Variable selection in climatic data for agro-ecological models

Internship - Master 2 (spring 2018)

Applied Mathematics and Informatics Unit, INRA Toulouse

To apply, send curriculum and cover letter to victor.picheny@inra.fr ; remi.servien@inra.fr ; nathalie.villa-vialaneix@inra.fr

Overview

Many recent works in agronomy deal with the numerical modeling of agro-ecological systems at different scales: landscape, plot, plant, etc., and the use of such models to help improving those systems with respect to various aspects, such as productivity, ecological impact or resilience to climate change. This internship will focus on the crop model called STICS [1], developed and used at INRA since 1996, which can be used in particular to predict the yield (in tons / hectar) of wheat and sunflower crops in the Mediterranean region.

Agro-ecological models account for climatic conditions, typically given as time series (daily recordings of rainfall, temperature, amount of sunshine..., over the crop season). These conditions usually have a critical influence on the behavior of the models. Hence, understanding this influence and the corresponding underlying mecanisms are of paramount importance in this context.

However, the complexity of the interactions between crop and climate does not allow us to study the systems explicitely. Instead, one can rely on approaches based on simulation, where the models are considered as *black-boxes*, and the inputs/outputs relationships are inferred from a sample generated by running several instances of the model. From a theoretical point of view, climatic data, seen as inputs of an **agro-ecological model**, can be treated either as **multivariate time series**, or as **functional random variables** (i.e., taking values in a functional space, see [4] for a gentle introduction). Hence, this belongs to the framework of **functional regression**, where a scalar value (e.g. the yield, output of the STICS model) is inferred from functional predictors (the climatic data).

The overall objective of this internship is to develop in this framework new methods to determine **critical time intervals within the climatic data** that explain the variations of the model outputs. This objective is closely related to classical tools of computational statistics: **sensitivity analysis, data mining, variable selection (LASSO)**, etc.

Work plan

This internship follows two recent works conducted at INRA, an internship focused on **random forests** approaches [2] and an approach based on **sliced inverse regression** [3]. The first task of the intern will be to understand those approaches and test them on data generated using the STICS model.

Then, the intern will **propose, implement and test new approaches** to search for intervals in climatic data. More precisely, he/she is expected to propose new heuristic algorithms interlinked with the mentioned approaches in order to design time intervals that decompose the climatic data in relevant and interpretable parts. To do so, the intern may realize a bibliographic study of existing approaches.

Finally, if time allows, the intern will study the impact of his new approaches in uncertainty propagation, risk analysis or optimization contexts.

The implementation of the methods, as well as the data analysis will be done in R.

Conditions

duration 4 to 6 months

place MIAT unit, INRA de Toulouse (Castanet-Tolosan, France)

remuneration approx. 550 euros / month

supervision Victor Picheny, Remi Servien, Nathalie Villa-Vialaneix
Contact : victor.picheny@inra.fr ; remi.servien@inra.fr ; nathalie.villa-vialaneix@inra.fr

Sought profile

- Master 2 student in applied mathematics or engineering;
- Good knowledge of either Matlab or R;
- Strong background in statistics or machine learning;
- No prerequisite knowledge in agronomy, however motivation to work on this field would be a plus;
- Good English or French level is mandatory.

References

[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.

[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.