POLITÉCNICA

E.T.S. DE INGENIERÍA AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

Título de las prácticas (Title of the internship):

Species traits quantitative estimation through phylogenetic information

Descripción de las funciones del alumno (Description of the student´s tasks)

In this internship, the student will work in a research group of the Department of Applied Mathematics, interested in theoretical approaches in the framework of computational biology. Along the internship, the student will apply the methodology and techniques learned in the MSc on Computational Biology relative to eco-evolutionary systems modeling and the corresponding computational analyses. Moreover, the student will learn new tools and techniques specialized to these kinds of complex systems. The student will be integrated in a collaborative and multidisciplinary environment, acquiring experience on scientific collaboration with researchers from both the theoretical and the experimental sides. As a result, the student will have acquired the skills needed to work in a group focused on the computational study of biological systems in a wide sense.

To be precise, the student will analyze data from biodiversity ecosystem functioning experiments with the aim of predicting a robust estimation of the number of traits that better fits biodiversity data in experiments. The approach is detailed below in the training project. The project is based on previous work of the group: see references (1) Serván, Capitán et al. (2018), Nature Ecology and Evolution; (2) Serván, Capitán et al. (2023), bioarxiv: https://doi.org/10.1101/2020.09.04.283507.

Requisitos (Prerequisites): *(indicar titulación y curso) (give Grade and academic year); otros requisitos adicionales (idiomas, informática, otros conocimientos, etc) (other aditional prerequitistes (languages, informatics, other knowledge, etc)*

Enrolled in the MSc on Computational Biology. Programming skills in R or python, eager to become proficient on them at the end of the internship. Background in statistics and machine learning. Interest on pursuing a research career will be a plus.

Proyecto formativo (Training Project)

PRESENTATION

The relation between biodiversity and function in ecosystems has been analyzed through Biodiversity Ecosystem Functioning (BEF) experiments, which are among the most complex ones in ecology and conservation. In those experiments, biodiversity is controlled by growing a number of species, which follow their dynamical trajectories, and, at the end of the experiment, a global productivity index of the ecosystem (typically global biomass) is measured.

Current approaches to model BEF experimental data are based on simple linear regression techniques, which sometimes incorporate aggregated measures of phylogenetic diversity, aimed at relating total biomass with phylogeny. Understanding the relation between ecosystem biodiversity and phylogenetic diversity is crucial to design efficient conservation protocols. This proposal aims at statistically analyzing BEF data to make experimental results be compatible with models of population dynamics.

OBJECTIVES

The objectives of the proposed training project are threefold:

- Statistical analysis of dynamical models based on the phylogenetic tree structure of the community to represent species interactions, allowing to differentiate the goodness-of-fit to data due to model features that make it able to reproduce trends in data. The models will make minimal assumptions to estimate trait values as a random evolutionary process over tree structure. To analyze data, machine learning techniques, statistical analysis, and data visualization will be used.
- Analyze the effect of explicitly incorporating the phylogenetic tree as a determinant of species interactions, and what are the corresponding implications this has in species abundances and biomass. By doing so, any level of similarity between species that allow for stable species coexistence will be validated. The models will be analyzed theoretically and computationally as function of the number of traits used to define a community.
- The optimal number of traits compatible with data and with a fixed tree structure will be estimated, through maximum likelihood estimation of model biodiversity predictions to data.

METHODOLOGY

Different statistical and computational techniques will be applied to analyze BEF datasets, which will be mainly provided by the group of Prof. Stefano Allesina (Department of Ecology and Evolution, University of Chicago), which is an on-going collaborator of our group at UPM. Machine learning techniques will pervade this statistical analysis, either as regression models and/or classification algorithms, or through the dimensional reduction of the multivariate dataset via principal component analyses. These machine learning techniques are now becoming useful due to large dataset generation on ecological communities in two aspects: experimental (reporting biomass and abundances) and computational (through genomic and phylogenetic analyses).

The research group which will mentor the student has a long experience in dynamical, mathematical models aimed at predicting species coexistence and feasibility in ecological communities. The main theoretical results in which this proposal is based are already obtained, so the student will be involved in model simulation, dynamic model fitting to data via maximum likelihood parameter estimation, and the validation of the corresponding results and goodness of fit.

PROSPECTIVE

The fundamental goal of the external practices is to guide the student in applying his previously acquired knowledge to real tasks in an environment that realistically represents the working conditions the students will encounter in their future roles. In this way, the student will get familiar with working schedule, responsibility, attitude, organization, etc., as well as with the adequate methodology in professional teams. This will be achieved by collaborating daily in the research group directed by the academic tutor.

At the end of the training, the work conducted by the student will result in an end-of-master dissertation, which summarizes the results obtained from the statistical analysis conducted and the estimations of the number of traits for the communities analyzed. Ideally, those results might

be publishable in a scientific article. It is expected that, after the training period, the student will have a basic initial formation to start over a professional activity in the field of computational biology.

Actividades a desarrollar en la práctica académica (Activities that will be performed in the academic internship):

The skills and competences that the student will acquire during the academic internship will be:

- Literature review to gain insight on the statement of the proposed problem to study, as well as the up-to-date approaches to solve it.
- Preparing and curating different datasets of BFE experiments. Retrieving phylogenetic trees from available sources.
- Analyzing the datasets through machine learning techniques. Statistical modeling of the data, and parameter inference (number of traits that better represent an ecological community) from fitting dynamical models to data.
- Evaluation of theoretical predictions of biomass per species and total biomass according to these models.
- Critical evaluation of the results, and potential errors detection. Goodness-of-fit evaluation of the fitting procedure. Biological and ecological interpretation of the results.
- Organizing the results obtained in a report (end-of-master dissertation), or potentially a manuscript for publication in specialized journals.

Nº de plazas: **(Nr. of places)**	1
¿El alumno tendrá trato habitual con menores? **(Has the student dealings with underage persons?)**	No
Fecha de inicio: **(Starting date)**	8/1/2024

Fecha de fin: **(End date)**	30/6/2024
Horas semanales: **(Weekly hours)**	25 hours
Horario jornada laboral: **(Working hours)**	Flexible, between 10:00 to 17:00
Importe Ayuda/Bolsa de estudio: **(Amount of fellowship / remuneration)**	- **€/mes**
Tutor académico: (Academic tutor (UPM)) Email: juancarlos.nuno@upm.es	Juan Carlos Sanz Nuño
Departamento tutor académico: **(Dept. of academic tutor)**	Department of Applied Mathematics
Tutor empresa: **(External tutor)**	José Ángel Capitán Gómez
Email tutor empresa: **(Email external tutor)**	ja.capitan@upm.es
Departamento tutor empresa: **(Dept. of external tutor)**	Department of Applied Mathematics

POLITÉCNICA

E.T.S. DE INGENIERÍA AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

Ubicación de la estancia de las practicas **(Location of the internship)**	ETS de Ingeniería Agronómica, Alimentaria y de Biosistemas Av. Puerta de Hierro, 2. 28040, Madrid.
ENTIDAD COLABORADORA: **(Collaborating Entity)**	
A cumplimentar por Oficina Prácticas ETSIAAB: **Créditos a reconocer (Nº ECTS)**:	

Enviar por email a: OFICINA DE PRÁCTICAS ACADEMICAS EXTERNAS – ETSIAAB
secretaria.pei.etsiaab@upm.es – Secretarias: Visitación Pérez / Susana Pardo - Tfno: 913363686)