

# Spatial data (Day 3)

## Objectives and introduction

Compared to Day 2, we will consider a slightly larger scale, considering communities of several species, distributed over a large spatial extent. The typical spatial dataset would consist in a community matrix (sites\*species), with either presence/absence data, detection/non detection data (if there is imperfect detection), or abundance data (density or number of individuals). Possibly, one may have presence only data, which causes specific issues. Ideally, this community matrix is observed repeatedly at several points in time (time series), which significantly expands the range of methods and models we can mobilize, compared to static (snapshot) datasets. This basic data set is often complemented with other relevant information describing species (traits, phylogenetic relationships...) and/or locations (coordinates, environmental descriptors...). The typical questions we want to address are: where are species distributed? What is the relative importance of environmental attributes versus dispersal in shaping communities? Can we infer interactions (competition, facilitation...) between species?

We briefly present three general frameworks used in this context: -1- The habitat filtering approach; -2- The metacommunity framework; -3- Vellend's community ecology theory framework, and discuss their strengths/weaknesses. We refer to this cool website presenting reflections from Vellend and Leibold on the fundamental publications which introduced frameworks (2) and (3).

We then go over a repertoire of methods and models that can be used to describe and analyze such datasets. We make a distinction between (1) data-driven approaches and (2) process-driven approaches.

## Data-driven approaches

We first consider community-level approaches, i.e. the ones that consider community composition/diversity as a patch metric. Unconstrained methods do not directly use environmental variables: the associations with environmental descriptors is done *a posteriori* (e.g. PCA). Constrained methods simultaneously consider species composition and environmental descriptors (e.g. RDA). A short focus is made on Cottenie's approach of variance partitioning (partialRDA + PCNM) with examples from mollusk communities in the French Antilles. We then turn to species-level methods, starting with permutation approaches, the Cscore, with some connection here to networks and Day 4. We mention the fact that even without interactions, several processes may yield spurious species associations (e.g. Calcagno et al. 2022). After that, we go to Species Distribution Models, and finally Occupancy models, whose hallmark is to add a detection layer (probability to detect a species when it is present). These will be discussed more in the next section (mechanistic approaches).

## Process-driven (mechanistic) approaches

We start by introducing patch-occupancy models, and the two simplest and historically most important models: Levins (1969) metapopulation model and McArthur & Wilson (1967) IB model. We highlight the mathematical connection between Levins' model and the logistic model that was studied on the previous day. We show how an intrinsic covariation of  $r$  and  $K$  shows up in this formulation. We walk-through the computation of equilibrium occupancies in the two models and their natural generalization. We discuss the IB model and its predictions, reflecting on why the curves are non linear. We go through the example of fitting this model to island presence/absence data, inspired from data in Manne et al. (1998).

We discuss possible extensions of patch-occupancy models: -1- explicit space (IFM models from Hanski and followers); -2- competition/colonization trade-offs; -3- spatial networks; -4- trophic webs and others. A more profound extension is to include an explicit description of population dynamics within patches. Models quickly become complicated and need to be simulated. One exception is Hubbel's neutral model, which has a dynamics for species abundances in patches and remains reasonably tractable. An example of simulation framework for metacommunities is Thompson et al. (2020) Ecology Letters. This is the one we will use to simulate data in the afternoon practical.

## Conclusion

We conclude by reviewing the different approaches possible and compare their advantages and drawbacks. We discuss how one approach is not universally better than the others, but rather a set of approaches can yield complementary insights, and different approaches are more appropriate depending on the underlying question and the type/amount of data available.