

Contrat post-doctoral – 18 mois

Domaine : statistique/data science

Sujet : *Analyse des déterminants de trajectoires pluriannuelles de dépérissement du vignoble par régression linéaire bayésienne pour données fonctionnelles*

Recrutement : dès que possible

Localisation : UMR MISTEA et ABSys, INRAE Montpellier SupAgro France.

[English description available on page 4.](#)

Les UMR MISTEA et ABSys proposent un post-doc de 18 mois à temps complet, cofinancés par le LabEx Numev et l'Institut de Convergence #DigitAg. La rémunération se fera sur la base de la grille de rémunération en vigueur à l'Université de Montpellier, pour un INM (échelon) 546 ou 564 selon le profil du/de la candidat.e.

Contexte

Depuis quelques années, des baisses de rendement et de longévité des ceps de vigne sont déplorées par les professionnels de la viticulture dans de nombreux bassins viticoles. Un caractère particulier du dépérissement viticole est qu'il se manifeste sur des temps longs, de quelques années à quelques décennies. Peu de données sont disponibles sur des temps aussi longs. Le Bureau National Interprofessionnel du Cognac (BNIC) a réalisé un suivi de 55 parcelles depuis 1977. De l'ordre de 90 variables décrivent les caractéristiques des parcelles, de la vigne, de la conduite technique, les caractéristiques de la récolte. Ces variables peuvent être associées à des indicateurs climatiques tirés de la base de données Safran de Météo-France. Cette base de données unique offre l'opportunité d'analyser les déterminants de trajectoires pluriannuelles de dépérissement du vignoble.

Objectifs

Une première phase d'exploration du jeu de données du BNIC a été effectuée en sélectionnant des années particulières à rendements ou mortalités remarquables (stage de C. Bial, 2020). Cette approche par sélection d'années particulières est d'usage limité et une caractérisation "dynamique" du rendement et de la mortalité de la vigne doit à présent être entreprise (ie, avec des variables explicatives qui évoluent au cours du temps). L'objectif de ce projet de post-doc est d'identifier (1) les facteurs et les interactions de facteurs biotiques, abiotiques et techniques qui contribuent à la baisse de rendement parcellaire et à la mortalité de certains ceps de vigne et (2) la période à laquelle ces facteurs

ont un impact, à court terme à l'échelle du cycle de culture et à long terme depuis la plantation de la parcelle.

Les produits attendus sont :

- Un diagnostic des principaux déterminants du dépérissement viticole (rendement et mortalité) et de la temporalité de leurs interventions, dans le contexte régional du Cognac ;
- L'évaluation de méthodologies innovantes d'analyse de données permettant de traiter le même type de bases de données produites par des organisations professionnelles ou des entreprises dans d'autres régions, en viticulture ou dans d'autres filières agricoles.
- L'extension d'un package R existant (BLISS) avec les modèles statistiques génériques qui seront développés dans le projet.

Programme de travail

Des analyses statistiques exploratoires seront nécessaires, ainsi que des développements méthodologiques en régression linéaire bayésienne pour données fonctionnelles.

Les analyses statistiques à prévoir dans ce projet sont de deux types différents. D'une part, il s'agira de reprendre les analyses exploratoires en utilisant des méthodes classiques ou développées récemment qui ont montré de bonnes aptitudes en apprentissage, comme les forêts aléatoires (Breiman, 2001). D'autre part, il s'agira de faire des adaptations de méthodes existantes, notamment dans le modèle BLISS qui est un modèle de régression linéaire « interprétable » pour données fonctionnelles dans un cadre de statistique bayésienne (Grollemund, 2019). BLISS ne prend par exemple en compte que des variables explicatives fonctionnelles. Des développements méthodologiques seront nécessaires pour son utilisation dans l'étude des données du BNIC. Dans un premier temps, il faudra ajouter dans le modèle des variables explicatives qualitatives (comme un mode de conduite ou le type de sol). Nous pourrions également ajouter des effets aléatoires dans le modèle, pour obtenir un modèle mixte (effet parcelle ou génotype aléatoire par exemple). En présence de plusieurs variables explicatives, une manière de quantifier les contributions respectives de chacune des variables à la variabilité totale de la variable à expliquer serait également bienvenue pour les experts agronomes. Enfin, d'autres extensions de la méthode peuvent encore être envisageables, comme expliquer une variable binaire et non plus scalaire, imaginer un modèle qui ne soit plus linéaire, ou prendre en compte des erreurs du modèle qui soient corrélées (temporellement ou spatialement).

L'exploration des données guidera les questions à affiner autour du modèle.

Voici les tâches à réaliser pendant le post-doc (y compris rédaction d'articles) :

1. Prise en main du sujet, des données, analyses exploratoires.
2. Adaptation méthodologique autour de BLISS.
3. Analyse des déterminants du rendement, la mortalité étant une variable explicative, avec le climat, le sol et les pratiques culturales.
4. Analyse des déterminants de la mortalité, avec un focus sur le climat, l'état sanitaire et les pratiques culturales.

Premiers éléments bibliographiques:

Aubert, I., Mahé, H., Prêtet-Lataste, C., 2017. Le plan dépérissement en 31 fiches d'actions https://www.vignevin.com/wp-content/uploads/2019/03/2_Les_carnets_du_Plan_Deperissement_Avri12018.pdf

Breiman, L. (2001) Random Forests. *Machine Learning* 45, 5–32.

Grollemund P.M. (2017), Régression linéaire bayésienne sur données fonctionnelles, Thèse Biostatistique, ED I2S, Université Montpellier. (MISTEA-IMAG).

Grollemund PM, Abraham C, Baragatti M, Pudlo P. Bayesian Functional Linear Regression with Sparse Step Functions, *Bayesian Analysis*, 2019, 14(1):111:135.

James, G. M., Wang, J., and Zhu, J. (2009). Functional linear regression that's interpretable. *Annals of Statistics*, 37(5A):2083–2108.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer Series in Statistics, Springer.

Encadrement et moyens mis à disposition

Le travail de post-doctorat sera encadré par Meïli Baragatti (UMR MISTEA, Montpellier SupAgro), Nadine Hilgert (UMR MISTEA, INRAE) et Christian Gary (UMR ABSys, INRAE). Il bénéficiera également de l'appui scientifique de Nathalie Smits (UMR ABSys, INRAE) et de Paul-Marie Grollemund (LMBP, université de Clermont-Auvergne) qui a développé le package BLISS durant sa thèse. Le lieu de travail principal sera situé dans les locaux de l'UMR MISTEA ou dans ceux de l'UMR ABSys, sur le campus de la Gaillarde de Montpellier SupAgro.

Compétences attendues et candidatures

Le ou la candidat(e) doit être titulaire d'une thèse en statistique/data science. Le sujet est appliqué, mais nécessite des développements méthodologiques en statistique bayésienne. Une connaissance des modèles de régression fonctionnelle serait un plus. De plus le ou la candidat(e) devra maîtriser les langages R et C++ (pour l'extension du package BLISS).

Les candidatures doivent être adressées jusqu'au 31 mai 2021 par mail avec CV détaillé, lettre de motivation et une ou deux lettres de recommandation à Meïli Baragatti: meili.baragatti@supagro.fr.

Pour toute information complémentaire, contacter Meïli Baragatti ou Christian Gary (christian.gary@inrae.fr).

Post-Doctoral position in Statistics / Data Science

Subject: *Long-term Analysis of Dieback Drivers*. Analysis of the determinants of multi-year vineyard dieback trajectories by Bayesian linear regression for functional data.

Location: Montpellier, France

Duration: 18 months

Context and aims:

In recent years, the wine industry has been complaining about a drop in yields and vine longevity. Vineyard dieback is a syndrome that results in a drop in productivity over several years and the premature death of certain vines. A particular feature of vine dieback is that it occurs over long periods, from a few years to a few decades. Few data are available over such long periods.

The Bureau National Interprofessionnel du Cognac (BNIC) has monitored 55 plots since 1977. This unique database offers the opportunity to analyse the determinants of multi-year vine dieback trajectories, notably by undertaking a "dynamic" characterisation of vine yield and mortality (i.e. with explanatory variables that evolve over time).

The objective of this post-doc project is to identify (1) the factors and interactions of biotic, abiotic and technical factors that contribute to the decline in plot yield and to the mortality of individual grapevines and (2) the time period over which these factors have an impact, both in the short term at the scale of the crop cycle and in the long term since the plot was planted. Exploratory statistical analyses will be performed, as well as methodological developments in Bayesian linear regression for functional data.

The statistical analyses to be conducted in this project will be of two different types: on the one hand, classical or recently developed exploratory analyses (such as random forests, Breiman, 2001) will be used to get a good handle on the data, On the other hand, adaptations of existing methods will be developed, notably around the BLISS model which is an "interpretable" linear regression model for functional data in a Bayesian framework (Grollemund, 2019). Indeed, BLISS only considers functional explanatory variables. Analysing the BNIC dataset requires methodological adaptations. For example, it will be necessary to add qualitative explanatory variables or random effects to the model. Data exploration will guide the questions to be refined around the model.

Beyond the analysis of this case study, the post-doc will produce a generic protocol for the analysis of temporal data on agricultural systems that will fill a gap in statistical tools for the analysis of agroecological transition trajectories.

Young Researcher skills required

- The applicants should have a Phd in Statistics or Data Science, and strong experience with programming (R and if possible C++), good communication skills and interest in working in a cross-disciplinary team.
- A successful post doc candidate might have experience in the development and application of Bayesian statistics. Knowledge in linear regression with functional data should be a plus.

Supervision and Contact

Meili Baragatti (meili.baragatti@supagro.fr) at [UMR MISTEA](#) and Christian Gary (christian.gary@inrae.fr) at [UMR ABSys](#), with the collaboration of Nadine Hilgert (MISTEA), Nathalie SMITS (ABSys) and Paul-Marie Grollemund (LMBP, universit  de Clermont-Auvergne).

Requested documents for application

A CV, a cover letter, and one or two recommendation letters should be sent to Meili Baragatti (before May 31, 2021).