Introduction to theoretical models and link to data (Day 1)

Why mathematical models?

In this introduction, we first discuss if we still need theory at the age of big data. We define what is theory and what is a model in a general way, and more specifically mathematical models as used in ecology. We develop on the necessary trade-offs involved in the construction of a model, in a relation of its specific goals in science: understanding or predicting. We then discuss to which extent we can reach these goal by instead using artificial intelligence informed by big data. Then we come back to the principles of the scientific method and elaborate on how mathematical models participate in this cycle of scientific knowledge production involving data and theory. We finally advocate for the necessity of deeper interactions between data and theory despite some identified current weakness in this link.

What types of theoretical models in ecology?

In this second part of the first day, we first consider that a model is necessarily defined in relation to a given question. The biological system and the associated research question jointly define the scale of interest. Depending on our question and on the scientific goal underlying this question (understand vs predict), we can use a range of different model types which features vary along two axes: from simple to complex and from phenomenological to process-based. We give the main reasons why focusing on process-based simple models and stressed out that the different types of assumptions we make while building models constrain the interpretation of model predictions. In a second time we present the different model formalisms regarding stochasticity (deterministic vs stochastic models), time (static models vs dynamical models, with either discrete or continuous time) and space (non-spatial vs spatially implicit vs explicit models). We discuss the situations in which some formalisms are more appropriate than others. Finally we present the main technical choices to make: using either agent-based modelling or dynamical equations, highlighting their pros and cons. Furthermore, we discuss tractability of dynamical models in relation to their complexity.

How to build a model?

In this third part, we will first try to apply what we've just discuss on a specific system and question in an interactive way, that is to define the variables,the processes to be included in the model, the formalism we can choose. We then will identify together the main assumptions underlying a classical simple model (Rosenzweig & McArthur revisited) from the mathematical formulation of the processes. Finally, we will show the principle of numerical integration to simulate a system of differential equations.

How to analyze a model?

In the fourth part, we see how we classically analyse a mathematical model (ODEs): search for equilibria, from analytical expression to graphical representations; run a local stability analysis of these equilibria to study the behaviour of the model; look for the dependence to initial conditions. In a second time we see the different ways to use a theoretical model to answer different types of research questions (parameter exploration, change the model structure, *in sillico* experiments), and we finally present how to assess the robustness of our conclusions (sensitivity analysis).

Making a link between models and data

On the last part of the day, we discuss how one can make a connection between models (of any type) and data. This is an overview of a broad field in statistics. We first discuss how to make a qualitative comparison of model predictions and data, and then consider the ways to make the comparison quantitative. First, finding the best parameter values (model fitting) with least squares, maximum likelihood, or Bayesian approaches. Second, quantifying the uncertainty in parameter estimates (intervals), and the goodness of fit. Finally,comparing different models and selecting the 'best' (model selection), with the possibility of using alternative models rather than just one (multimodel inference).