

What can Joint Species Distribution Models tell us  
about beta diversity?

Linking Hierarchical Modelling of Species Communities (HMSC) with  
nestedness and turnover dissimilarities

Organismal and Evolutionary Biology Research Programme, Faculty of Biological and  
Environmental Sciences, University of Helsinki

Lluís Serra Domínguez

May 2021



<b>Tiedekunta – Fakultet – Faculty</b> Faculty of Biological and Environmental Sciences		<b>Koulutusohjelma – Utbildningsprogram – Degree Programme</b> Ecology and Evolutionary Biology	
<b>Tekijä – Författare – Author</b> Lluís Serra Dominguez			
<b>Työn nimi – Arbetets titel – Title</b> What can Joint Species Distribution Models tell us about beta diversity?			
<b>Oppiaine – Läroämne – Subject</b> Community Ecology			
<b>Työn laji – Arbetets art – Level</b> Master's Thesis	<b>Aika – Datum – Month and year</b> May, 2021	<b>Sivumäärä – Sidoantal – Number of pages</b> 30 (+Appendix)	
<b>Tiivistelmä – Referat – Abstract</b> <p>Beta diversity (total dissimilarity) can be partitioned into two components: dissimilarity attributed to turnover and nestedness-resultant dissimilarity. Turnover refers to the variation in species identities among sites and implies the replacement of some species by others. In contrast, nestedness occurs when species-poor sites have a subset of the biota present in species-richer sites. Although disentangling the relative contribution of these two antithetic components from beta diversity can characterize species assemblages, the dissimilarity indices do not provide information on the processes generating the patterns. Conversely, Hierarchical Modelling of Species Communities (HMSC), which unifies many of the recent advantages of Joint Species Distribution Models, has proved to be the one of the best performing frameworks for unravelling the underlying mechanisms structuring ecological communities.</p> <p>The aim of this research is to explore the relationship between the outputs of the HMSC model and the dissimilarity indices in different communities with a wide range of parameterizations. As the observed patterns measured by the beta-diversity indices result from the underlying processes which HMSC attempts to capture, I hypothesized that both frameworks are at least partially linked to each other.</p> <p>To achieve this aim, I simulated the community data by following the structure of the HMSC model. For simplicity, only one environmental covariate was considered, which was scaled to 0 mean. The intercept of the HMSC model accounted for the baseline occurrence probability of the species, while the slope modeled the species responses to the environmental covariate. The HMSC-intercept and the HMSC-slope, which represent the species multivariate niches, were summarized in terms of center and spread. Simultaneously, the beta diversity indices (total, turnover and nestedness dissimilarity) were calculated from the community data. Finally, the outputs of both frameworks were related in terms of linear modelling and variation partitioning.</p> <p>As hypothesized, the results of this study suggest that outputs of the HMSC model are able to explain most of the variation in the beta-diversity indices, indicating that both frameworks are strongly related. By plotting the species niches (intercept and slope coefficients of the HMSC model) it is possible to determine the main axes of niche variation producing the nestedness and turnover patterns. While nestedness is generated by a shared response of the species to the environmental covariate(s), turnover is produced by variation in the species responses. Finally, the total dissimilarity index is driven by species rarity. In conclusion, the most comprehensive evaluation of the structure of ecological communities and the processes determining the diversity patterns can be achieved by combining the outputs of beta-diversity indices and the HMSC model.</p>			
<b>Avainsanat – Nyckelord – Keywords</b> HMSC, joint species distribution models, beta diversity, nestedness, turnover, dissimilarity			
<b>Ohjaaja tai ohjaajat – Handledare – Supervisor or Supervisors</b> Otso Ovaskainen			
<b>Säilytyspaikka – Förvaringställe – Where deposited</b> Helsinki University Digital Repository Helda			
<b>Muita tietoja – Övriga uppgifter – Additional information</b>			

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Beta diversity, nestedness and turnover . . . . .	3
1.2	Baselga (2010) framework . . . . .	5
1.3	Hierarchical Modelling of Species Communities (HMSC) . . . . .	7
1.4	Aims and hypotheses . . . . .	8
<b>2</b>	<b>Methods</b>	<b>9</b>
2.1	Simulation . . . . .	9
2.2	Statistical analysis . . . . .	12
2.2.1	Species richness and prevalence . . . . .	12
2.2.2	HMSC summaries and beta diversity indices . . . . .	13
2.2.3	Linear models and variance partitioning . . . . .	13
2.3	Replication and confidence intervals . . . . .	14
<b>3</b>	<b>Results</b>	<b>15</b>
3.1	Exploring the data . . . . .	15
3.2	Linear models and variance partitioning . . . . .	17
<b>4</b>	<b>Discussion</b>	<b>20</b>
4.1	How does the HMSC model relate to the beta diversity indices? . . .	20
4.2	Guidelines and considerations for empirical studies . . . . .	22
4.3	Concluding remarks and future directions . . . . .	25
<b>5</b>	<b>Acknowledgements</b>	<b>26</b>
<b>A</b>	<b>Appendix (simulation R code)</b>	<b>31</b>

# Introduction

## 1.1 Beta diversity, nestedness and turnover

Beta diversity describes the variation in species composition among communities. In its simplest form (true beta diversity) it can be calculated as the ratio between gamma diversity (regional or landscape scale) and alpha diversity (local, within-site or within-habitat scale):  $\beta = \gamma/\alpha$ . Although this concept was first introduced by Whittaker (1960), many authors had already developed different indices to gauge the (dis)similarity between species assemblages, e.g.: Jaccard (1912), Simpson (1943), Sørensen (1948). Quantifying the differences between communities is key for understanding how and why biodiversity is distributed in the way it is. Knowledge on beta diversity patterns can measure and reveal the scale of biodiversity loss, give new insights on how diversity is maintained, or aid the design of robust protected areas systems, among other longstanding conservation problems (Socolar et al., 2016).

Beta diversity can be further understood as the interplay between nestedness and turnover (Baselga, 2010, 2012; Legendre, 2014). Nestedness occurs when species-poor sites have a subset of the species present in species-rich sites, reflecting non-random species loss as a consequence of any process that favors the assembly or disassembly of communities from a common species pool (Fig. 1.1A; Ulrich & Almeida-Neto, 2012). Turnover, on the other hand, refers to the variation in species identities among sites, and implies the replacement of some species by others (Fig. 1.1B; Koleff et al., 2003). Turnover may result from species sorting mediated by environmental variation (different species are favoured in different habitat patches based on their niche requirements; Leibold et al., 2004).

Nestedness has been well studied in island systems, where the species occurring in

small and isolated patches tend to be a nested subset of those occurring in mainland or larger (and therefore, richer) areas. This is often a product of deterministic processes of environmental filtering such as, for example, dispersal limitation: only the species with higher mobility and dispersal rates may be able to colonize the more isolated patches (Patterson, 1987; Ulrich et al., 2008). On the other hand, species turnover has always been a central topic in biogeography. For example, species replacement under future scenarios of land use or climate change are among the most widely studied patterns (e.g.: Gibson-Reinemer et al., 2015; Virkkala & Lehikoinen, 2017). In nature any combination of nestedness and turnover is possible (Fig. 1.1C), making these antithetic processes difficult to disentangle from community data.

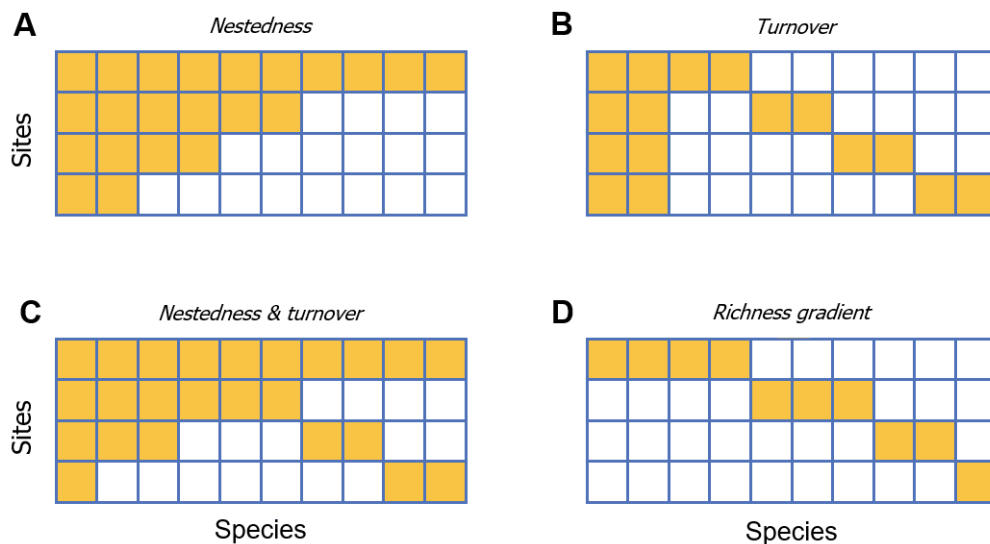


Figure 1.1: Presence (in yellow) absence matrices from 4 hypothetical communities (A-D) with 4 sample sites in each. Community A is completely nested, community B presents a pattern of spatial turnover, community C presents a combination of both patterns, and community D presents a pattern of spatial turnover with differences in richness (but there is no nestedness, since the assemblages in species-poorer sites are not subsets of the assemblages in species-richer sites).

Partitioning beta diversity into nestedness and turnover can therefore provide further insights into the patterns of spatio-temporal variability in communities compared with beta diversity indices alone. For example, Baselga (2010) found that the structural patterns of European longhorn beetle communities are different despite having similar beta-diversity values. In southern Europe beta diversity is mainly

driven by spatial turnover, whereas in Northern Europe it is driven by both spatial turnover and nestedness. Moreover, disentangling the components of beta diversity is also essential for conservation. While nestedness might allow practitioners to prioritize conservation efforts into a small number of richest sites, turnover might require to extend conservation efforts to a larger number of sites, not necessarily the richest ones.

## 1.2 Baselga (2010) framework

Baselga (2010) was the first to unravel the contribution of spatial turnover and nestedness on beta diversity. As for measuring total dissimilarity, he proposed the Sørensen index ( $\beta_{sor}$ , Table 1.1), since it does not discern between differences in composition attributable to richness gradients (nestedness) from differences attributable to species replacement (turnover) (Koleff et al., 2003). The Simpson dissimilarity index ( $\beta_{sim}$ , Table 1.1), on the other hand, describes compositional differences without the influence of richness gradients, so it can be used exclusively to measure turnover. Baselga suggested that the difference between these two indices can therefore be used as a measure of the nestedness component of the beta diversity, that is,  $\beta_{nes} = \beta_{sor} - \beta_{sim}$ . This holds because the Simpson and the Sørensen dissimilarity indices yield the same values when two sites have the same number of species (in the absence of nestedness), since  $b$  and  $c$  must also be equal and  $b/(a+b) = 2b/(2a+2b)$ . It is also evident that if any dissimilarity exists in these scenarios, it is solely due to turnover.

Therefore, after rearrangements, the pairwise nestedness resultant dissimilarity can be defined as (Baselga, 2010):

$$\begin{aligned}
 \beta_{nes} &= \beta_{sor} - \beta_{sim} \\
 &= \frac{b+c}{2a+b+c} - \frac{\min(b,c)}{a+\min(b,c)} \\
 &= \frac{\max(b,c) - \min(b,c)}{2a + \min(b,c) + \max(b,c)} \times \frac{a}{a + \min(b,c)} \\
 &= \frac{|b,c|}{2a+b+c} \times \frac{a}{a + \min(b,c)}
 \end{aligned} \tag{1.1}$$

Table 1.1: Overview of the traditional pairwise dissimilarity indices for presence/absence data mentioned in this thesis. The indices are re-expressed in terms of matching components following Koleff et al. (2003):  $a$  comprises the total number of species present in both sites;  $b$  comprises the total number of species present in the neighbouring site but not in the focal one; and  $c$  comprises the total number of species present in the focal site but not in the neighbouring one.

Measure	Notation	Formula	References
Sørensen	$\beta_{sor}$	$\frac{b+c}{2a+b+c}$	Sørensen (1948); Koleff et al. (2003)
Simpson	$\beta_{sim}$	$\frac{\min(b,c)}{\min(b,c)+a}$	Simpson (1943); Koleff et al. (2003)

Baselga (2010) also proposed a multiple-site generalization of the Sørensen ( $\beta_{SOR}$ ), Simpson ( $\beta_{SIM}$ ) indices and, following the same logic, derived the multiple-site nestedness ( $\beta_{NES}$ ) resultant dissimilarity index (hereafter, capital letters are used to distinguish multiple-site indices from pairwise indices). For the Sørensen index:

$$\beta_{SOR} = \frac{\left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right] + \left[ \sum_{i<j} \max(b_{ij}, b_{ji}) \right]}{2 \left[ \sum_i S_i - S_T \right] + \left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right] + \left[ \sum_{i<j} \max(b_{ij}, b_{ji}) \right]} \quad (1.2)$$

where  $S_T$  is the total number of species in all sites,  $S_i$  is the total number of species in site  $i$ , and  $b_{ij}$  and  $b_{ji}$  are the number of species exclusive to sites  $i$  and  $j$  respectively. By comparing the Sørensen pairwise dissimilarity index ( $\beta_{sor}$ ) to its multiple-site generalization ( $\beta_{SOR}$ ) one can see that  $\left[ \sum_{i<j} \min(b_{ij}, b_{ji}) \right]$  and  $\left[ \sum_{i<j} \max(b_{ij}, b_{ji}) \right]$  correspond to the matching components "b" and "c", and  $\left[ \sum_i S_i - S_T \right]$  corresponds to "a" (Baselga, 2010).

Likewise, the multiple-site Simpson dissimilarity index is given by:

$$\beta_{SIM} = \frac{\left[ \sum_{i < j} \min(b_{ij}, b_{ji}) \right]}{\left[ \sum_i S_i - S_T \right] + \left[ \sum_{i < j} \min(b_{ij}, b_{ji}) \right]} \quad (1.3)$$

Finally, as with the pairwise measures, the multiple-site nestedness resultant dissimilarity can be derived from  $\beta_{SOR}$  by simply subtracting  $\beta_{SIM}$  (Baselga, 2010):

$$\begin{aligned} \beta_{NES} = & \frac{\left[ \sum_{i < j} \max(b_{ij}, b_{ji}) \right] - \left[ \sum_{i < j} \min(b_{ij}, b_{ji}) \right]}{2 \left[ \sum_i S_i - S_T \right] + \left[ \sum_{i < j} \min(b_{ij}, b_{ji}) \right] + \left[ \sum_{i < j} \max(b_{ij}, b_{ji}) \right]} \\ & \times \frac{\sum_i S_i - S_T}{\left[ \sum_i S_i - S_T \right] + \left[ \sum_{i < j} \min(b_{ij}, b_{ji}) \right]} \end{aligned} \quad (1.4)$$

### 1.3 Hierarchical Modelling of Species Communities (HMSC)

Joint Species Distribution Models (JSDMs) are an extension of single Species Distribution Models (SDMs) that account for the multivariate nature of community data by simultaneously analysing all species. That is, they assume shared responses of the species to the environment instead of species-specific responses, allowing to derive both species and community level inference (Ovaskainen et al., 2017). Compared to other approaches (e.g.: distance-based ordinations), JSDMs outputs are the most informative to disentangle the underlying community assembly mechanisms that generate the patterns in the data (Ovaskainen et al., 2019).

Hierarchical Modelling of Species Communities (HMSC) is a recent framework that aims to unify many of the recent advances in JSDMs. The required input data for HMSC-analyses include a matrix of species occurrences (e.g.: presence-absence,



count, or biomass) and a matrix of environmental covariates. Information about species-specific traits, phylogenetic relationships, and the spatio-temporal context of the study design can also be included (Ovaskainen et al., 2017). Apart from modelling the species response to their environment, HMSC is able to relate community patterns to species-specific traits and phylogenies (Abrego et al., 2017), and to quantify co-occurrences among species beyond those generated by species responses to environmental covariates (Ovaskainen et al., 2016). In a comprehensive evaluation of the predictive power of a large number of SDMs and JSDMs methods, HMSC proved to be one of the best performing approaches for modelling communities, even with sparse data (many rare species) (Norberg et al., 2019).

## 1.4 Aims and hypotheses

The aim of this research is to bring new insights into the processes structuring ecological communities and to improve our understanding on how such processes can be inferred by different statistical methods. In particular, I aim to compare the outputs of two alternative approaches that an empirical researcher may apply to their data, namely beta diversity analyses and Joint Species Distribution Modelling (HMSC framework). While the former provides a definite measure of the relative contribution of nestedness and turnover to total dissimilarity, the latter provides correlative inference on the drivers behind species distributions and community composition.

As the observed patterns measured by beta-diversity indices result from the underlying processes which HMSC attempts to capture, I hypothesize that the outputs of these two approaches are at least partially linked to each other. However, as they are developed from very different conceptual starting points, I hypothesize that their outputs are not fully correlated, and thus the most comprehensive evaluation of species communities can be achieved by combining the outputs of the dissimilarity indices and the HMSC model.

# Methods

## 2.1 Simulation

The community data was simulated following the structure of the HMSC model (Ovaskainen et al., 2017). Although working with simulated data may not be as exciting as with real data, in the present study it has some advantages. First, it allows to analyse the behaviour of the diversity metrics under wide range of scenarios with previously chosen parameters. Second, because the true ecological relationships are known, the error arising from the data collection and the estimation of the parameters by fitting the HMSC model is avoided. Third, simulating data under a certain model ensures that one understands each component of that model and makes the link between equations and computer code explicit.

One thousand (1000) communities were stochastically simulated with the same parameterization (Fig. 2.1A). For simplicity, I assumed that the entire metacommunity was embedded within the same environmental context and only one environmental covariate was considered. For each assemblage I generated presence-absence data on  $ns = 100$  species in  $n = 50$  sites, modelled by a probit regression. The R code can be found in the appendix.

First I simulated the regression parameters.  $\beta_{jk}$  denotes the response of species  $j$  to covariate  $k$ . In HMSC the intercept is modelled by the first explanatory variable, that is, the intercept is included within the  $\beta$  parameters. Therefore, in the present study, each species has two regression coefficients: the intercept  $\beta_{j1}$  and a slope  $\beta_{j2}$ .

The intercept is the expected mean value of the response variable when all the predictors are equal to zero. However, considering that the environmental covariate was scaled to zero mean, the intercept  $\beta_{j1}$  models the probability that species  $j$

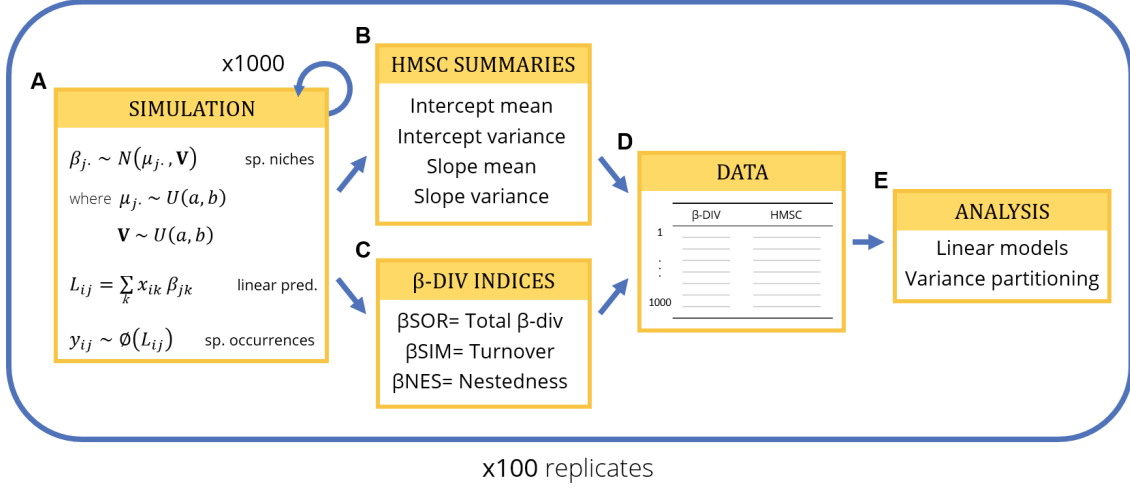


Figure 2.1: Conceptual illustration of the methodology. Panel A illustrates the simulation process for one metacommunity (consisting of a 1000 communities), which follows the structure of the HMSC model; panel B indicates the summaries calculated from the simulated species niches; panel C indicates the diversity indices calculated from the simulated occurrence data (total dissimilarity, dissimilarity attributed to turnover, and nestedness-resultant dissimilarity, respectively); panel D illustrates the resulting data set; and panel E indicates the statistical analyses performed. In order to compute confidence intervals, the entire simulation process and analysis was replicated a 100 times.

occurs at mean environmental conditions. The baseline occurrence probability of a species becomes greater than 0.5 if  $\beta_1 > 0$ , and becomes smaller than 0.5 if  $\beta_1 < 0$ . The slope  $\beta_{j2}$  models how the occurrence of species  $j$  depends on the environmental covariate. If the slope is positive increases with the explanatory variable, if negative, it decreases. A zero slope indicates that the response variable does not depend on the explanatory variable.

The regression coefficients can be further interpreted as the species' niches, since they measure the species response to the environmental conditions. HMSC builds on the assumption of continuous variation in the species niches, so following Ovaskainen and Soininen (2011), I assumed that the regression parameters adhere to a multivariate normal distribution:

$$\beta_{j\cdot} \sim N(\mu_{j\cdot}, \mathbf{V}) \quad (2.1)$$

where  $\beta_{j\cdot}$  is the vector of all regression parameters for species  $j$  and characterises the species entire multivariate niche. Vector  $\mu_{j\cdot}$  is the expectation and  $\mathbf{V}$  is the

variance-covariance matrix of the multivariate normal distribution. If there are  $n_c$  covariates (including the intercept, in this case  $n_c = 2$ ), then vectors  $\beta_j$  and  $\mu_j$  have length  $n_c$  and  $\mathbf{V}$  is a  $n_c \times n_c$  matrix.

For simplicity I assumed a common expected niche  $\mu_j$  for all species within each assemblage. The expectations of the intercept and the slope were drawn from a random uniform distribution. The intercept was set to range both positive and negative values since it influences the degree of filling of the community matrix. This is not trivial because the dissimilarity indices do not treat species rarity and commonness equally (Baselga, 2010). In order to avoid completely filled or empty matrices the mean intercept minimum and maximum were set to -2 and 2, which in the probit model corresponds to a baseline occurrence probability of 0.02 and 0.98, respectively. On the other hand, since the environmental covariate was scaled to zero mean and unit variance, the expectation of the slope was only limited to positive values (0, 5). Negative values would simply produce the same mirror pattern in the response variable. To generate variation in the species niches, the variances of the intercept and the slope (diagonal of the variance-covariance matrix  $\mathbf{V}$ ) were independently drawn from a random uniform distribution with range (0, 2).

The species niches were therefore modelled to be fully independent, since no covariance was assumed among species (for example due to species traits or phylogenetic relationships) nor covariates.

The linear predictor was then generated as:

$$L_{ij} = \sum_k x_{ik} \beta_{jk} \quad (2.2)$$

where the term  $x_{ik}$  denotes the covariate  $k$  measured at site  $i$ , with  $x_{i1} = 1$  modelling the intercept. The standard normal distribution was used to simulate variation in the environmental covariate ( $x_{i2}$ ) with zero mean and unit variance.

Finally I modelled the occurrence (i.e.: presence-absence) of each species  $j$  at each site  $i$  with the probit model:

$$y_{ij} \sim \text{Bernoulli}(\Phi(L_{ij})) \quad (2.3)$$

where  $y_{ij} = 1$  indicates that the species  $j$  is present at site  $i$  and  $y_{ij} = 0$  indicates

absence. The reason why I used the probit model instead of the logistic model is that the former is easier to apply to data than the latter. Namely, the probit model can be written mathematically equivalently as:

$$y_{ij}^* = L_{ij} + \epsilon_{ij} \quad (2.4)$$

where  $\epsilon_{ij} \sim N(0, 1)$ . Then  $y_{ij}$  can be viewed as an indicator for whether this latent variable is positive:

$$y_{ij} = \left\{ \begin{array}{ll} 1 & y_{ij}^* > 0 \\ 0 & \text{otherwise} \end{array} \right\} \quad (2.5)$$

that is,  $y_{ij} = 1$  if  $y_{ij}^* > 0$ , and alternatively, if  $y_{ij}^* \leq 0$  then  $y_{ij} = 0$ .

If  $n$  is the number of sites ( $n = 50$ ) and  $n_s$  the number of species ( $n_s = 100$ ), the end result of the simulation is a  $n \times n_s$  matrix of presence-absences for each community. In total, a 1000 communities were simulated under different parameterizations.

## 2.2 Statistical analysis

### 2.2.1 Species richness and prevalence

Species richness  $S_i$  (number of species present in site  $i$ ) and prevalence  $P_j$  (fraction of occupied sites by species  $j$ ) can be calculated for each assemblage as:

$$S_i = \sum_j y_{ij} \quad (2.6)$$

$$P_j = \frac{\sum_i y_{ij}}{n} \quad (2.7)$$

By visually exploring the variation in species richness and species prevalences one can already get an initial insight on how the environmental niches are structured among the species.

### 2.2.2 HMSC summaries and beta diversity indices

The species niches of each simulated community (i.e the intercepts and the slopes) were summarized in terms of center and spread, that is, mean and variance (Fig. 2.1B). Because positive and negative slopes generate the same mirror pattern, in the analysis I used the absolute value of the mean slope.

The multiple-site nestedness-resultant  $\beta_{NES}$ , turnover  $\beta_{SIM}$ , and total dissimilarity  $\beta_{SOR}$  indices were calculated for each assemblage using Baselga's framework (Baselga, 2010) (Fig. 2.1C). The analysis was done with the R-package "betapart" (Baselga & Orme, 2012), which implements the formulas already discussed in the introduction.

The final data set consists of a matrix with the beta-diversity indices and HMSC summaries for each of the 1000 communities: HMSC-intercept mean, HMSC-intercept variance, HMSC-slope mean, HMSC-slope variance,  $\beta_{SOR}$ ,  $\beta_{NES}$ , and  $\beta_{SIM}$  (Fig. 2.1D). Note that the prefix "HMSC-" will hereafter be used to avoid confusing the HMSC intercept and slope summaries with the intercepts and slopes obtained from the subsequent linear regression analysis.

### 2.2.3 Linear models and variance partitioning

Two related approaches were used to examine the links between beta diversity indices and the HMSC summaries (Fig. 2.1E): First, a linear model was fitted to each beta diversity metric with the HMSC-intercept mean, HMSC-intercept variance, HMSC-slope mean, and the HMSC-slope variance as explanatory variables. In order to compare the effect sizes of the means and variances, the explanatory variables were normalized to a common scale (zero mean and unit variance). The Sørensen beta diversity index had a left-skewed distribution, so I used the reciprocal transformation of  $1 - \beta_{SOR}$  to improve normality. After this, all three linear models met the assumptions of normality, homoscedasticity and independence of residuals. Furthermore, the adjusted  $R^2$  was reported as a measure of goodness-of-fit.

Second, the total variation explained by the model ( $\text{adj-}R^2$ ) was partitioned into HMSC-intercept and HMSC-slope components. That is, each HMSC parameter's mean and variance were grouped together to explain their relative contribution to the beta diversity indices. The variation partitioning analysis is based on three multiple regressions with (i) both sets as predictors (HMSC-intercept and HMSC-slope), (ii) one set as predictor (HMSC-intercept), (iii) the remaining set as predictor (HMSC-slope). The following fractions can then be obtained by simple subtractions of the regressions'  $\text{adj-}R^2$ : [a], the variation uniquely explained by the HMSC-intercept; [c], the variation uniquely explained by the HMSC-slope; [b], the variation jointly explained by the HMSC-intercept and the HMSC-slope; and [d], the residual variation (Legendre, 2007). The variation partitioning was conducted using the R-package "vegan" (Oksanen et al., 2020).

## 2.3 Replication and confidence intervals

In simulation studies, the sample size (in this case number of communities) can be arbitrarily chosen with the only burden of computational resources. Although large sample sizes yield more precise estimates, the p-values should be interpreted with caution. Any effect size, however small it may be, can become significant by just increasing the sample size (unless there is no effect whatsoever; Sullivan & Feinn, 2012). Therefore, it is important to distinguish statistical significance from biological relevance. To avoid falling into interpretation pitfalls, the regression coefficients (with the 95% confidence intervals) will be reported instead.

The 95 % confidence intervals were calculated from a 100 replicates of the entire simulation process and statistical analyses. The 2.5% and 97.5% quantiles of the mean estimates were used to establish the end points of the confidence intervals.

# Results

## 3.1 Exploring the data

Plotting the variation in species richness  $S_i$  and prevalence  $P_j$  indices can help to visualize some key structural differences between highly nested communities and communities with high turnover. Furthermore, by also plotting the variation in species niches ( $\beta$  matrix, i.e.: the generated intercepts and slopes) one can get an initial idea on what is driving nestedness and turnover. For this reason, two of the simulated communities, hereafter called A and B, were used to represent these two contrasting patterns. Community A (Fig. 3.1A) has high turnover  $\beta_{SIM} = 0.936$  and community B (Fig. 3.1B) is highly nested  $\beta_{NES} = 0.722$ . However, both have similar beta diversity values:  $\beta_{SOR} = 0.951$  and  $\beta_{SOR} = 0.927$ , respectively.

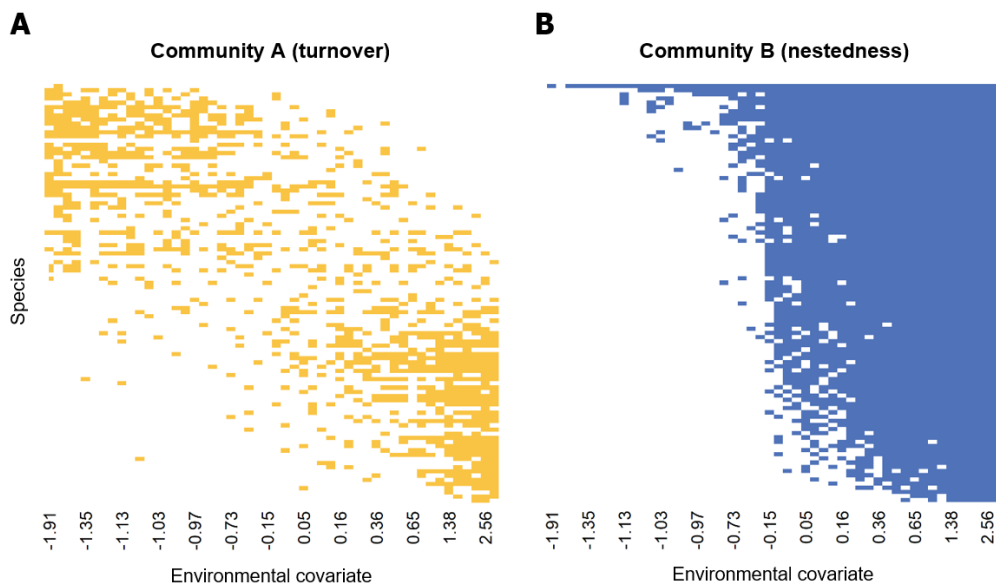


Figure 3.1: Maximally packed presence (coloured) absence matrices from two simulated assemblages (A, in yellow; and B, in blue).



In community A, nearly all generated intercepts are negative ( $\mu_{int} = -1.141$ ,  $\sigma_{int} = 0.245$ ; Fig. 3.2A), making the occurrence probability of these species less than 0.5. As a result, most of the species are only present in a small number of sampling units (Fig. 3.4A). The slopes, on the other hand, take both positive and negative values ( $\mu_{slope} = 0.261$ ,  $\sigma_{slope} = 1.342$ ; 3.2a), indicating that the occurrence probability of some species decreases with increasing environmental covariate while increases for others, thus generating turnover. Furthermore, Fig. 3.3A suggests that sites with either low or high covariate values have the highest species richness whereas at intermediate environmental conditions species richness is minimized.

In contrast, in community B, the simulated slopes ( $\mu_{slope} = 4.036$ ,  $\sigma_{slope} = 0.613$ ; Fig. 3.2B) are positive, indicating that the occurrence probability of these species increases with an increasing value of the environmental covariate, and so does species richness (Fig. 3.3B). This produces a nested pattern, where species-poor sites have a subset of the species present in species-rich sites. Finally, the majority of the intercepts ( $\mu_{int} = 0.916$ ,  $\sigma_{int} = 2.238$ ; 3.2B) are positive, meaning that most species are rather common (Fig. 3.4B).

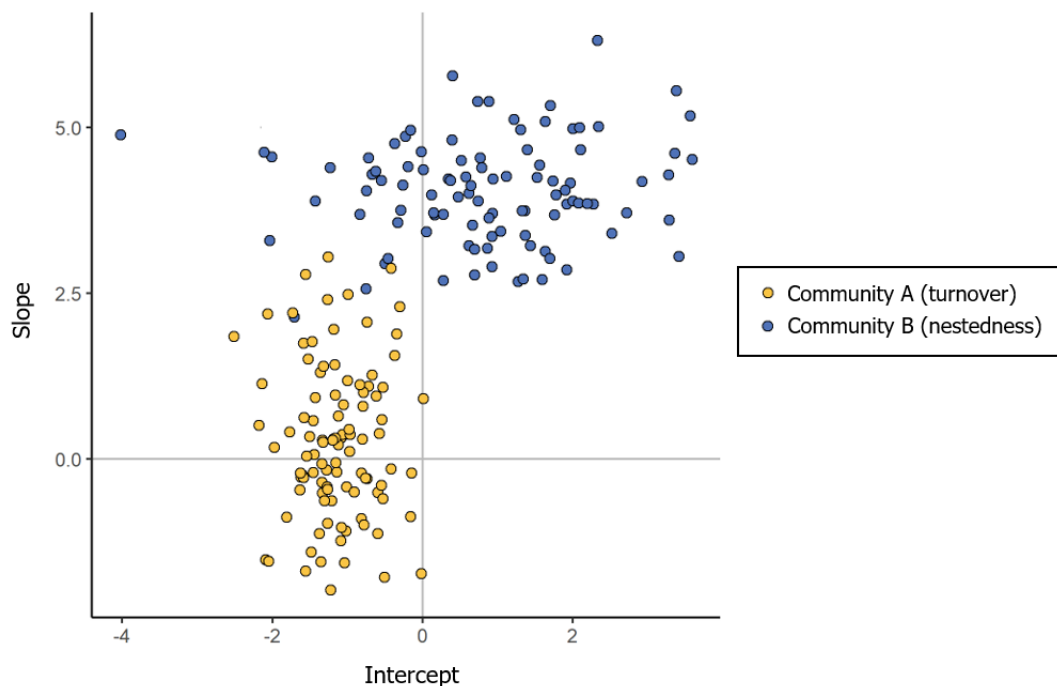


Figure 3.2: Intercept-slope plots, illustration of the variation in species niches. Each dot corresponds to one species.

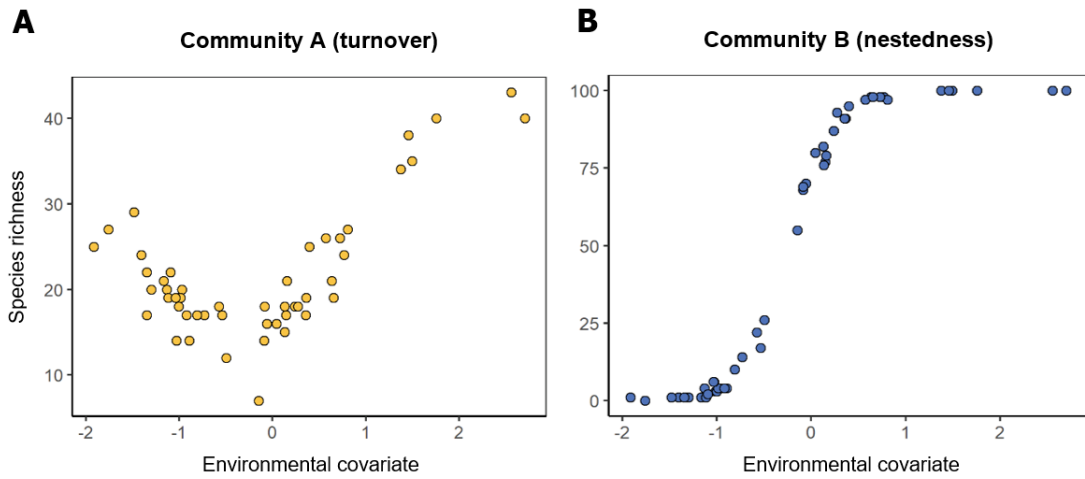


Figure 3.3: Species richness as a function of the environmental covariate.

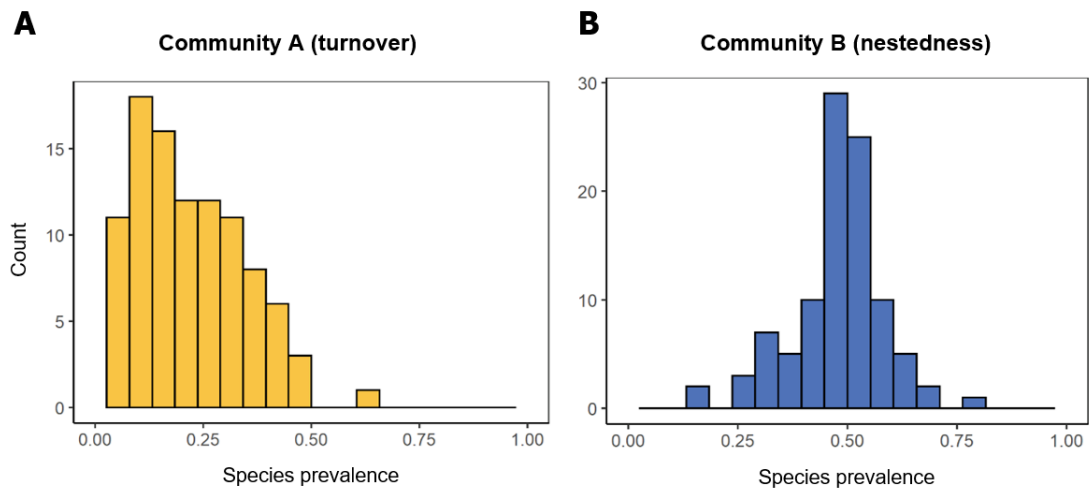


Figure 3.4: Histograms of species prevalences for the simulated Communities A and B. The y-axis (Count) corresponds to the number of species, and the x-axis shows the proportion of sites in which each species is present.

## 3.2 Linear models and variance partitioning

The HMSC summaries explained 89.48% (95% CI, 85.86 to 92.51) of the variation in the total dissimilarity index, 94.18% (95 CI, 90.04 to 96.42) of the variation in the dissimilarity attributed to turnover index, and 96.08% (95 CI, 93.95 to 97.58) of the variation in the nestedness-resultant dissimilarity index. These results suggest that both frameworks (HMSC and beta-diversity measures) are strongly related.

The variation partitioning further revealed that the HMSC-intercept component is the most important factor for determining the total dissimilarity index (86% of variance explained; Table 3.1) while the HMSC-slope component is the most important determinant for the turnover and nestedness indices (81% and 92% of variance explained, respectively; Table 3.1). Because the simulated HMSC intercepts and slopes are orthogonal (independent), there is no collinearity and the shared variation component in all three partitions is close to 0 ([b], Table 3.1).

Table 3.1: Fractions of variation obtained by variance partitioning of the response variables against the HMSC-intercept (mean and variance) and the HMSC-slope (mean and variance). Fraction [a] refers to the variation uniquely explained by the HMSC-intercept, [c] refers to the variation uniquely explained by the HMSC-slope, [b] refers to the variation jointly explained by the HMSC-intercept and the HMSC-slope, [d] refers to the residual variation. The 95% Confidence intervals are reported in brackets.  $\beta_{SOR}$  was transformed to  $1/(1 - \beta_{SOR})$  to reach normality.

Resp.	[a]	[b]	[c]	[d]
$\beta_{SOR}^*$	86.20% [79.19, 90.90]	0.02% [-2.24, 2.02]	3.31% [0.03, 12.46]	10.52% [7.48, 14.13]
$\beta_{SIM}$	13.26% [3.08, 28.30]	-0.27% [-4.81, 3.39]	81.19% [68.60, 89.51]	5.82% [3.58, 9.97]
$\beta_{NES}$	4.54% [0.16, 12.76]	-0.16% [-3.22, 2.34]	91.70% [84.03, 96.55]	3.92% [2.42, 6.05]

Additionally, the regression coefficients of the linear models give important insights into the direction and strength of the relationship between the HMSC summaries and the three beta-diversity indices. (i) The total dissimilarity index ( $\beta_{SOR}$ ) is negatively affected by the HMSC-intercept mean and variance, while the effects of the HMSC-slope are not significant (a coefficient estimate is considered to be significant if the 95% confidence interval does not include the null value 0; Fig. 3.5A). (ii) The dissimilarity attributed to turnover index ( $\beta_{SIM}$ ) is negatively affected by the HMSC-intercept mean, HMSC-intercept variance, and HMSC-slope mean, but positively affected by the HMSC-slope variance. However, the effect of the HMSC-intercept variance appears to be small (Fig. 3.5B). (iii) The HMSC-intercept mean, HMSC-intercept variance, and HMSC-slope mean have a positive effect on the nestedness resultant dissimilarity index ( $\beta_{NES}$ ), while the HMSC-slope variance has a

negative effect. Again, the effect of the HMSC-intercept variance appears to be mild (Fig. 3.5C).

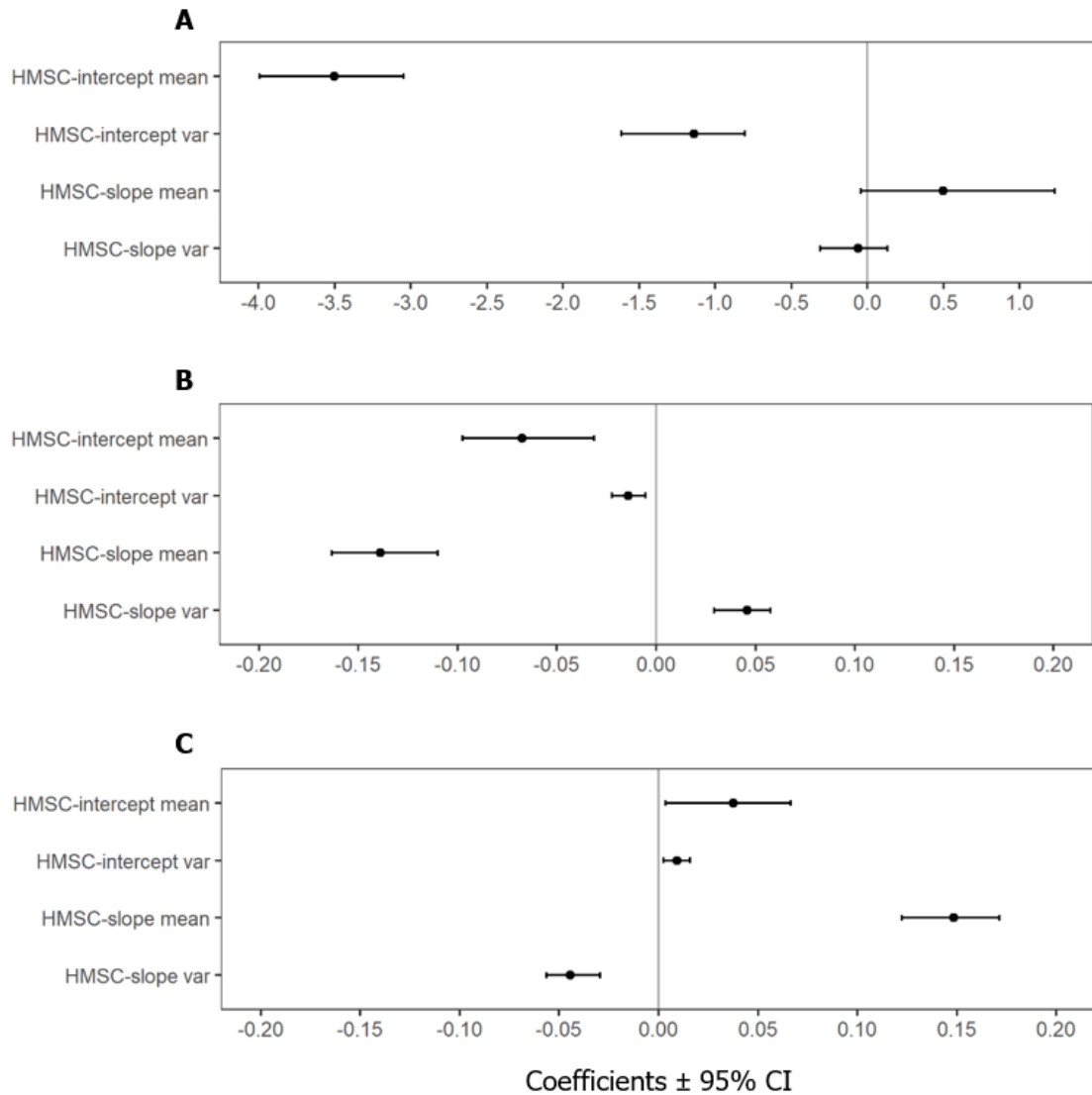


Figure 3.5: Coefficient estimates and 95% confidence intervals of the linear models for the total dissimilarity index (A), turnover (B), and nestedness (C).  $\beta_{SOR}$  was transformed to  $1/(1 - \beta_{SOR})$  to reach normality.

It is worth noting that the coefficient estimates of the linear models for the turnover and nestedness resultant dissimilarities show the opposite pattern and, in fact, both indices are strongly correlated (Pearson's  $r$  -0.980 [-0.988, -0.961]). This is because  $\beta_{NES}$  is calculated by subtracting  $\beta_{SIM}$  from  $\beta_{SOR}$  (Eq. 1.1 and 1.4) and, therefore, they are not independent (Baselga, 2010).

# Discussion

## 4.1 How does the HMSC model relate to the beta diversity indices?

The results of this master's thesis support the hypothesis that Joint Species Distribution Models, in particular Hierarchical Modelling of Species Communities, are strongly related to the beta-diversity partitioning framework from Baselga (2010). The HMSC outputs, which describe the species niches, can therefore identify the main axes of environmental variation generating the nested and turnover patterns.

The Sørensen dissimilarity index (total dissimilarity) is mainly linked to the HMSC-intercept, increasing in communities with low species prevalences. This is because  $\beta_{SOR}$  yields the proportion of unique species per site, so fuller matrices have more chances of having sites with shared species (i.e.: similar composition). The coefficient estimates of the HMSC-slope (mean and variance) are not significant and explain very little variation, suggesting that  $\beta_{SOR}$  is independent of the species responses to the environmental covariate.

Turnover refers to the variation in species identities among sites and implies species replacement (Baselga, 2010; Koleff et al., 2003). The results of the variance partitioning analysis indicated that  $\beta_{SIM}$  is primarily linked to the HMSC-slope. The linear model further suggested that high turnover is achieved in communities with a small HMSC-slope mean with a comparatively large variance. This produces a positive response for some species and a negative response for others, therefore generating species replacement along the environmental gradient. Although the variation explained by HMSC-intercept is small, the coefficient estimates are negative,

indicating that  $\beta_{SIM}$  increases with species rarity. This is to be expected since the differences in species composition between sites are mostly driven by the presence and absence of rare species (common species are shared in most sites).

Nestedness describes a pattern where species-poor sites have a subset of the biota present in species-richer sites, resulting in ordered species gain (or loss) along the environmental gradient. (Baselga, 2010; Ulrich & Almeida-Neto, 2012).  $\beta_{NES}$  is also principally linked to the HMSC-slope. However, in contrast to turnover, it is generated by a shared response of the species to the environmental covariate. That is, a large HMSC-slope mean with a comparatively small variance is needed for all species to respond in the same direction. In this way, the occurrence probability of all species increases (or decreases) with the value of the environmental covariate, and so does species richness. The variation explained by the HMSC-intercept is very small and thus plays a minor role.

It should be noted that the detailed interpretation of the coefficient estimates of the HMSC-intercept has to be made with caution. The species occurrences are modeled by a non-linear link function and the HMSC-intercept mean is not always 0. In practice what it means is that, for example, a larger HMSC-intercept variance may not increase common and rare species in a symmetrical way. This is likely the reason why the Sørensen dissimilarity model had a lower coefficient of determination ( $\text{adj-}R^2$ ) compared to the other models, since this index is largely influenced by the species prevalences. If the focus of the thesis were on understanding the relationship between the variation in species rarity/commonness and the beta-diversity measures, the mean and variance of the species prevalences should be used in the models instead of the HMSC-intercept.

A preliminary HMSC analysis to explore the typical variation in the species niches could not be performed due to time and resource constraints. This could have aided in the determination of the input parameters for the simulation. Nevertheless, the HMSC-intercept was parameterized so the simulated species would span the entire range of baseline occurrence probabilities, which in turn produced communities with different degrees of filling. In contrast, the parameterization of the HMSC-slope could have taken any other values since it models the direction and strength of the species responses to the environmental covariates (the HMSC-

slope could theoretically range from  $-\infty$  to  $+\infty$ ). Larger HMSC-slopes than the ones considered in this study would produce stronger species responses to the environmental covariates, while negative HMSC-slopes would just generate the same mirror pattern. In the linear models, the contribution of the mean HMSC-slope coefficient estimates to the nestedness and turnover dissimilarities would remain in the same direction or change to the opposite direction, respectively. For example, if the HMSC-slopes were modelled to range negative values, the nestedness pattern would be driven by a shared negative response of the species to the environmental covariate, instead of positive as in this study. In conclusion, although the simulated communities do not encompass the entire range of variation of real communities, the interpretation of the results are qualitatively applicable to any other community.

An obvious limitation of this research is that, while simulations can be a great tool for exploring some of the underlying mechanisms that structure ecological communities, they typically rely on highly simplified assumptions. Here, the species niches (HMSC-intercept and HMSC-slope) were modelled by a single environmental covariate and simulated to be fully random and independent. However, in real communities, species niches are considered to be an n-dimensional space (Hutchinson, 1957) and they are influenced by the species functional traits and, to some extent, structured by phylogenetic relationships.

## 4.2 Guidelines and considerations for empirical studies

This section aims at providing a short recommendation for empirical community ecologists interested in disentangling the contribution of nestedness and turnover to beta-diversity patterns.

In this study the community data was simulated by following the structure of the HMSC model and later analysed in terms of total, turnover and nestedness-resultant dissimilarities. However, the workflow for an empirical community ecologist would consist on calculating the dissimilarity indices (Fig. 4.1A) and inferring the HMSC parameters from the community data (Fig. 4.1B-G). That is, the workflow is in the reverse direction. Nevertheless, the interpretation of the niche parameters and

their relationship with the beta-diversity indices remains the same as in the present study. The results of this thesis can therefore be used to relate the outputs of these complementary approaches to each other, enabling one to link the observed patterns (as described by beta-diversity indices) to the underlying processes (as identified by HMSC) in real communities.

When analysing empirical community data with the HMSC model we may include many environmental covariates that we hypothesize that are important in explaining the species occurrences. The covariates may affect in different strengths and directions, making the link between the species niches and the dissimilarity patterns not so evident. In order to determine the axes of environmental variation with the highest influence on the community structure, we can partition the variation explained by the model (Fig. 4.1B; Ovaskainen & Abrego, 2020).

The posterior mean estimates of the intercept and slopes can then be visualized with the built-in heatmap function of the `hmsc` R package (Tikhonov et al., 2019). As suggested by the results of this thesis, negative intercepts (independently of the slopes) indicate species rarity, which in turn suggests high total dissimilarity between sites (Fig. 4.1C). Communities with high turnover are expected to show a pattern of positive and negative responses of the species to the main driver(s) of community structure, indicating variation in the species niches (Fig. 4.1D). Finally, a highly nested community would be characterized by the species having a similar response to the environmental covariate (i.e.: high niche overlap; Fig. 4.1E). Similarly, we can also visualize the raw species co-occurrences, that is, the associations between species. This plot reflects the differential habitat preferences of the species, together with their biotic interactions. In assemblages with high turnover we would expect to see positive and negative associations, since some species are replaced by others along the environmental gradient and therefore do not tend to occur together (Fig. 4.1F). On the other side, in a nested system we would expect to see mostly positive associations, since most species respond in the same direction (Fig. 4.1G).

Apart from deterministic niche-based processes, stochasticity may also play an important role in structuring some communities (Gilbert & Levine, 2017; Hubbell, 2001). In particular, ecological drift is expected to increase beta diversity in small assemblages, with the dissimilarity being unrelated to the environmental conditions



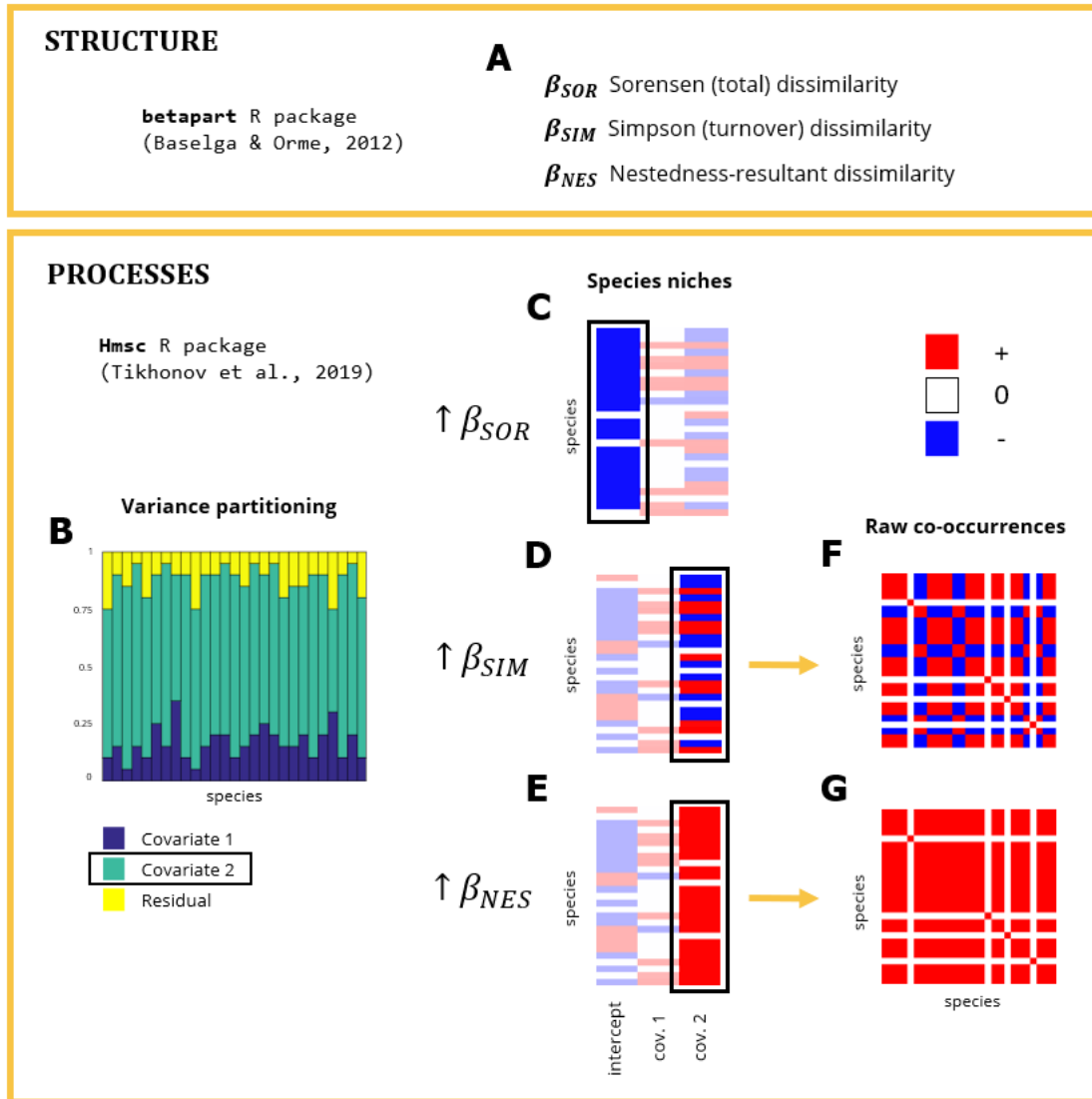


Figure 4.1: Hypothetical example of an empirical study. Panel A describes the dissimilarity indices used to describe the structure of the community. The other panels correspond to the HMSC analysis. Panel B illustrates the variance partitioning of the HMSC model, covariate 2 explains most of the variation of the response variable (and therefore it is the most important variable). Panels C, D and E illustrate the species niches for a hypothetical community with high total dissimilarity, high dissimilarity attributed to turnover, and high nestedness-resultant dissimilarity, respectively. The important covariates for each scenario are highlighted. Panels F and G illustrate the raw species co-occurrences for a community with high turnover and high nestedness resultant dissimilarity, respectively. In panels C-G positive values (positive responses to the model covariates and positive species associations) are colored in red, negatives values in blue.

(Hubbell, 2001). The stochastic variability in dispersal might also cause different degrees of nestedness, specially in communities mainly assembled by colonization (Ulrich & Zalewski, 2007). Since it is very difficult to accurately model all the deterministic processes in a system, the unexplained variation attributed to stochasticity

and missing covariates can not be separated, hampering the interpretation of the processes driving the dissimilarity.

Certainly, this study has also demonstrated the importance of reporting the intercept when applying the HMSC model (in generalized linear modeling the intercept is rarely given any interpretation). To emphasize this, let us consider two hypothetical communities, A and B, in both of which all species appear at random. This means that the species occur independently of each other (no species co-occurrences) and show no habitat preferences. However, most of the species in community A are rare (e.g.: they have a prevalence of 5%) while the species in community B are more common (e.g.: prevalence of 10%). From the point of HMSC, the only difference between these assemblages is in the intercept, which measures the overall baseline occurrence probability of the species. Therefore, a researcher applying the HMSC model and focusing only in the environmental covariates may equivocally conclude that there are no profound differences between both communities. Instead, a researcher applying the beta-diversity indices will conclude that both communities differ in their total dissimilarity ( $\beta_{SOR}$ ) and thus may be different.

### 4.3 Concluding remarks and future directions

As hypothesized, the results of this simulation study demonstrate that the beta diversity indices and the HMSC framework are strongly related to each other, despite being developed from very different conceptual starting points. While the dissimilarity indices provide a descriptive measure of the community structure, the HMSC analysis provides information on the underlying mechanisms generating the observed patterns. Therefore, the most comprehensive evaluation of the structure of ecological communities and the processes determining the diversity patterns can be achieved by combining both frameworks (as illustrated in Fig. 4.1).

The outputs of the HMSC model (intercept and slope) have proved to be able to characterize the different axes of niche variation generating nestedness and turnover. Further research, especially empirical studies, should focus on including species-specific traits and phylogenies in the HMSC analysis so as to achieve a better understanding on the underlying mechanisms shaping the species niches.

# Acknowledgements

I would like to thank my supervisor Otso Ovaskainen for mentoring me and providing invaluable assistance. This research was funded by the Organismal and Evolutionary Research Programme of the University of Helsinki and by The University of Helsinki Alumni Association Fund.

Finally, I would like to thank my parents for their immense patience and support during this whole process.

# Bibliography

- Abrego, N., Norberg, A., & Ovaskainen, O. (2017). Measuring and predicting the influence of traits on the assembly processes of wood-inhabiting fungi. *Methods in Ecology and Evolution*, *105*(4), 1070–1081. <https://doi.org/10.1111/1365-2745.12722>
- Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*(1), 134–143. <https://doi.org/10.1111/j.1466-8238.2009.00490.x>
- Baselga, A. (2012). The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, *21*(12), 1223–1232. <https://doi.org/10.1111/j.1466-8238.2011.00756.x>
- Baselga, A., & Orme, C. D. L. (2012). Betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution*, *3*(5), 808–812. <https://doi.org/10.1111/j.2041-210X.2012.00224.x>
- Gibson-Reinemer, D. K., Sheldon, K. S., & Rahel, F. J. (2015). Climate change creates rapid species turnover in montane communities. *Ecology and Evolution*, *5*(12), 2340–2347. <https://doi.org/10.1002/ece3.1518>
- Gilbert, B., & Levine, J. M. (2017). Ecological drift and the distribution of species diversity. *Proceedings of the Royal Society B*, *284*(1855), 20170507. <https://doi.org/10.1098/rspb.2017.0507>
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
- Hutchinson, G. (1957). *Concluding remarks* (Vol. 22). <https://doi.org/10.1101/SQB.1957.022.01.039>
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, *11*, 37–50.

- Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, *72*(3), 367–382. <https://doi.org/10.1046/j.1365-2656.2003.00710.x>
- Legendre, P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, *23*(11), 1324–1334. <https://doi.org/10.1111/geb.12207>
- Legendre, P. (2007). Studying beta diversity: Ecological variation partitioning by multiple regression and canonical analysis. *Journal of Plant Ecology*, *1*(1), 3–8. <https://doi.org/10.1093/jpe/rtm001>
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*(7), 601–613. <https://doi.org/10.1111/j.1461-0248.2004.00608.x>
- Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O’Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., ... Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monograph*, *89*(3), e01370. <https://doi.org/10.5281/zenodo.2637812>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O’Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *Vegan: Community ecology package* [R package version 2.5-7]. <https://CRAN.R-project.org/package=vegan>
- Ovaskainen, O., & Abrego, N. (2020). *Joint Species Distribution Modelling: With applications in R*. Cambridge University Press.
- Ovaskainen, O., Abrego, N., Halme, P., & Dunson, D. (2016). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, *7*(5), 549–555. <https://doi.org/10.1111/2041-210X.12501>

- Ovaskainen, O., Rybicki, J., & Abrego, N. (2019). What can observational data reveal about metacommunity processes? *Ecography*, *42*(11), 1877–1886. <https://doi.org/10.1111/ecog.04444>
- Ovaskainen, O., & Soininen, J. (2011). Making more out of sparse data: Hierarchical modeling of species communities. *Ecology*, *92*(2), 289–295. <https://doi.org/10.1890/10-1251.1>
- Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F. G., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecology Letters*, *20*(5), 561–576. <https://doi.org/10.1111/ele.12757>
- Patterson, B. D. (1987). The principle of nested subsets and its implications for biological conservation. *Conservation Biology*, *1*(4), 323–334. <https://doi.org/10.1111/j.1523-1739.1987.tb00052.x>
- Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*, *241*, 1–31. <https://doi.org/10.2475/ajs.241.1.1>
- Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should beta-diversity inform biodiversity conservation? *Trends in Ecology and Evolution*, *31*(1), 67–80. <https://doi.org/10.1016/j.tree.2015.11.005>
- Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, *5*, 1–34.
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>
- Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2019). Joint Species Distribution Modelling with the r-package hmsc. *Methods in Ecology and Evolution*, *11*(3), 442–447. <https://doi.org/10.1111/2041-210X.13345>
- Ulrich, W., & Almeida-Neto, M. (2012). On the meanings of nestedness: Back to the basics. *Ecography*, *35*(10), 865–871. <https://doi.org/10.1111/j.1600-0587.2012.07671.x>

- Ulrich, W., Almeida-Neto, M., & Gotelli, N. J. (2008). A consumer's guide to nestedness analysis. *Oikos*, *118*(1), 3–17. <https://doi.org/10.1111/j.1600-0706.2008.17053.x>
- Ulrich, W., & Zalewski, M. (2007). Are ground beetles neutral? *Basic and Applied Ecology*, *8*(5), 411–420. <https://doi.org/10.1016/j.baae.2006.08.002>
- Virkkala, R., & Lehikoinen, A. (2017). Birds on the move in the face of climate change: High species turnover in northern europe. *Ecology and Evolution*, *7*(20), 8201–8209. <https://doi.org/10.1002/ece3.3328>
- Whittaker, R. (1960). Vegetation of the siskiyou mountains, oregon and california. *Ecological Monographs*, *30*, 280–338. <https://doi.org/doi/10.2307/1943563>

# Appendix (simulation R code)

```
library(betapart)
library(MASS)

values= array(dim=c(1000, 7, 1))
#1000 communities, 7 columns (HMSC+beta-div indices), 100 replicates
for (j in 1:1){ #100 replicates
  ns = 100 #number of species
  beta= list()
  mean_int= vector()
  var_int= vector()
  mean_slope= vector()
  var_slope= vector()
  for (i in 1:1000){ #1000 communities
    V2= matrix(c(runif(1, 0, 2), 0, 0, runif(1, 0, 2)), nrow=2)
    #variance covariance matrix
    mu = rbind(rep(runif(1, -2 , 2), 100), rep(runif(1, 0, 5), 100))
    #intercept and slope expectations
    B= matrix(mvrnorm(n = 1, mu = as.vector(mu), Sigma = kronecker(diag(ns),
      V2)), ncol = ns) #pecies niches (intercept and slope)
    beta[[i]]= rbind(B[1,], B[2,])
    mean_int[i]= mean(beta[[i]][1,]) #intercept mean
    var_int[i]= var(beta[[i]][1,]) #intercept variance
    mean_slope[i]= abs(mean(beta[[i]][2,])) #slope mean
    var_slope[i]= var(beta[[i]][2,]) #slope variance
  }

  n = 50 #number of sites
  env_covariate = cbind(rep(1, n), rnorm(n, 0, 1)) #environmental covariate
  beta_div= list()
  for(i in 1:length(beta)){
    L= env_covariate %*% beta[[i]] #linear predictor
    Y= 1*((L + matrix(rnorm(n*ns), ncol = ns)) > 0)
    #community matrix. Presence-absence data modelled with the probit
    link function.
    beta_div[[i]]= beta.multi(Y, index.family = "sorensen")
    #beta-diversity indices
  }
}
```



```
}  
  
matrix_beta= t(matrix(unlist(beta_div), nrow=3))  
da_1000= cbind(mean_int, var_int, mean_slope, var_slope, matrix_beta)  
#HMSC summaries + beta-diversity indices  
values[,j]= da_1000  
colnames(values)= c("mean_int", "var_int", "mean_slope", "var_slope",  
                    "beta.SIM", "beta.SNE", "beta.SOR")  
}
```