

Master 2nd year - research internship

Model comparison in multivariate hidden semi-Markov models with covariates

Application to modelling flowering sequences in apple trees.

Supervisors: Sandra Plancade⁽¹⁾ and Jean-Baptiste Durand⁽²⁾

⁽¹⁾ sandra.plancade@inrae.fr, MIAT, INRAE, Auzeville-Tolosane

⁽²⁾ Jean-Baptiste.Durand@univ-grenoble-alpes.fr, Inria and Laboratoire Jean Kuntzmann, St Martin d'Hères

Location of internship: INRAE, MIAT, Auzeville-Tolosane (near Toulouse)

Allowance: about 550 euros per month

Duration: 4-6 months (starting February - early April 2022)

Context

The development of automatic tools for plants phenotyping produces a large amount of data which can be analysed to guide agriculture practice, provided that reliable models are available. Notably, fruit tree trunks produce lateral branches characterized by various features (number of flowers...). The classical modelling assumes that tree growth occurs through successive unobservable (or latent) phases, whose duration varies from one tree to the other, and that impact the measured variables on branches [1]. This assumption enables to infer the latent phases indirectly from the observations. The corresponding class of mathematical model is the hidden semi-markov models (HSMMs) with multivariate observations (Multivariate HSMMs or MHSMMs; see Yu, 2010 [2]). Currently used MHSMMs assume development phases to have an effect on all observed variables simultaneously.

We are specifically interested in comparing the tree development of various classes of trees (cultivar, treatment, etc), introduced in the model through covariate effects on the parameters, with an identical covariate effects on the observed features through a given latent phase. But in some context, this assumption can lead to biased conclusions and variants MHMMs should be considered to assess which assumptions are appropriate for a given data set.

Tasks

The objective of this internship is to propose a procedure to analyse the validity of different variants of MHSMMs in presence of covariates, based on a data set collected on juvenile apple trees.

First of all the candidate will determine the models to be compared, through a reflexion jointly on the mathematical and biological assumptions. Then comparison criteria will be chosen based on the existing references. Therefore, models will be compared on experimental data, notably using a dedicated software (vplant). Complementary analyses could be implemented on simulated data, in which assumptions are controlled, to validate the methodology.

In a first time, the mathematical questions could be studied via Hidden Markov Models (HMM), which represent a particular case of HSMM, and for which more tools are available.

Prerequisites

The candidate should possess a solid mathematical background, with a specialisation in statistics, as well as programming skills. Knowledge on stochastic processes would be appreciated. The candidate should also be interested in application to agronomy / plant science (but no experience is required on this aspect).

Remark

This work may be continued as a PhD thesis, depending on the team's success in obtaining fellowships.

References

[1] Mészáros, M., Guédon, Y., Krška, B. and Costes, E., *Modelling the bearing and branching behaviors of 1-year-old shoots in apricot genotypes* (2020), PLoS ONE, **17**(7), e0235347.

[2] Yu, S.-Z., *Hidden semi-Markov models* (2010), Artificial intelligence, **174**(2), 215–243.