Proposal of master internship

Title: Study of deer behavior based on accelerometry data: generation of realistic data to evaluate the performances of a hidden semi-Markov model.

keywords: Statistics, Hidden Semi-Markov Models, Simulations, Accelerometry Data, Behavioral Ecology.

Context: High-frequency sensor data, particularly accelerometers, allow for the study of animal behavior over long durations and in conditions that are not directly observable, helping to better understand their interactions and adaptation strategies. The CEFS unit (behavior and ecology of wild fauna) has developed a large study on deer behaviour, including accelerometry data from long-term monitoring of a wild roe deer population, as well as supervised data that combines both accelerometry and video data, enabling much finer exploration.

In this context, classical analysis typically focuses on the proportion of time spent in each behavior (feeding, running, etc.), ignoring the temporal aspect of the data. Alternatively, we have developed an approach based on Hidden Semi-Markov Models (HSMM), a class of stochastic processes that provides a natural framework for studying these behavioral dynamics and allow for the exploration of new applied questions.

Internship objective: The goal of this internship is (i) to develop a framework for generating realistic data to study the performance of our approach, (ii) to use these generated data to evaluate the performance of our approach, particularly its ability to distinguish behavioral differences between two conditions (e.g., the level of vigilance in animals may differ between a covered and an open environment). However, the proposed model includes classical but unrealistic assumptions, particularly the independence of accelerometry measurements conditioned on the underlying behavior chain. Therefore, a simple simulation based on this model would be insufficient.

Two approaches can be considered for generating datasets: (i) Data can be simulated under a more complex HSMM, which would include autocorrelation in the observations; this work would combine an exploration of existing models in the literature with modeling to propose meaningful enrichment from a behavioral perspective. (ii) Alternatively, a data-driven approach could be developed in the spirit of the pseudo-observations developed in survival analysis. By using the large amount of available accelerometry time series data and restoring the associated behavior sequences, we can generate a large set of supervised sequences from which samples with the desired characteristics can be drawn. The choice between one or both approaches will take into account their advantages and disadvantages, as well as the intern's interests and skills.

These generated data will first allow us to assess the robustness of our model when the parameters governing the dynamics to be restored deviate from those learned on the training dataset. Then, the ability to detect differences in dynamics between two conditions will be analyzed by identifying behaviorally relevant differences.

Required Skills. The intern should have a strong background in mathematics/statistics to understand complex stochastic processes, as well as proficiency in programming. The main supervisor primarily uses R, so this language will be preferred, although the intern may also work in Python.

Supervisors.

Sandra Plancade, INRAE Toulouse, unité MIAT
Nicolas Morellet, INRAE Toulouse, unité CEFS
Nathan Ranc, INRAE Toulouse, unité CEFS
Contact : sandra.plancade@inrae.fr

Internship Conditions. The intern will be hosted in the MIAT unit (INRAE Toulouse) for a period of 4 to 6 months starting from April 1, 2025. The internship will be remunerated at the current hourly rate (4.35 € per hour in 2024).