiques** permettant d'expliquer les variations des sorties des modèles. Cette problématique est en lien avec différents outils classiques de la statistique computationnelle : **analyse de sensibilité**, **fouille de données (data mining)**, **sélection de variables (LASSO)**, etc.

### Travail à effectuer

Ce stage s'inscrit dans la suite d'une série de travaux effectués récemment à l'INRA, en particulier d'un stage centré sur des approches par **forêts aléatoires** [2] et une approche basée sur la **régression inverse par tranche** [3]. L'étudiant aura pour première tâche de s'approprier ces travaux et de tester les méthodes développées sur les données issues du modèle STICS.

Dans un second temps, l'étudiant pourra **proposer et expérimenter de nouvelles approches** pour la recherche d'intervalles dans les données climatiques. Plus précisément, il s'agira de proposer des heuristiques intégrées aux méthodes précédentes permettant de construire des intervalles temporels qui décomposent les données climatiques en segments explicatifs. Pour cela, l'étudiant devra réaliser une étude bibliographique des approches existantes.

Enfin, l'étudiant pourra, selon le temps imparti, étudier le bénéfice de telles approches dans un cadre d'application à la propagation d'incertitude, l'analyse de risque ou l'optimisation de modèles.

L'implémentation des méthodes ainsi que le travail d'analyse seront effectués 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 (environ 550 euros)

**encadrement** Victor Picheny, Rémi Servien, Nathalie Villa-Vialaneix

Contact : [victor.picheny@inra.fr](mailto:victor.picheny@inra.fr); [remi.servien@inra.fr](mailto:remi.servien@inra.fr); [nathalie.villa-vialaneix@inra.fr](mailto:nathalie.villa-vialaneix@inra.fr)

## Profil recherché

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

## Références

- [1] Nadine Brisson, Christian Gary, Eric Justes, R Roche, Bruno Mary, Dominique Ripoche, Daniel Zimmer, Jorge Sierra, Patrick Bertuzzi, and P. Burger. An overview of the crop model stics. *European Journal of agronomy*, 18(3) :309–332, 2003.
- [2] R. Kpekou-Tossou. Analyse par simulations de l'interaction climat-rendement. *Master 2 internship report*, 2015. [http://www.nathalievilla.org/doc/pdf/RKpekouTossou\\_report.pdf](http://www.nathalievilla.org/doc/pdf/RKpekouTossou_report.pdf).
- [3] V. Picheny, R. Servien, and N. Villa-Vialaneix. Interpretable sparse SIR for functional data. *Preprint*, 2016. <https://hal.archives-ouvertes.fr/hal-01325090>.
- [4] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. New York, 2nd edition edition, 2005.