Modeling the Dynamics of Plants Physiology with Statistical and Deep Learning Approaches

Context

Statistical models for the dynamics of plant physiology aim to describe and analyze the process underlying plant development. In its simplest formulation, the model is a non-linear mixed-effects model in which a limited number of physiological parameters control the shape and rate of the dynamic traits. These physiological parameters can in turn be described as random variables whose distributions depend on a set (possibly high-dimensional) of explanatory cofactors. The inference of such models is rather complex, as it often requires the use of sampling techniques such as SAEM or MCEM [1,2].

Subject

The aim of this post-doc is twofold:

1. First, to develop an alternative inference strategy based on recent tools and techniques from deep learning to fit state-of-the-art non-linear mixed-effect models for physiological dynamics. To this end, the inference will be presented as an optimization problem where the target is a neural network, and where the loss function will be chosen according to the nature of the physilogical feature (continuous or discrete).

2. Second, more refined statistical models involving differential equations will be considered. These differential equations will be incorporated into the neural network model using physics-informed neural networks (PINN, [3]).

All developments will be implemented using libraries for high-performance numerical computing and optimization (PyTorch, Jax).

The models will be applied to the data collected in the G2WAS ANR project. The data consist of 250 grape varieties that have been dynamically phenotyped during 3 weeks for vegetative biomass production by imaging at the PhenoArch platform. Each variety underwent 3 different hydric scenarios (from well watered to severe drought stress). A set of $\approx$ 60K genetic markers will be used as explanatory variables in the statistical model to explain and predict the dynamics of plant physiology.

Required skills

- Phd in one of these domains: computer science, statistics or machine learning,
- Proficiency in Python programming, experience in working with large datasets,
- Experience in working with reproducible research methods (GitLab, versioning, testing of code),
- Fluency in written English.

Previous experience in working with biologists and/or basic knowledge in genetics is a plus.

Working environment

The Post-Doc position is funded by the ANR G2WAS Project. You will work in the SOLsTIS team of the MIA Paris-Saclay unit, located at AgroParisTech (Palaiseau). The Post-Doctoral researcher will be supervised by Tristan Mary-Huard, Laure Sansonnet and Julien Chiquet, in close collaboration with Vincent Segura and Timothée Flutre for the plant physiology and genetics aspects.
This research project corresponds to a 1 year position, the starting can be as early as September 2023.

Contact

Interested candidates should apply by sending a CV and a motivation letter at
tristan.mary-huard@agroparistech.fr
laure.sansonnet@agroparistech.fr
julien.chiquet@inrae.fr

References

[1] Kuhn & Lavielle (2005). Maximum likelihood estimation in nonlinear mixed effects model, Comput. Stat. and Data Analysis.

[2] Liu & Wu (2007). Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. Biometrics.

[3] Raissi, Perdikaris & Karniadakis (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. of Comput. Phys.