Ecography

ECOG-04444
Ovaskainen, O., Rybicki, J. and Abrego, N. 2019. What can observational data reveal about metacommunity processes? – Ecography doi: 10.1111/ecog.04444

Supplementary material

Appendix 1

Supporting information for "What can observational data reveal about metacommunity processes?"

July 20, 2019

1 Technical details on the simulation model

We simulated metacommunity dynamics using an agent-based resource-consumer model in continuous space and time. The model can be technically described as a spatiotemporal point-process [1]. The parameters of the model and their numerical values are summarized in Table S1. In the following, we describe the model in detail. We use Tophat$(a, b; d)$ to denote the top-hat kernel with integral r and length-scale ℓ, that is, we have

$$\text{Tophat}(r, \ell; d) = \begin{cases} r/(\pi\ell^2) & \text{if } 0 \le d \le \ell \\ 0 & \text{otherwise.} \end{cases}$$

1.1 Resource generation

To generate spatiotemporal variation in resource availability, we assumed that there are $P = 4$ different habitat patch types that generate a total of $R = 24$ resource types, so that each patch generates $L = R/P = 6$ distinct resource types. The resource particles are generated within circular patches which are described by their size (radius λ) and quality (per-patch rate of resource production σ). A patch of type $j \in \{1, \ldots, P\}$ centered at location $\mathbf{x}$ will generate resource particles of type (j, k), where $k \in \{1, \ldots, L\}$ according to the kernel Tophat(σ_j, λ_j). That is, resource particles appear at rate σ_j uniformly at random within a disk of radius λ_j centered at $\mathbf{x}$. Each resource particle is removed at rate q unless they are consumed by some individual (and hence removed when consumed).

1.2 Species dynamics

In each scenario, we consider a community of $S = 24$ species which utilize the resource particles for their survival and reproduction. Each individual can be at two states: resource-deprived or resource-satiated. When at the satiated state, the individual changes to the deprived state at rate h. When at the deprived state, the individual can change to the satiated state by consuming a resource particle from its proximity. If the consumption of resource type is within the niche of species, we assume that resource consumption can take place at maximum distance θ between the individual and resource particle. Within this distance, per-particle consumption rate is u. A resource-deprived individual may die, which takes place at per-individual rate m.

Satiated individuals produce propagules at rate f, which disperse so that they are equally likely to be deposited anywhere within distance δ from their mother. In other words, a satiated individual produces a propagules onto a point at distance d at per-unit area rate Tophat$(f, \delta; d)$. The deposited propagules emerge without delay as new offspring individuals that are in the resource-deprived

state. Moreover, we assume that every species has immigration of new offspring individuals outside the simulation area occurring at rate I per unit area. The new offspring individuals are initially in the resource-deprived state, and consequently those that immigrate in an area with no resource production will eventually die and never produce new offspring.

Niche scenarios. Recall that we have $P = 4$ patch types each producing $L = 6$ distinct resource types so that there are $S = R = 24$ different resource types in total. We consider three different niche scenarios for the species community:

N1. generalists (no specialization): species of any type can consume any type of resource,

N2. partial specialization: species of type i only consumes resource particles produced by a single patch type (out of all 4 patch types), i.e., species i consumes any of the $L = 6$ resources types produced by patch type $j = (i \bmod P) + 1$,

N3. strict specialization: there is a one-to-one correspondence between species the $S = 24$ species and the $R = 24$ resource types so that each species is associated to a unique resource type.

Dispersal scenarios. There are three dispersal scenarios for the propagules:

D1. all species have short dispersal distance, that is, $\delta = 1$,

D2. all species have long dispersal distance, that is, $\delta = 5$, and

D3. half of the species have short dispersal distance ($\delta = 1$) and other half have long dispersal distance ($\delta = 5$), i.e., species i has short dispersal distance if $i \bmod 2 = 0$ and long dispersal distance otherwise.

1.3 Patch dynamics

The landscape structure can be either static or dynamic. In the latter case, new patches of type j are generated with per-unit-area rate b_j and existing patches of type j disappear with per-patch rate d_j. We consider three cases for patch dynamics:

P1. ephemeral short-term patches,

P2. ephemeral long-term patches,

P3. static patches.

For static patches, we have $b_j = d_j = 0$, and thus, the habitat structure is determined by the initial configuration of the system.

Short-term and long-term patches. For short-term ephemeral patches, we set the dissipation rate d_j of all patch types to be $d_j = 1/4$ so that the expected life time of a patch is four time units; note that this is the same as the attrition parameter $h = 1/4$ that determines how fast a satiated individual turns back into the resource-deprived state. For long-term ephemeral patches, we set the dissipation rate to be $d_j = 1/20$ so that the expected life time of a patch is 20 time units. We set b_j such that the expected density of patches of type j is ρ_j.

Resource generation rate and habitat quality. The resource generation rate of a single resource (j,k) at location $\mathbf{x}$ is given by

$$\omega_j(t,\mathbf{x}) = \sum_{\mathbf{p}\in\mathscr{P}_j(t)} \text{Tophat}(\sigma_j,\lambda_j;\text{dist}(\mathbf{p},\mathbf{x}))$$

with $\mathscr{P}_j(t)$ denoting the locations of patches of type j at time t and $\text{dist}(\mathbf{p},\mathbf{x})$ the distance between the locations $\mathbf{p}$ and $\mathbf{x}$ in the two-dimensional torus. While $\omega_j(\cdot)$ may vary over t and $\mathbf{x}$, we ensure that its expected value (over both time and space) remains constant in all landscapes and scenarios.

To this end, consider patches of type j and any point $\mathbf{x}$ in the landscape. Let X be the number of patch centers of type j within distance λ_j of $\mathbf{x}$ at time t. The expected value of X is

$$\mathbb{E}[X] = \pi\lambda_j^2\rho_j,$$

where ρ_j is the density of patch centers. Since each patch within distance λ_j from $\mathbf{x}$ contributes to the resource generation rate (of a single resource type) at that particular point by σ_j/A, where $A=\pi\lambda_j^2$ is the area of the patch, the expected resource generation rate at point $\mathbf{x}$ is

$$\mathbb{E}[\omega_j(t,\mathbf{x})] = \mathbb{E}[X]\cdot\frac{\sigma_j}{\pi\lambda_j^2} = \rho_j\sigma_j.$$

Note that for ephemeral patches the density $\rho_j(t)$ at time t itself is a random variable. However, we choose the rates b_j and d_j so that the expectation of $\rho_j(t)$ remains constant.

1.4 Environmental gradients

We consider two cases for the large-scale structure in habitat types:

E1. no environmental gradient: all patch types are distributed uniformly at random in the landscape,

E2. environmental gradient: patch types have spatial autocorrelation.

In the first case (E1), when a new patch appears, it is equally likely to appear anywhere in the landscape. In the second case (E2), the relative location of a new patch of type $j\in\{1,2,3,4\}$ is sampled according to the density function $g_j(x,y)$ on the unit square, where

$$\begin{aligned}
g_1(x,y) &= 1+\sin(2\pi x)\\
g_2(x,y) &= 1+\sin(2\pi x+\pi)\\
g_3(x,y) &= 1+\sin(2\pi y)\\
g_4(x,y) &= 1+\sin(2\pi y+\pi).
\end{aligned}$$

The relative locations are then mapped to actual coordinates in the $U\times U$ torus, where U is the parameter controlling landscape size. See Figure S1 for an example of a landscape with a environmental gradient generated according to the above functions. The expected patch density is the same for each patch type, as we have that

$$\int_{\mathbf{x}\in[0,1]^2} g_j(\mathbf{x})d\mathbf{x} = 1.$$

Figure S4 (environmental gradient) and Figure S5 (no gradient) illustrate how the distribution of species and resource units is influenced by the environmental gradient scenario.

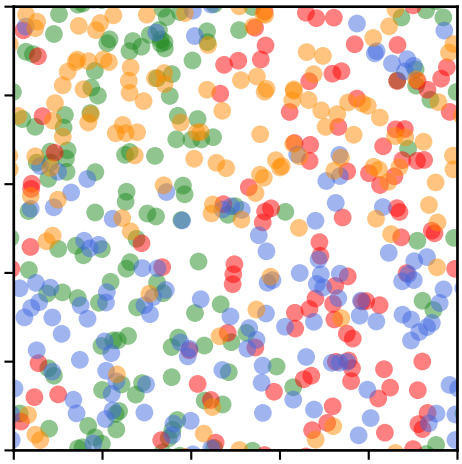


Figure S1: A snapshot of a landscape, where the patches have environmental gradients. The coloured circles are patches and the colours indicate the four different patch types. The green patches are predominantly on the left-hand side of the landscape, whereas the red patches are on the right. The orange patches are most likely to occur at the top half of the landscape, whereas blue patches are likely to appear in the bottom half.

1.5 Landscape structure and heterogeneity in habitat types

We consider different spatial patterns for the habitat structure:

H1. patchy landscape, and

H2. continuous landscape.

Furthermore we vary the two scenarios above even further by assuming either

Q1. uniform patch quality, that is, all patches have the same resource production rate, or

Q2. varying patch quality, that is, there are "source" patches with high resource production rate and "sink" patches with low resource production rate.

Normalizing the habitat quality. In all scenarios with uniform patch quality (Q1), we fix $\sigma_j \cdot \rho_j$ to be a constant for each patch type j. That is, the *expected* rate of resource production per-unit-area is

$$\omega = R \sum_{j=1}^{P} \sigma_j \rho_j.$$

For static landscapes (scenario P3), ω remains constant throughout time, but for ephemeral patches (scenarios P1 or P2) we choose the parameters so that the expectation of $\omega(t)$ remains ω for all $t > 0$. Note that even though the parameter ω remains the same in all scenarios, the resource particle generation rate (i.e. the habitat quality) can vary at different spatial locations. Having low length scale parameter λ_j produces patchy environments, whereas high λ_j results in more (locally) homogeneous habitat quality; see Figure S2 and Figure S3 for examples.

Source-sink scenario. In the source-sink scenario (Q2) with both high and low quality patches, we assume that there are $2P$ patch types (j,s), where $s \in \{0,1\}$ and $j \in \{1,\ldots,P\}$ as before. Patches

of type $(j,0)$ denote source patches of type j, whereas $(j,1)$ are sink patches of type j. Both patches otherwise behave exactly the same as in case Q1 except that source patches have higher resource production rate than sink patches, that is, $\sigma_{j,0} > \sigma_{j,1}$. In particular, both produce the same types of resource particles. In the source-sink scenario, we use $\sigma_j(\cdot)$ as a short-hand for $\sigma_{j,0}(\cdot) + \sigma_{j,1}(\cdot)$. The expected total resource production rates remain the same as in the scenario Q1.

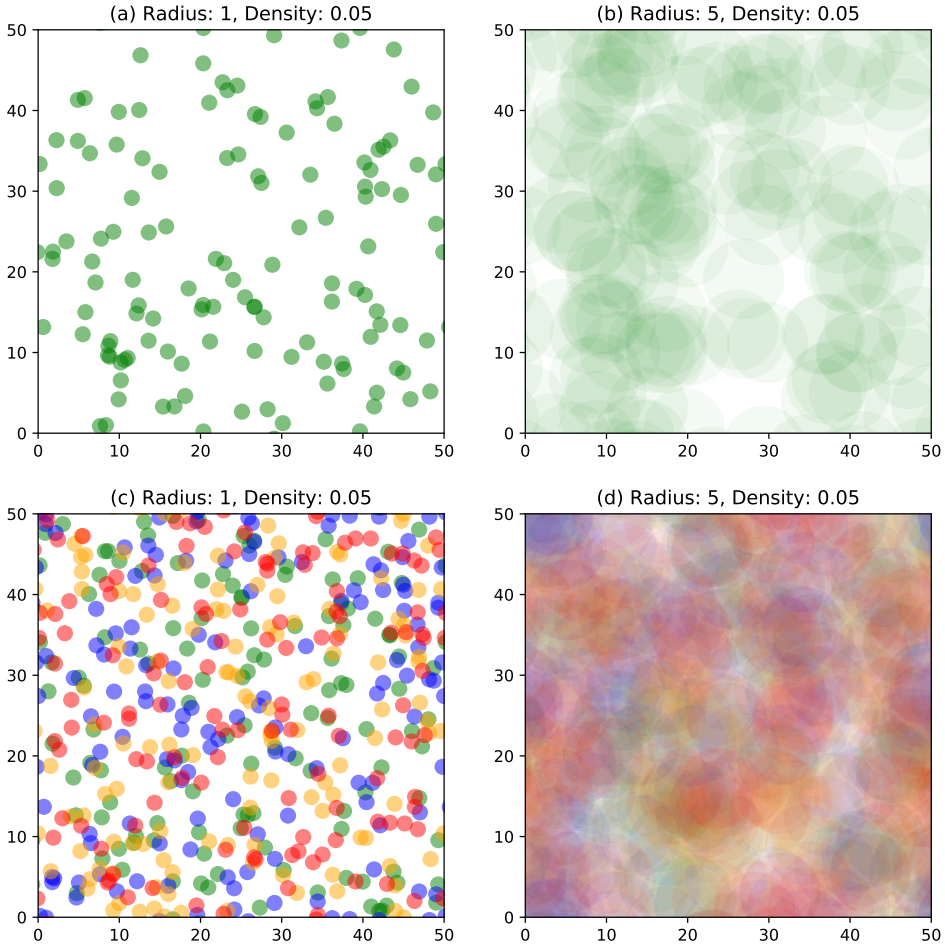


Figure S2: Examples of the patchy and continuous habitat scenarios. Here there is no environmental gradient, i.e., the patches follow complete spatial randomness. The top panels illustrate the distribution of a single patch type and the lower panels show the landscape with all $P = 4$ patch types plotted. The panels on the left give the patchy case $\lambda = 1$. The panels on the right illustrate the continuous case $\lambda = 5$. Each patch type has density $1/20$.

1.6 Initial conditions

At time $T = 0$, the system is initialized as follows. Patches of type j are placed randomly with density ρ_j. If the patches follow an environmental gradient, the patch types are sampled according to the density functions described in Section 1.4. Otherwise, the patch locations are distributed uniformly at random in space. Initially, no species is present, i.e., there are no resource-deprived or

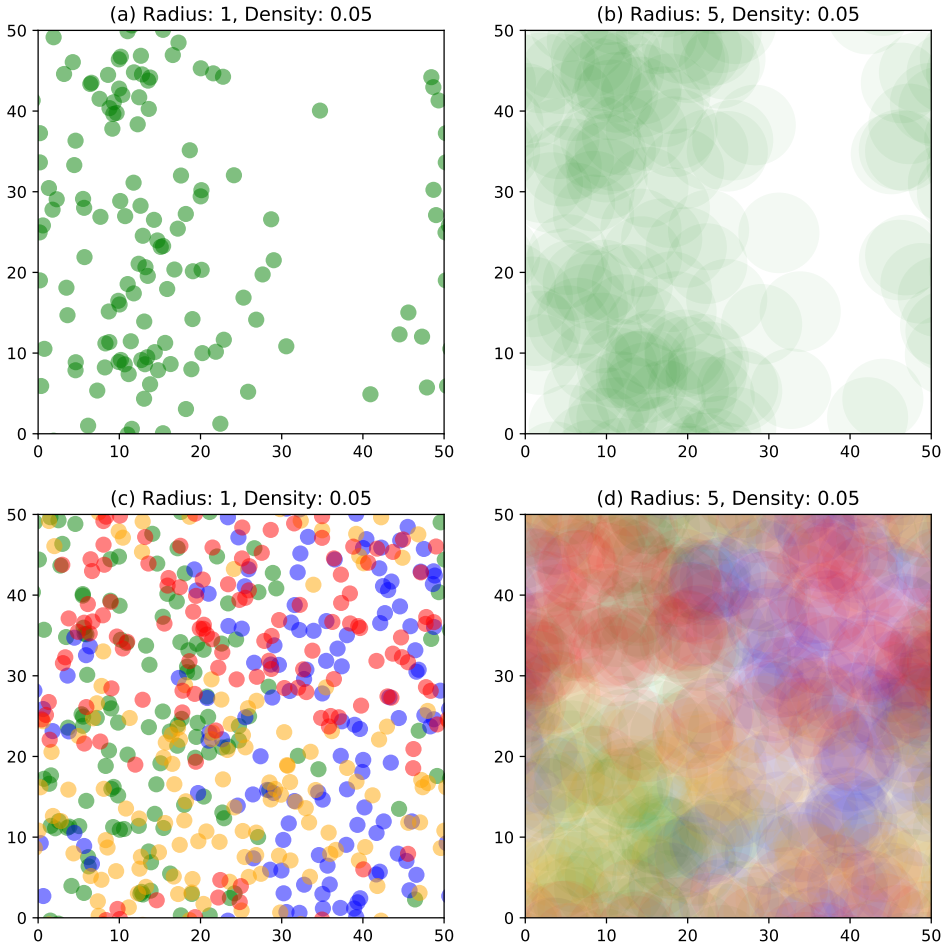


Figure S3: Examples of the patchy and continuous habitat scenarios with environmental gradient. Otherwise, the four panels are as in Figure S2.

resource-satiated individuals. However, the species populations' are eventually established by the fact that resource-deprived individuals immigrate into the landscape at per-unit-area rate I.

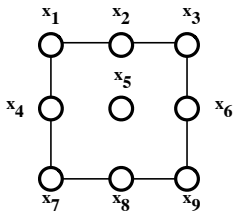
2 Details on simulation and data sampling

We simulated the dynamics of the models on a torus of size $U \times U$, with $U = 50$, for $T = 1000$ time units. As earlier work (Smith and Lundholm, 2010; Münkemüller et al., 2012; Tucker et al., 2016; Clappe et al., 2018) we wished to ignore transient dynamics and our preliminary tests indicated that $T = 1000$ time units was sufficient for reaching the stationary state. We assumed that a virtual ecologist acquired data from the final state of the simulation, placing 100 study plots of size 1×1 placed into a regular 10×10, grid, with the distance between the plot centers being $D = 2$ spatial units (Figure 1B in the main text and Figure S4D in the supplementary information).

Additionally, the virtual ecologist collected validation data to be used to test the predictive ability

of the JSDM approach, acquired using the same design, but plots being located with respect to the main data (used to fit the JSDM) as white and black squares in a chess board. The researcher scored the presence-absence of resource-satiated individuals for each species in each study plot, resulting in the matrix $\mathbf{Y}$, where the columns correspond to species and rows to sampling locations.

Assaying for habitat quality. The virtual ecologist collected data also on habitat quality, separately for each of the $P = 4$ habitat types, assaying the resource generation rate σ_j as the average over 9 distinct equally spaced points $\mathbf{x}_1, \ldots, \mathbf{x}_9$ placed on a 3×3 grid within each study plot:



The assayed quality for habitat type j was the average of $\omega_j(t, \mathbf{x}_1)$, ..., $\omega_j(t, \mathbf{x}_9)$ at time $t = 1000$.

Sampling scenarios. We assumed two sampling scenarios for the virtual ecologist:

M1. either the virtual ecologist acquired covariate data on all $P = 4$ habitat types, or

M2. only on three out of the four habitat types, the remaining one thus being missing covariate.

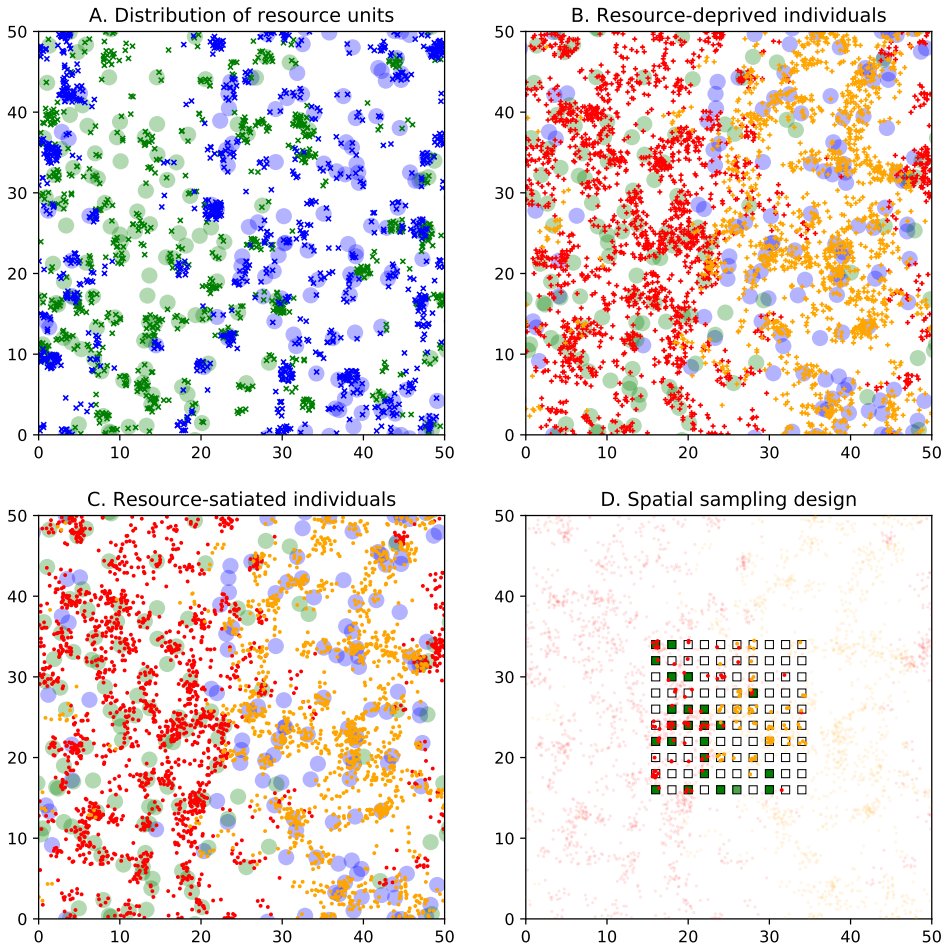


Figure S4: An example scenario with two species (red and orange points) and two patch types (green and blue circles). Here, the red species uses resources of green patches and orange species uses resources of blue patches. The panels depict a snapshot of a patchy landscape with fast patch turnover and an environmental gradient. (A) The distribution of resource particles. The × symbols represent resource particles. (B) The distribution of resource-deprived individuals of both species. (C) The distribution of resource-satiated individuals of both species. (D) Example of the sampling design using a 10×10 grid of study plots of size 1×1. The colour of the study plot represents assayed habitat quality of the *green* patch type only. The red and green points represent satiated individuals.

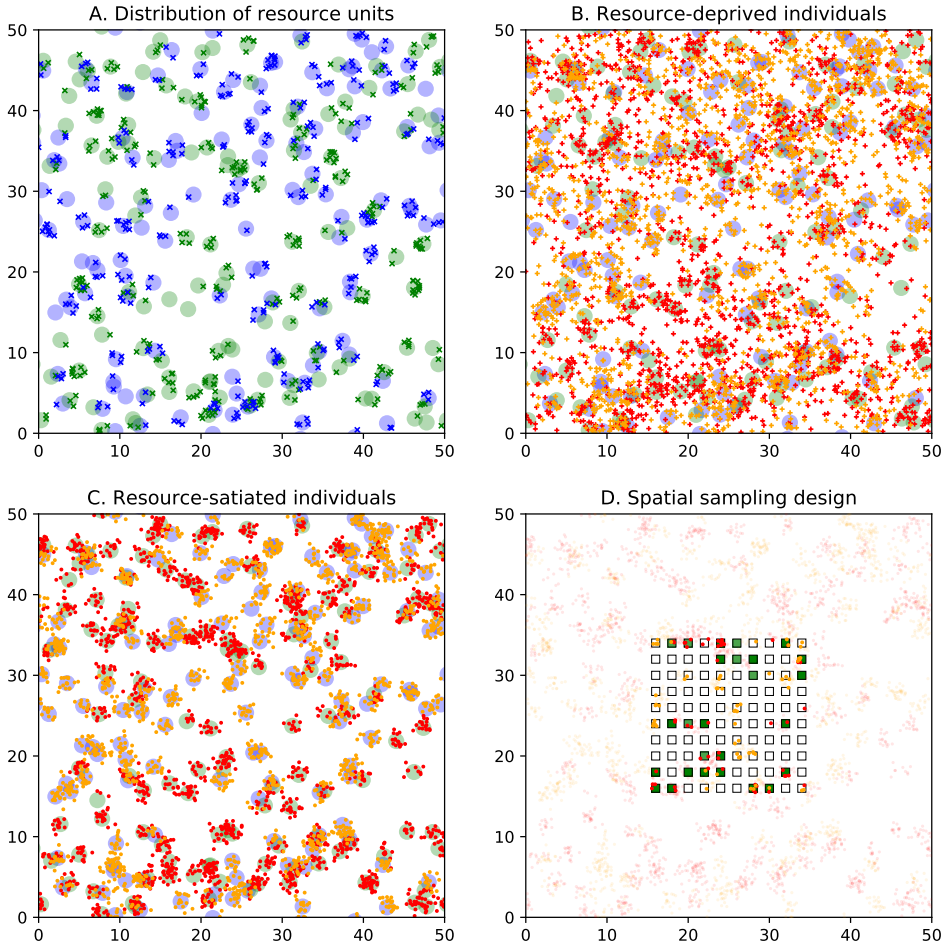


Figure S5: An example scenario similar to Figure S4. Here the patches are static (no patch turnover) and there is no environmental gradient (both patch types follow complete spatial randomness).

3 Details on the simulated scenarios

To generate scenarios that span the range of the classic metacommunity paradigms, we considered a number of parameterizations of the individual-based model and sampling scenario for the virtual ecologist, presented with the help of the nine choices (C1–C9) that are described conceptually in Figures 2 and 3 of the main text and summarized numerically in Table S1. These binary choices C1–C9 determine the scenarios as follows:

C1. environmental gradient: no gradient (E1) vs. gradient (E2),

C2. landscape structure: patchy (H1) vs. continuous (H2),

C3. patch quality: uniform (Q1) vs. varying patch quality (Q2),

C4. temporal habitat structure: ephemerality (P1 or P2) vs. static patches (P3),

C5. patch turnover rate: slow (P1) vs. fast (P2),

C6. specialization: generalists (N1) vs. specialists (N2 or N3),

C7. level of specialization: partial (N2) vs. strict specialists (N3),

C8. dispersal strategy: uniform dispersal (D1 or D2) or variation in dispersal strategy (D3),

C9. dispersal distance: short (D1) vs. long (D2),

The resulting metacommunity scenarios are determined by 9 binary Choices C1-C9 described above, which yield a total number of $2^9 = 512$ combinations. However, some of the combinations are not relevant: C5 has an effect only if C4 defines a dynamic landscape, C7 has an effect only if C6 defines specialization in resource use, and C9 has effect only if C8 determines that all species follow an identical dispersal scenario. Consequently, the total number of different metacommunity scenarios is 216. Finally, Choice C10 determines which covariates the virtual ecologist samples:

C10. sampling scenario: all covariates (M1) vs. missing covariates (M2).

References

[1] Ovaskainen, O., Finkelshtein, D., Kutoviy, O., Cornell, S., Bolker, B. and Kondratiev, Y., 2014. A general mathematical framework for the analysis of spatiotemporal point processes. *Theoretical Ecology 7*, 101–113.

[2] Clappe S. Dray, S., Peres-Neto, P., 2018. Beyond neutrality: disentangling the effects of species sorting and spurious correlations in community analysis. *Ecology 99*, 1737–1747.

[3] Münkemüller, T., de Bello, F., Meynard, C.N., Gravel, D., Lavergne, S., Mouillot, D., Mouquet, N., Thuiller, W., 2012. From diversity indices to community assembly processes: a test with simulated data. *Ecography 35*, 468–480.

[4] Smith, T.W., Lundholm, J.T., 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography 33*, 648–655.

[5] Tucker, C.M., Shoemaker, L.G., Davies, K.F., Nemergut, D.R., Melbourne, B.A., 2016. Differentiating between niche and neutral assembly in metacommunities using null models of β-diversity. *Oikos 125*, 778–789.

Table S1: Parameters of the simulation model and their default values

Parameter	Description	value
P	Number of patch types	4
L	Number of resource types per patch	6
R	Number of resource types in total	$PL = 24$
S	Number of species	24
U	Width and height of the landscape	50
λ	Patch radius	1 (patchy) or 5 (continuous)
σ	Resource production rate of a patch	4 if scenario is Q1, otherwise 6.4 for source patches and 1.4 for sink patches
ρ	Density of each type of patches	1/20
d	Patch death rate (turnover rate)	1/4 (fast) or 1/20 (slow)
b	Per-unit-area birth rate of each type of patches	$d \cdot \rho$
q	Death rate of resource particles	1/10
h	Transition rate from satiated to hungry state	1/4
θ	Maximum distance for resource consumption	1
u	Rate of resource consumption	1 if the resource type is within the species niche, otherwise 0
m	Death rate of hungry individuals	1
f	Propagule production rate	1
δ	Maximal dispersal distance of propagules	1 or 5
I	Immigration rate from outside (for each species)	1/100

4 The output metrics derived from the statistical approaches

As described in the main manuscript, we considered eighteen output metrics that are derived from five different types of analyses: analysis of habitat variation (HAB), beta-diversity indices (BETA), distance-based variation partitioning (db-VP), distance-based redundancy analysis (db-RDA), and joint species distribution modelling (JSDM). Here we give the details on how these output metrics were computed, and how we hypothesized them to relate to the simulated metacommunity processes.

Analysis of habitat variation (HAB)

Analysis of habitat variation is not a standard approach in community ecology, as the structure of the habitat is typically considered as part of the study design, or as explanatory variable, rather than a variable of interest itself. However, we decided to include an analysis of habitat variation as it brings direct information about habitat structure and thus combining its information with species data can be expected to improve the understanding of the assembly processes.

Output metric O1: Variance in habitat quality (V_{HAB}). We calculated V_{HAB} as standard deviation of habitat quality, calculated first separately for each habitat type and then averaged over the habitat types. We expected V_{HAB} to increase with the heterogeneity of the landscape, and hence to be larger i) for those landscapes that show marked gradients than for those that do not, ii) for patchy landscapes than for continuous landscapes, and iii) for landscapes that show variation in patch quality than for those that do not.

Output metric O2: Distance decay in habitat similarity (D_{HAB}). D_{HAB} was calculated as the slope of the linear model where the correlation between habitat qualities measured in two locations was regressed against the distance between those two locations, for data obtained for 500 randomly selected pairs of sampling locations. As D_{HAB} measures spatial turnover of habitats, we expected it to decrease especially fast (i.e. to show higher turnover) for landscapes with a large-scale gradient and for patchy landscapes.

Beta-diversity indices (BETA)

We computed three index-based measures partition beta-diversity as defined by Baselga (2010).

Output metric O3: Sørensen-based multiple-site dissimilarity (β_{SOR}). Sørensen dissimilarity measures the proportion of species shared between two communities and it incorporates both true spatial turnover and differences in richness (Baselga, 2010). We thus expected it to be influenced by all processes that generate spatial heterogeneity in the communities, especially heterogeneity of the landscape, specialisation in resource use, and short-distance dispersal.

Output metric O4: Simpson-based multiple-site dissimilarity (β_{SIM}). Simpson dissimilarity differs from Sørensen dissimilarity by being influenced only by the true spatial turnover and not by the variation in species richness (Baselga, 2010). We thus expected it to be influenced otherwise by the same factors as the Sørensen dissimilarity except not by whether the landscape is continuous or patchy, as we expected the patchiness of the landscape to influence especially variation in species richness (high species richness in patches versus low species richness in matrix).

Output metric O5: Nestedness-resultant multiple-site dissimilarity (β_{NES}). As its name indicates, nestedness dissimilarity evaluates whether communities occurring in different sites are nested, i.e. whether species found from one site are a subset of the species found from another site. Nestedness dissimilarity is computed as the difference between the Sørensen and Simpson dissimilarities. We expected it to separate especially between the patchy and continuous landscapes, as the patchy landscapes can be expected to show a strong nestedness structure: species poor communities sampled in the matrix are expected to be subsets of those sampled in patches.

Distance-based variation partitioning (db-VP)

Distance-based variance partitioning decomposes variation in community dissimilarity into components that can be explained by environmental and spatial variables. Following Legendre et al. (2005) and Smith and

Lundholm (2010), we used the vegan R-package (Oksanen et al., 2018) to calculate community dissimilarity (applying Bray-Curtis distance) from the occurrence data **Y**, environmental dissimilarity (applying Euclidean distance) from the environmental covariate matrix **X**, and spatial distance (applying Euclidean distance) from the spatial coordinates. We defined the total explained variance (V_{ABC}) as the R^2 of the linear regression where community dissimilarity was explained by both environmental and spatial distances. The total environmental variance (V_{AB}) was the R^2 of the linear regression where community dissimilarity was explained by environmental dissimilarity, and the total spatial variance (V_{BC}) was the R^2 of the linear regression where community dissimilarity was explained by spatial distance. To avoid high correlation among these three output metrics, we divided the environmental and spatial variances by the total variance. Thus, we considered the following three output metrics.

Output metric O6: Total explained variance (V_{ABC}). We expected the total explained variance to be highest for communities where variation in community structure can be related to the environmental or spatial predictors. That is, for communities that show species sorting due to environmental filtering (high specialisation level combined with variation in habitat structure), static landscapes, and species that follow short-distance dispersal.

Output metric O7: Environmental proportion (V_{AB}/V_{ABC}). We expected the environmental proportion to be high for those communities that show species sorting due to environmental filtering (high specialisation level combined with variation in habitat structure) but that do not show spatial structure beyond that explained by the environmental predictors. That is, for communities that follow long-distance rather than short-distance dispersal.

Output metric O8: Spatial proportion (V_{BC}/V_{ABC}). We expected the spatial proportion to be high for those communities that follow short-distance dispersal but do not show species sorting due to environmental filtering.

The R-code by which we computed these measures is given below (the Y matrix includes the community data, the X matrix the environmental covariate data, and xy includes the spatial coordinates):

```
distY=vegdist(Y)
distX=vegdist(X,method="euclidean")
distS=vegdist(xy,method="euclidean")

m=lm(distY~distX+distS)
sm=summary(m)
VABC=sm$r.squared
m=lm(distY~distX)
sm=summary(m)
VAB=sm$r.squared
m=lm(distY~distS)
sm=summary(m)
VBC=sm$r.squared
c(VABC,VAB/VABC,VBC/VABC)
```

Distance-based redundancy analysis (db-RDA)

As an alternative method to db-VP, we considered distance-based redundancy analysis (db-RDA). We expected these two to give largely similar results, as db-RDA also evaluates the amount of variation in community structure that can be explained by environmental and/or spatial predictors. We performed db-RDA following McArdle and Anderson (2001). To do so, we applied the varpart function of the vegan R-package, where we included both environmental and spatial predictors. The candidate spatial predictors were

generated with the listw.candidates function of the adespatial R-package (Dray et al., 2019). The selected spatial predictors that were included in the db-RDA were chosen with the listw.select function of the adespatial R-package. We defined the following three output metrics.

Output metric O9: Total explained variance ($R^2.adj$). We hypothesized that $R^2.adj$ behaves similarly to V_{ABC} because they are conceptually similar .

Output metric O10: Environmental proportion $\left(\frac{X1|X2}{R^2.adj}\right)$. We hypothesized that $\frac{X1|X2}{R^2.adj}$ behaves similarly to V_{AB}/V_{ABC} because they are conceptually similar.

Output metric O11: Spatial proportion $\left(\frac{X2|X1}{R^2.adj}\right)$. We hypothesized that $\frac{X2|X1}{R^2.adj}$ behaves similarly to V_{BC}/V_{ABC} because they are conceptually similar .

The R-code by which we computed these measures is given below (the Y matrix includes the community data, the X matrix the environmental covariate data, and xy includes the spatial coordinates):

```
candidates = listw.candidates(xy,nb=c("gab"),weights = c("binary","flin"))
modsel.Y = listw.select(Y,candidates, method = "FWD",
          MEM.autocor = "positive",p.adjust = TRUE)
MEM.spe = modsel.Y$best$MEM.select
vY=vegdist(Y,method = "bray")
vY[is.na(vY)]=0
VP = varpart(vY,X,MEM.spe)
vp = VP$part$indfract$Adj.R.squared[1:3]
c(sum(vp),vp[1]/sum(vp),vp[3]/sum(vp))
```

Joint species distribution modelling (JSDM).

As a model-based method, we used Hierarchical Modelling of Species Communities (Ovaskainen et al., 2017), which model belongs to the class of Joint Species Distribution Models (Warton et al., 2015). We considered the matrix **Y** of species occurrences as the response variables, the matrix **X** of habitat qualities as the explanatory variables. The coordinates of the sampling locations were used to fit a random effect through spatial latent variables (following Ovaskainen et al., 2016). We fitted a probit-regression model, from which we generated the following seven output metrics.

Output metric O12: Predictive power of the model (AUC). We computed the AUC statistic based on two-fold cross-validation, and thus it measures predictive rather than explanatory power. We expected AUC to increase with the predictability of the community, which we expected to increase with the same factors as V_{ABC} and/or $R^2.adj$ (see above).

Output metric O13: Variance attributed to random effects (V_{RAND}). V_{RAND} was computed as the proportion of explained variance (at the level of the linear predictor; see Ovaskainen et al. 2017) attributed to the spatial random effect and that cannot thus be attributed to the measured variation in resource availability. As V_{RAND} measures the proportion of the variance explained by spatial predictors, we expected it to behave as V_{BC}/V_{ABC} and/or $\frac{X2|X1}{R^2.adj}$ (see above).

Output metric O14: Evidence for resource use specialization (RUS). To compute RUS, we computed for each species the proportion of explained variance (at the level of the linear predictor; see Ovaskainen et al. 2017) attributed to each of the resource types, and measured species-specific specialization as the variance among these proportions. We then defined RUS as the average specialization over the species. We assumed RUS to be highest when the species are specialized in their habitat use, and when the landscape is spatially heterogeneous but temporally static so that the species can sort according to their resource use preferences.

Output metric O15: Proportion of species pairs with positive association (POS). We computed POS as the proportion of species that showed a positive residual association supported by at least 95% posterior probability. As our agent-based model did not involve any facilitative interactions, we expected positive associations to arise only as an artefact due to species responding to a missing environmental covariate.

Output metric O16: Proportion of species pairs with negative association (NEG). We computed NEG as the proportion of species that showed a negative residual association supported by at least 95% posterior probability. As the species compete in the agent-based model for the same resources, we expected this to generate negative associations, especially for the case where species are specialized and the resource distribution in the landscape is heterogeneous.

Output metric O17: Posterior mean of spatial scale of residual variation ($E[\alpha]$). We computed $E[\alpha]$ as the posterior mean of the spatial scale of the leading spatial latent factor. As $E[\alpha]$ measures the spatial scale at which the species communities show turnover not explained by environmental covariates, we expected $E[\alpha]$ to be higher for communities where the species followed long-distance dispersal than for communities where the species followed short-distance dispersal.

Output metric O18: Posterior support for spatially structured residual variation ($Pr[\alpha > 0]$). We computed $Pr[\alpha > 0]$ as the posterior probability by which the spatial scale of the leading spatial latent factor is greater than zero, i.e. the data shows a spatial signal (in the prior distribution for α, half of the prior-mass was assigned to $\alpha = 0$ and half to $\alpha > 0$). We expected $Pr[\alpha > 0]$ to be higher for communities where species followed long-distance dispersal than for communities that followed short-distance dispersal.

We sampled the posterior distribution with 1500 MCMC iterations with HMSC-R 3.0 (Tikhonov et al., 2019), out of which we discarded the first 500 as transient. We note that this amount of sampling is unlikely to lead to good MCMC convergence, but we applied it due to computational constrains (note that we needed to fit and cross-validate the model for a very large number of replicates). While compromised MCMC convergence may influence the absolute performance of the approach, we did not expect it to influence how the output metrics relate to the underlying process, except possibly by adding noise. In this sense, the results from HMSC analyses are conservative.

The R-code by which we computed these measures is given below (Y include the community data, XData include environmental covariate data, and xy are the spatial coordinates):

```
rL.site = HmscRandomLevel(sData =xy)
m = Hmsc(Y=Y, XData = XData,  distr="probit",
             studyDesign=studyDesign, ranLevels={list(site=rL.site)})

m = sampleMcmc(m, samples = 1000, thin=1,
         adaptNf=400, transient = 500, nChains = 1)
partition=createPartition(hM=m, nfolds=2, column="site")
predYCV = computePredictedValues(m, partition=partition)
MFCV = evaluateModelFit(hM=m, predY=predYCV)
AUC = mean(MFCV$AUC,na.rm=TRUE)
OmegaCor = computeAssociations(m)
supportLevel = 0.95
POS = (sum(OmegaCor[[1]]$support>(supportLevel))-m$ns)/(m$ns*m$ns-m$ns)
NEG = (sum(OmegaCor[[1]]$support<(1-supportLevel)))/(m$ns*m$ns-m$ns)
group = 1:(m$nc-1)
groupnames = group
VP = computeVariancePartitioning(m, group=group, groupnames = groupnames)
VRAND = rowMeans(VP$vals)[m$nc]
RUS = mean(sqrt(apply(VP$vals[-m$nc,],MARGIN = 2,FUN = var)))
```

```
mpost = convertToCodaObject(m)
MEA = mean(mpost$Alpha[[1]][,1][[1]])
SS = mean(mpost$Alpha[[1]][,1][[1]]>0)
c(AUC,POS,NEG,VRAND,RUS,MEA,SS)
```

References

Baselga, A., 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecol.Biogeogr.* 19, 134-143.

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2019. adespatial: Multivariate Multiscale Spatial Analysis.

Legendre, P., Borcard, D., Peres-Neto, P., 2005. Analyzing Beta diversity: Partitioning the spatial variation of community composition data. *Ecol.Monogr.* 75, 435-450.

McArdle, B.H., Anderson, M.J., 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* 82, 290-297.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Wagner, H., 2018. vegan: Community Ecology Package.

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 Ecol. Evol.* 7, 549-555.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., 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. *Ecol.Lett.* 20, 561-576.

Smith, T.W., Lundholm, J.T., 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography* 33, 648-655.

Tikhonov, G., Opedal, Ø., Abrego, N., Lehikoinen, A., Ovaskainen, O., 2019. Joint species distribution modelling with HMSC-R. *bioRxiv* .

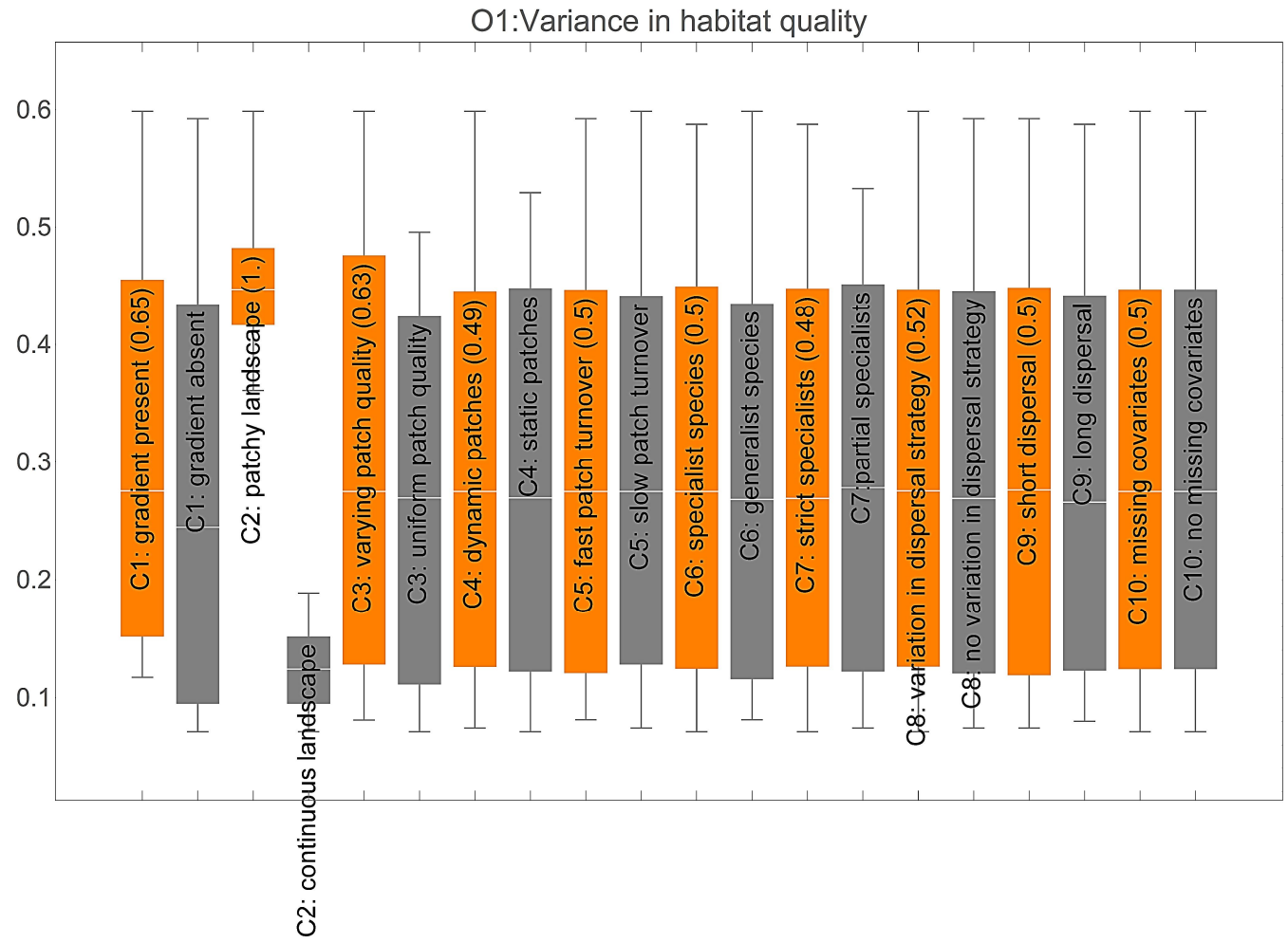
Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., Hui, F.K.C., 2015. So many variables: Joint modeling in community ecology. *Trends Ecol. Evol.* 30, 766-779.

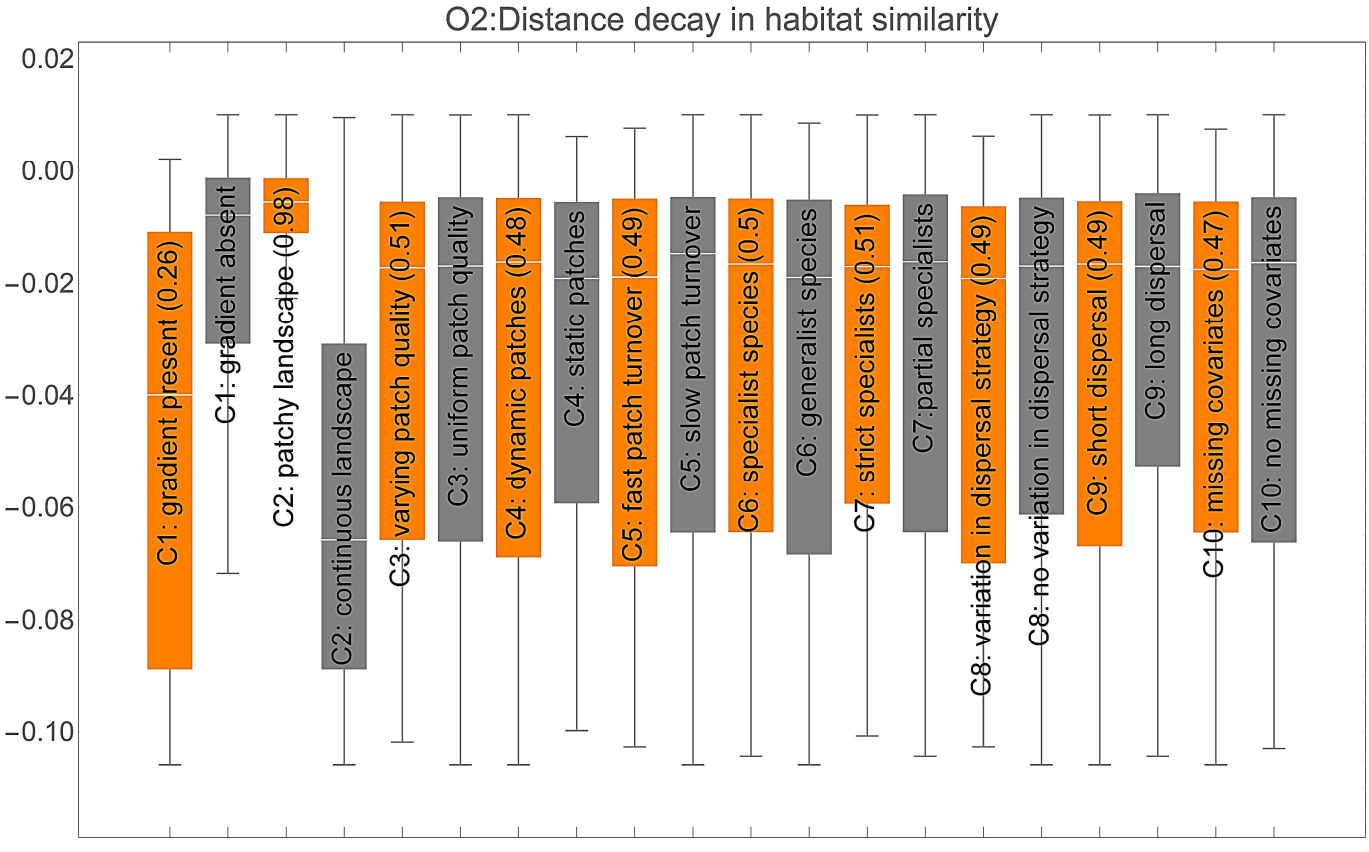
5 Supporting results

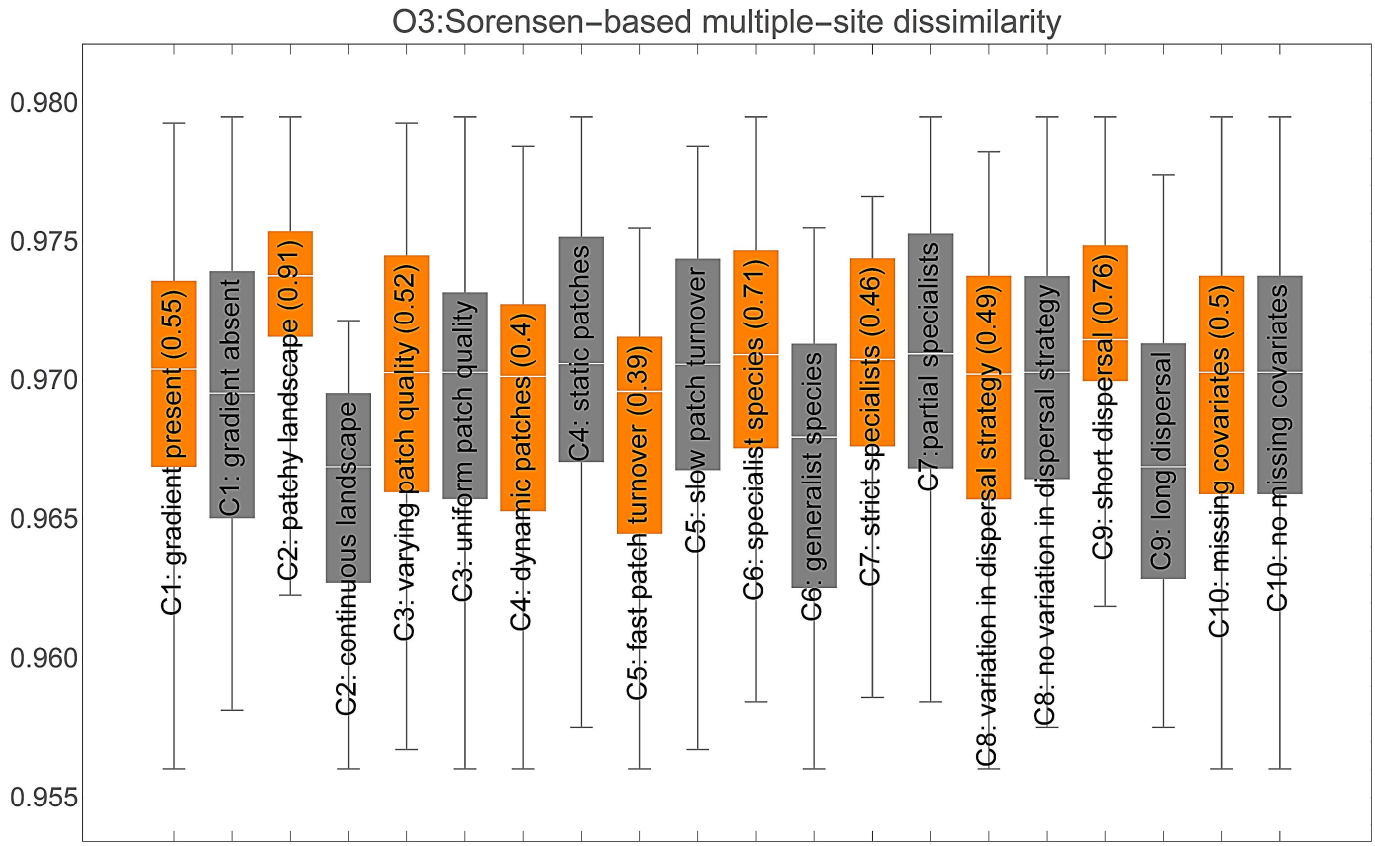
The raw results are provided numerically in the .csv file of Table S2 and graphically in Figures O1-O18. In Table S2, the columns A-J correspond to the Choices C1-C10 that determine the scenarios and the columns K-AB correspond to the output metrics O1-O18. For the scenarios, the value 0 corresponds to the baseline choice and the value 1 to the alternative choice, the baseline choice being shown inside brackets in the column name. For example, for the column "C1: gradient (no gradient)" the value 0 corresponds to the absence of the large-scale gradient, and the value 1 to the presence of the large-scale gradient.

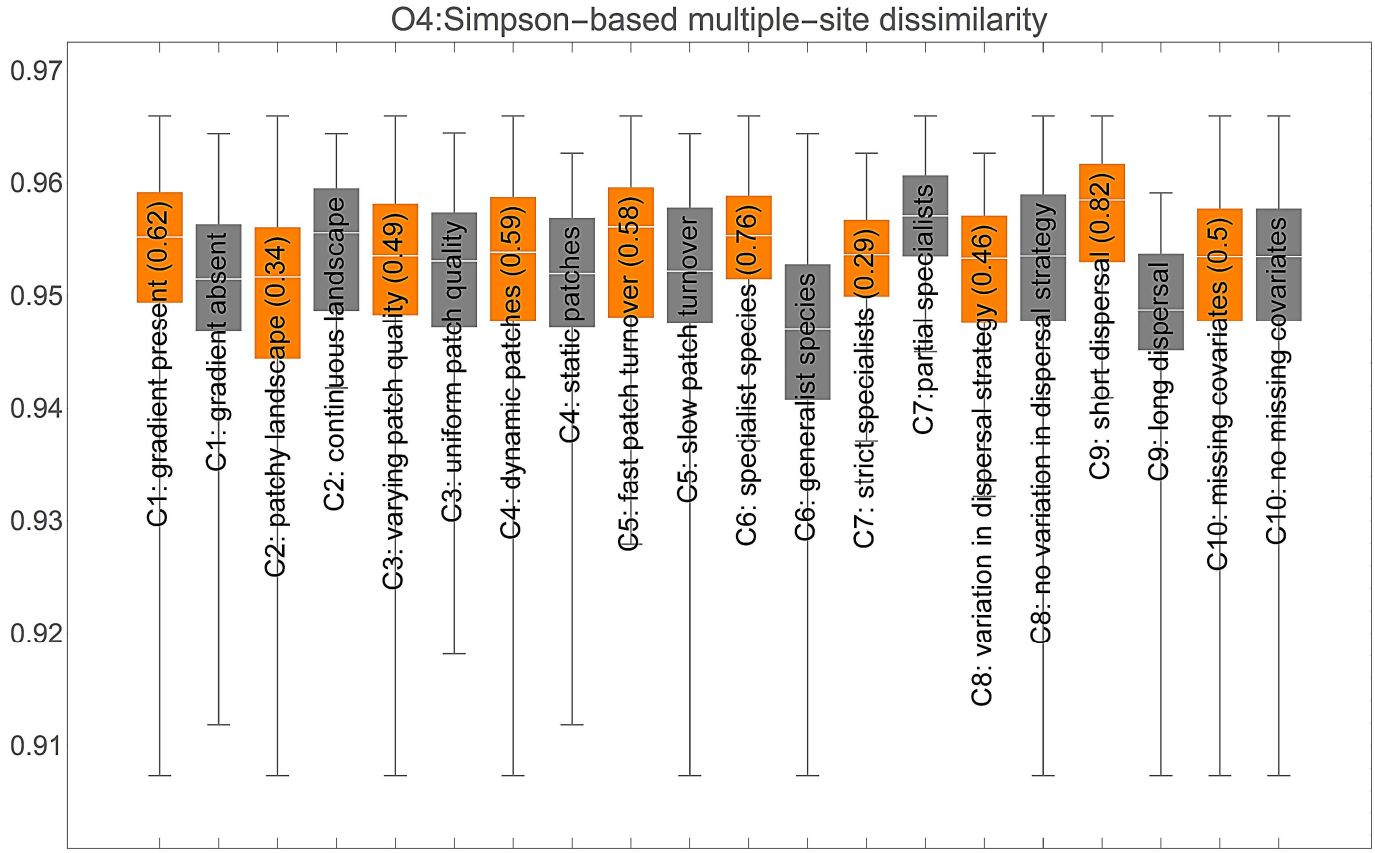
In the Figures O1-O18 below, we show the raw distributions of each of the output metrics O1-O18. For the definitions of the output metrics, see Table 1 in the main manuscript and the Section 4 above. In each figure, the distribution of the output metric is shown for scenarios split accordingly to each of the ten choices C1-C10, so that the range of values in scenarios based on one choice are shown by an orange boxplot, and the range of values in scenarios based on the other choice are shown by a grey boxplot.

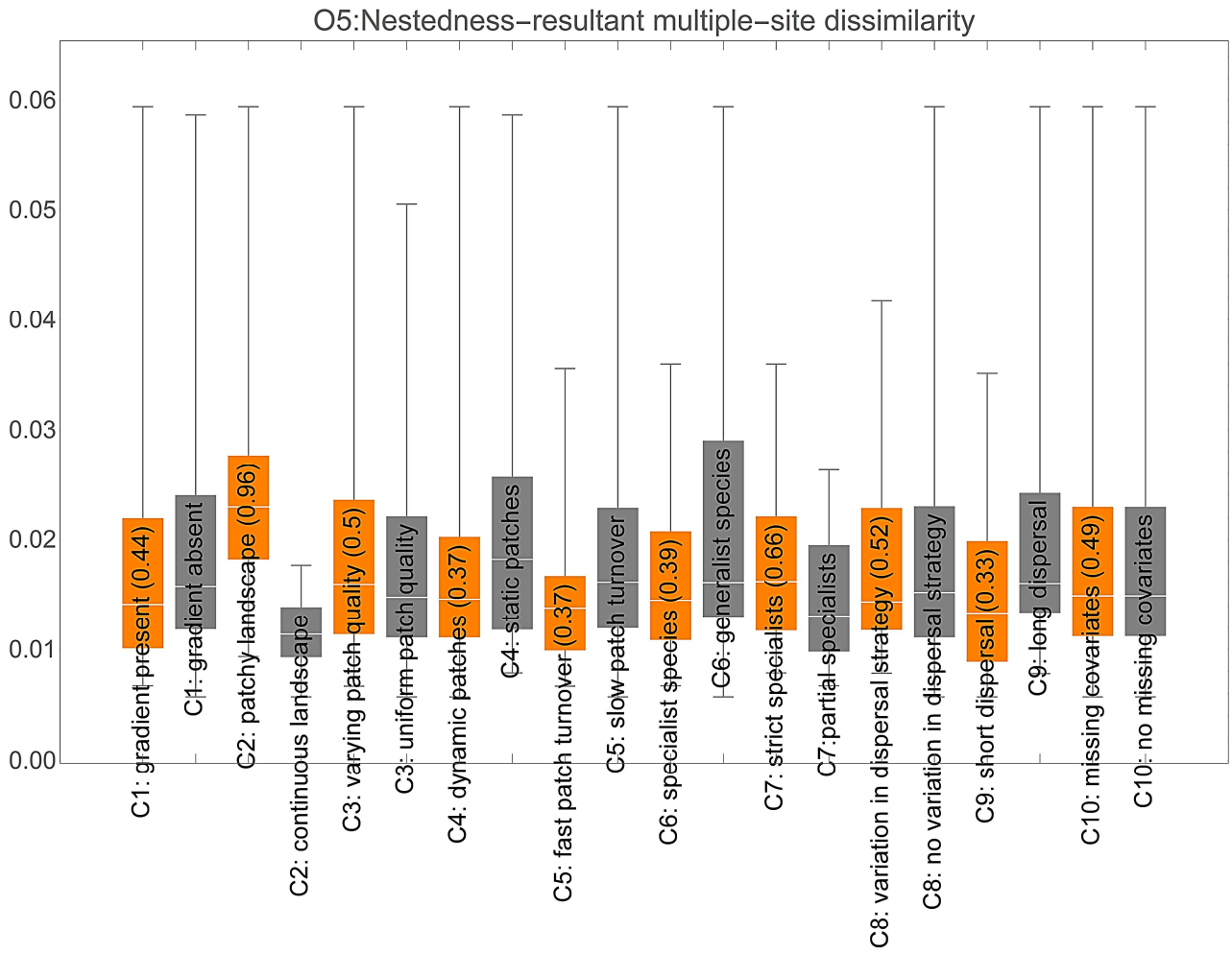
The numbers in brackets placed in labels of the orange box show the probability by which the output metric is greater for a scenario that is random selected among those that involve the choice of the orange box compared to a scenario that is random selected among those that involve the choice of the grey box. Thus, for example the output metric O1 was always greater (with probability 1) for scenarios simulated in patchy landscapes compared to scenarios simulated continuous landscapes. As another example, the output metric O4 was greater with probability 0.82 for scenarios where the species followed short-distance dispersal compared to those where the species followed long-distance dispersal.

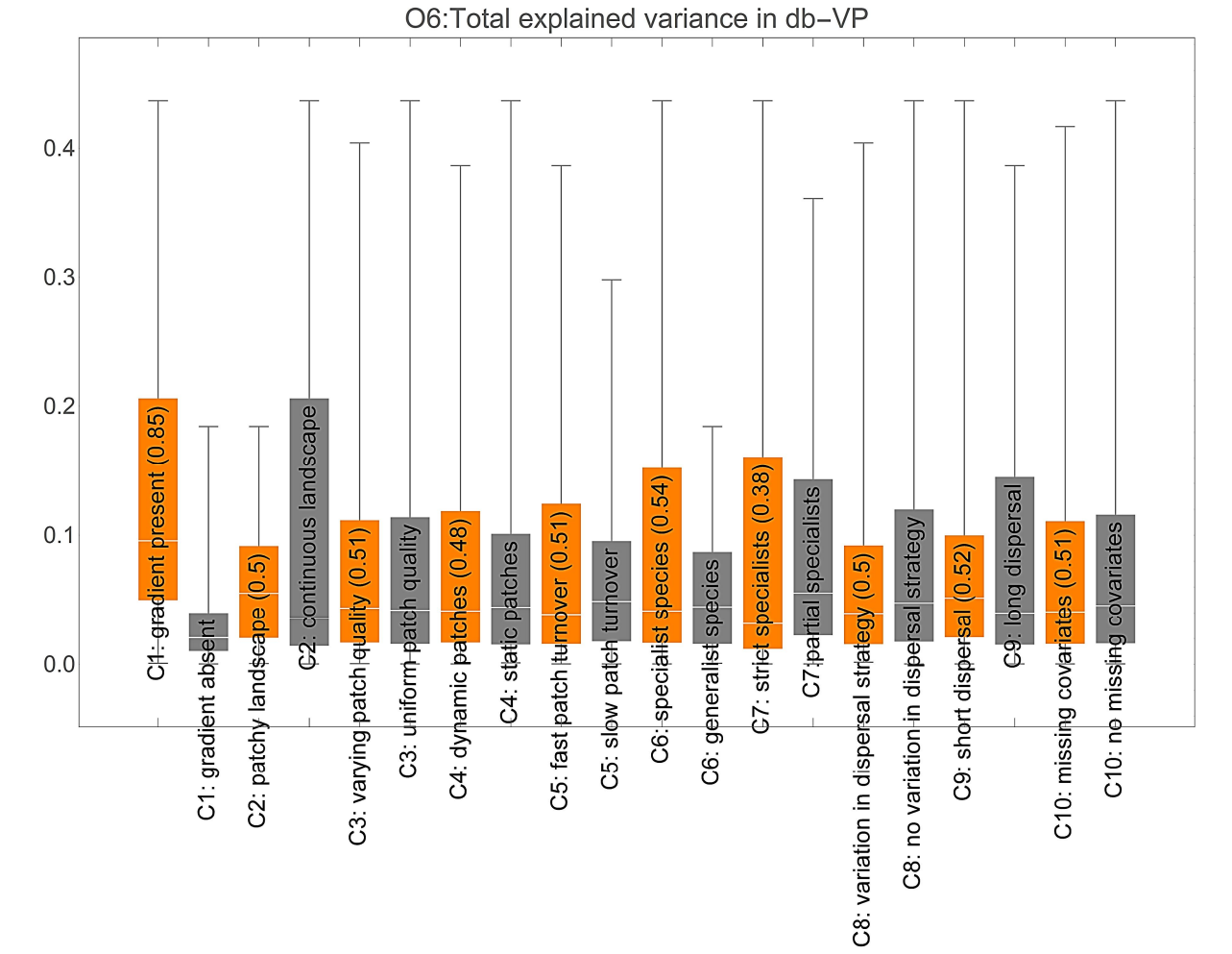

O1:Variance in habitat quality
0.6
0.5
0.4
0.3
0.2
0.1
C1: gradient present (0.65)
C1: gradient absent
C2: patchy landscape (1.)
C2: continuous landscape
C3: varying patch quality (0.63)
C3: uniform patch quality
C4: dynamic patches (0.49)
C4: static patches
C5: fast patch turnover (0.5)
C5: slow patch turnover
C6: specialist species (0.5)
C6: generalist species
C7: strict specialists (0.48)
C7:partial specialists
C8: variation in dispersal strategy (0.52)
C8: no variation in dispersal strategy
C9: short dispersal (0.5)
C9: long dispersal
C10: missing covariates (0.5)
C10: no missing covariates

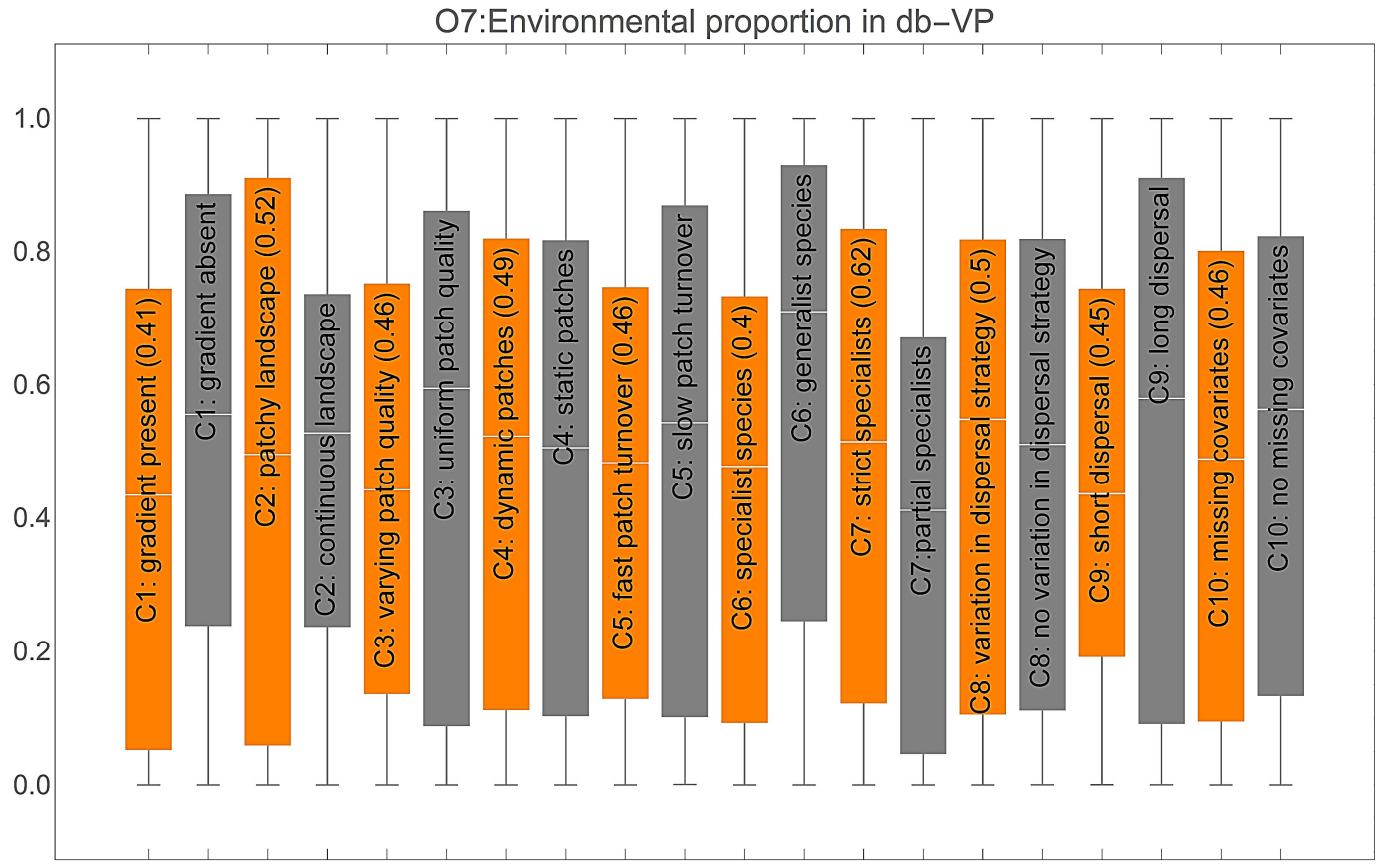

O2:Distance decay in habitat similarity
0.02
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
C1: gradient present (0.26)
C1: gradient absent
C2: patchy landscape (0.98)
C2: continuous landscape
C3: varying patch quality (0.51)
C3: uniform patch quality
C4: dynamic patches (0.48)
C4: static patches
C5: fast patch turnover (0.49)
C5: slow patch turnover
C6: specialist species (0.5)
C6: generalist species
C7: strict specialists (0.51)
C7:partial specialists
C8: variation in dispersal strategy (0.49)
C8: no variation in dispersal strategy
C9: short dispersal (0.49)
C9: long dispersal
C10: missing covariates (0.47)
C10: no missing covariates

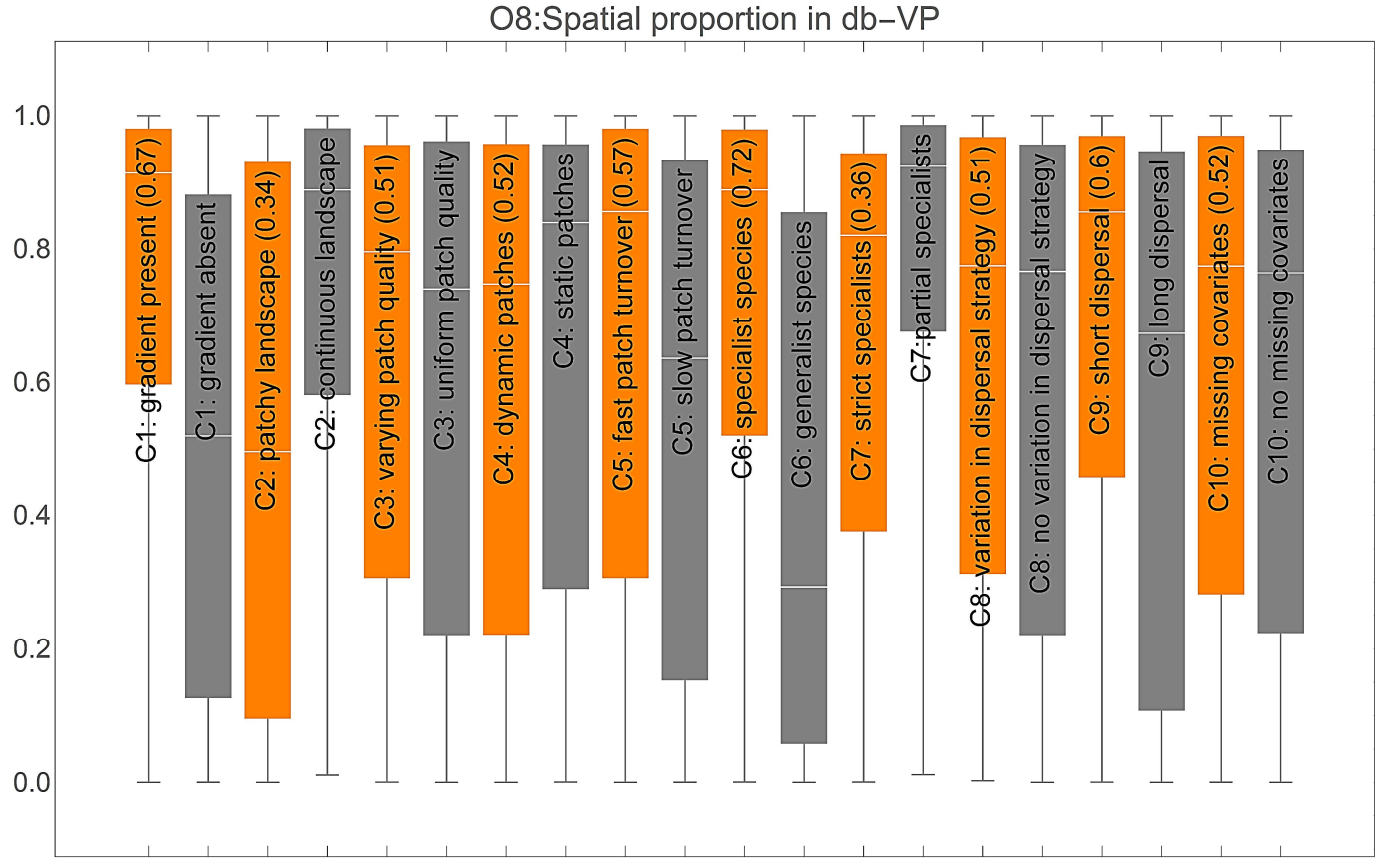

O3:Sorensen−based multiple−site dissimilarity
0.980
0.975
0.970
0.965
0.960
0.955
C1: gradient present (0.55)
C1: gradient absent
C2: patchy landscape (0.91)
C2: continuous landscape
C3: varying patch quality (0.52)
C3: uniform patch quality
C4: dynamic patches (0.4)
C4: static patches
C5: fast patch turnover (0.39)
C5: slow patch turnover
C6: specialist species (0.71)
C6: generalist species
C7: strict specialists (0.46)
C7: partial specialists
C8: variation in dispersal strategy (0.49)
C8: no variation in dispersal strategy
C9: short dispersal (0.76)
C9: long dispersal
C10: missing covariates (0.5)
C10: no missing covariates

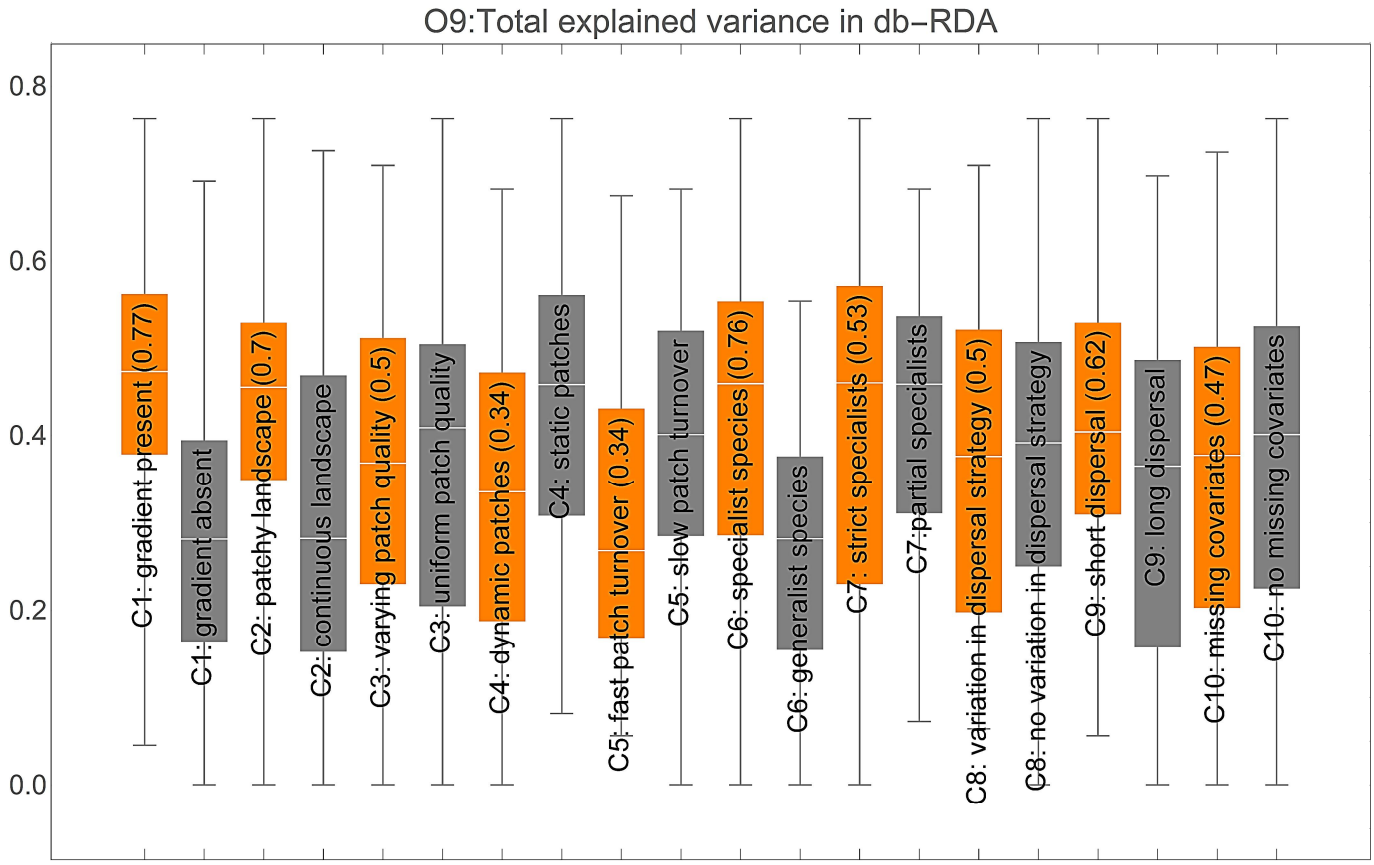

O4:Simpson−based multiple−site dissimilarity
0.97
0.96
0.95
0.94
0.93
0.92
0.91
C1: gradient present (0.62)
C1: gradient absent
C2: patchy landscape (0.34)
C2: continuous landscape
C3: varying patch quality (0.49)
C3: uniform patch quality
C4: dynamic patches (0.59)
C4: static patches
C5: fast patch turnover (0.58)
C5: slow patch turnover
C6: specialist species (0.76)
C6: generalist species
C7: strict specialists (0.29)
C7: partial specialists
C8: variation in dispersal strategy (0.46)
C8: no variation in dispersal strategy
C9: short dispersal (0.82)
C9: long dispersal
C10: missing covariates (0.5)
C10: no missing covariates

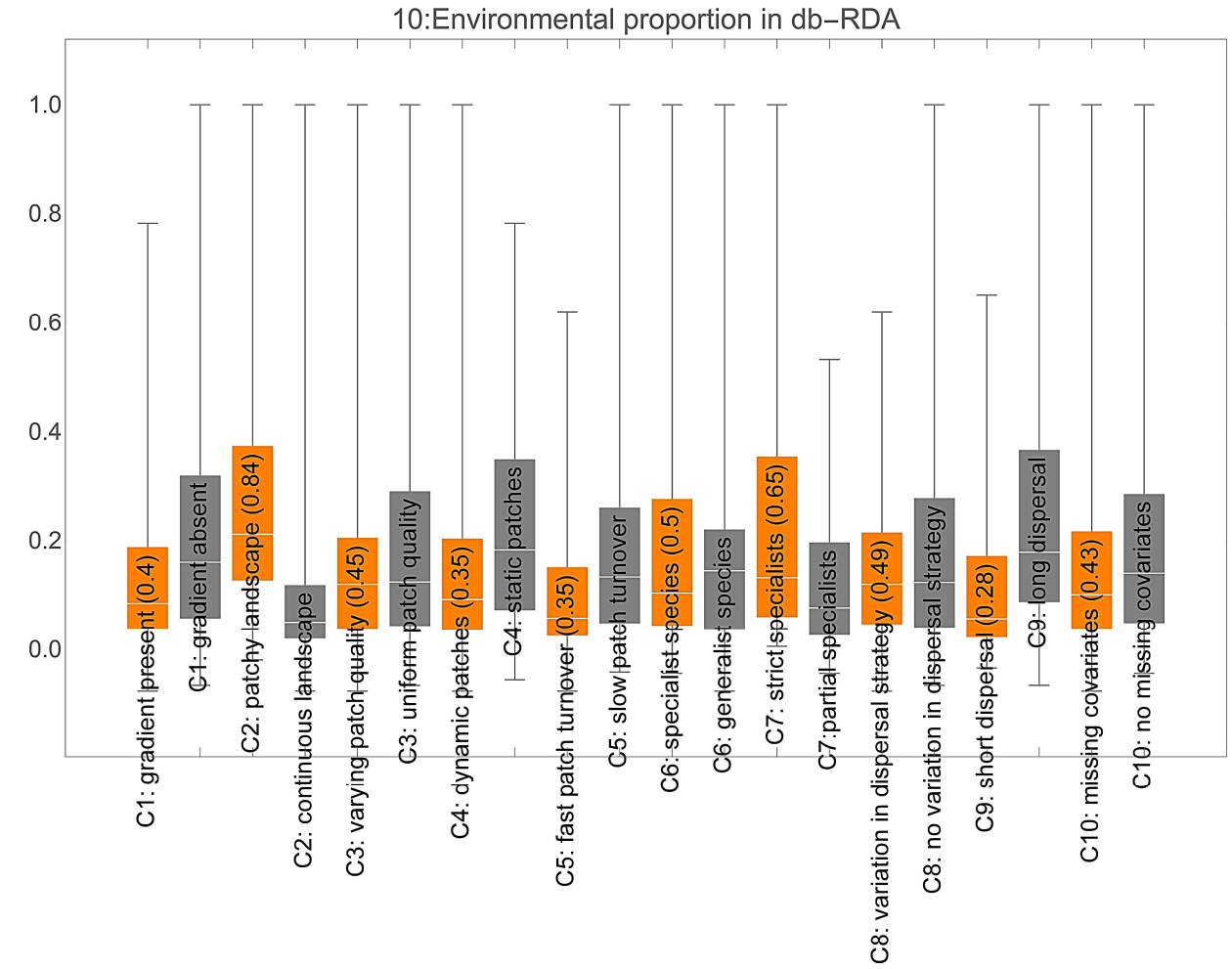

O5:Nestedness−resultant multiple−site dissimilarity
0.06
0.05
0.04
0.03
0.02
0.01
0.00
C1: gradient present (0.44)
C1: gradient absent
C2: patchy landscape (0.96)
C2: continuous landscape
C3: varying patch quality (0.5)
C3: uniform patch quality
C4: dynamic patches (0.37)
C4: static patches
C5: fast patch turnover (0.37)
C5: slow patch turnover
C6: specialist species (0.39)
C6: generalist species
C7: strict specialists (0.66)
C7: partial specialists
C8: variation in dispersal strategy (0.52)
C8: no variation in dispersal strategy
C9: short dispersal (0.33)
C9: long dispersal
C10: missing covariates (0.49)
C10: no missing covariates

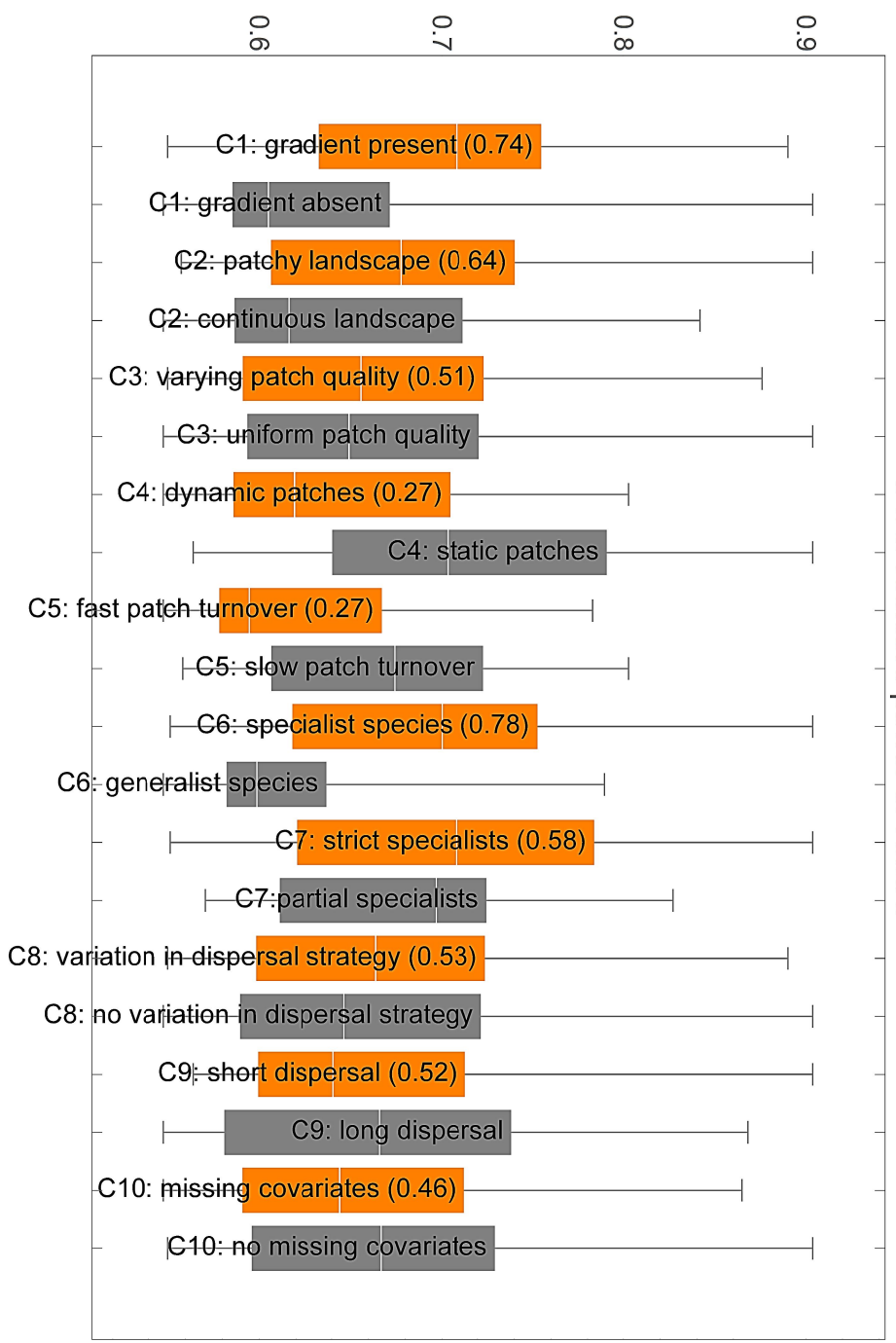

O6:Total explained variance in db−VP
0.4
0.3
0.2
0.1
0.0
C1: gradient present (0.85)
C1: gradient absent
C2: patchy landscape (0.5)
C2: continuous landscape
C3: varying patch quality (0.51)
C3: uniform patch quality
C4: dynamic patches (0.48)
C4: static patches
C5: fast patch turnover (0.51)
C5: slow patch turnover
C6: specialist species (0.54)
C6: generalist species
C7: strict specialists (0.38)
C7: partial specialists
C8: variation in dispersal strategy (0.5)
C8: no variation in dispersal strategy
C9: short dispersal (0.52)
C9: long dispersal
C10: missing covariates (0.51)
C10: no missing covariates

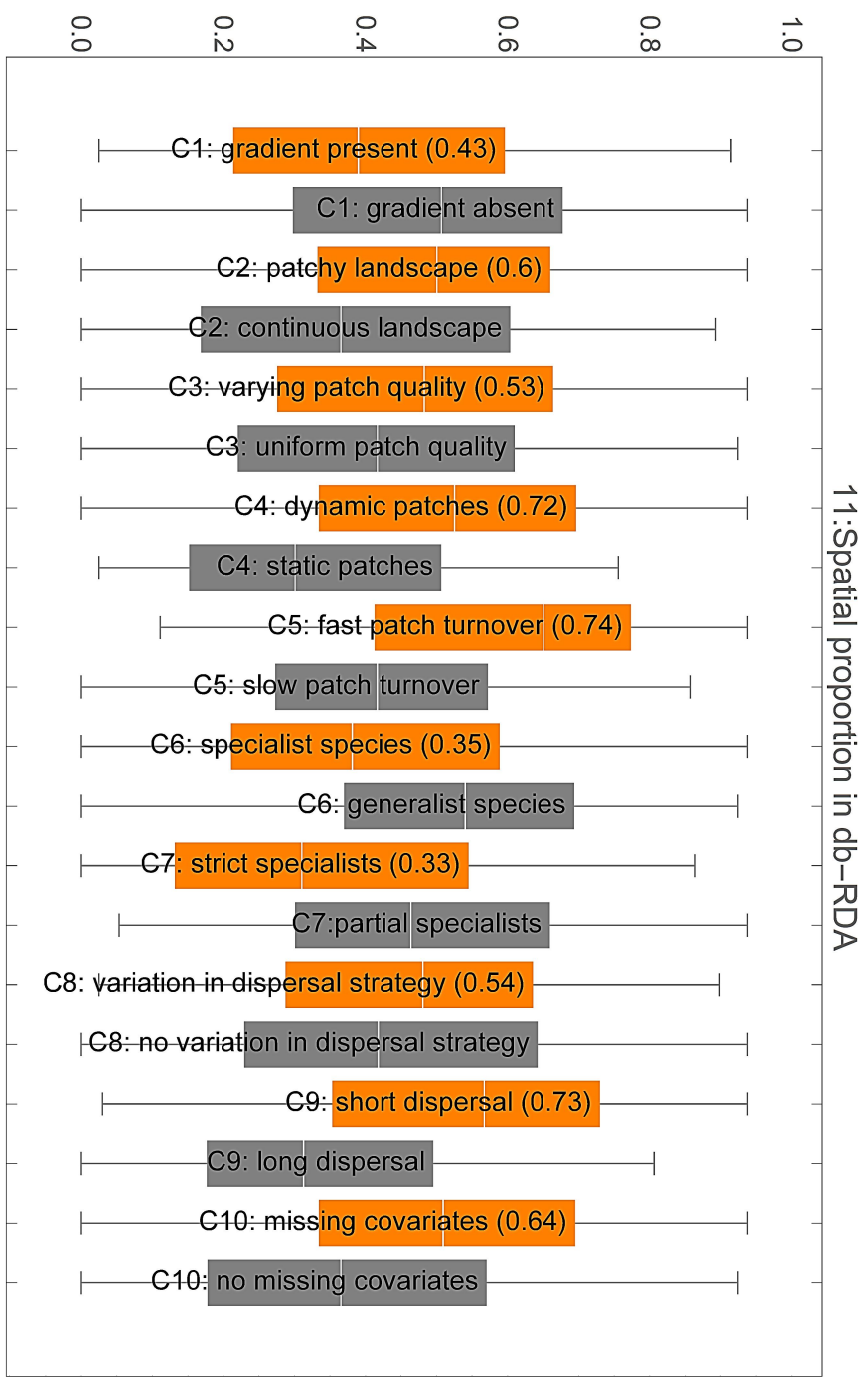

O7:Environmental proportion in db−VP
1.0
0.8
0.6
0.4
0.2
0.0
C1: gradient present (0.41)
C1: gradient absent
C2: patchy landscape (0.52)
C2: continuous landscape
C3: varying patch quality (0.46)
C3: uniform patch quality
C4: dynamic patches (0.49)
C4: static patches
C5: fast patch turnover (0.46)
C5: slow patch turnover
C6: specialist species (0.4)
C6: generalist species
C7: strict specialists (0.62)
C7:partial specialists
C8: variation in dispersal strategy (0.5)
C8: no variation in dispersal strategy
C9: short dispersal (0.45)
C9: long dispersal
C10: missing covariates (0.46)
C10: no missing covariates

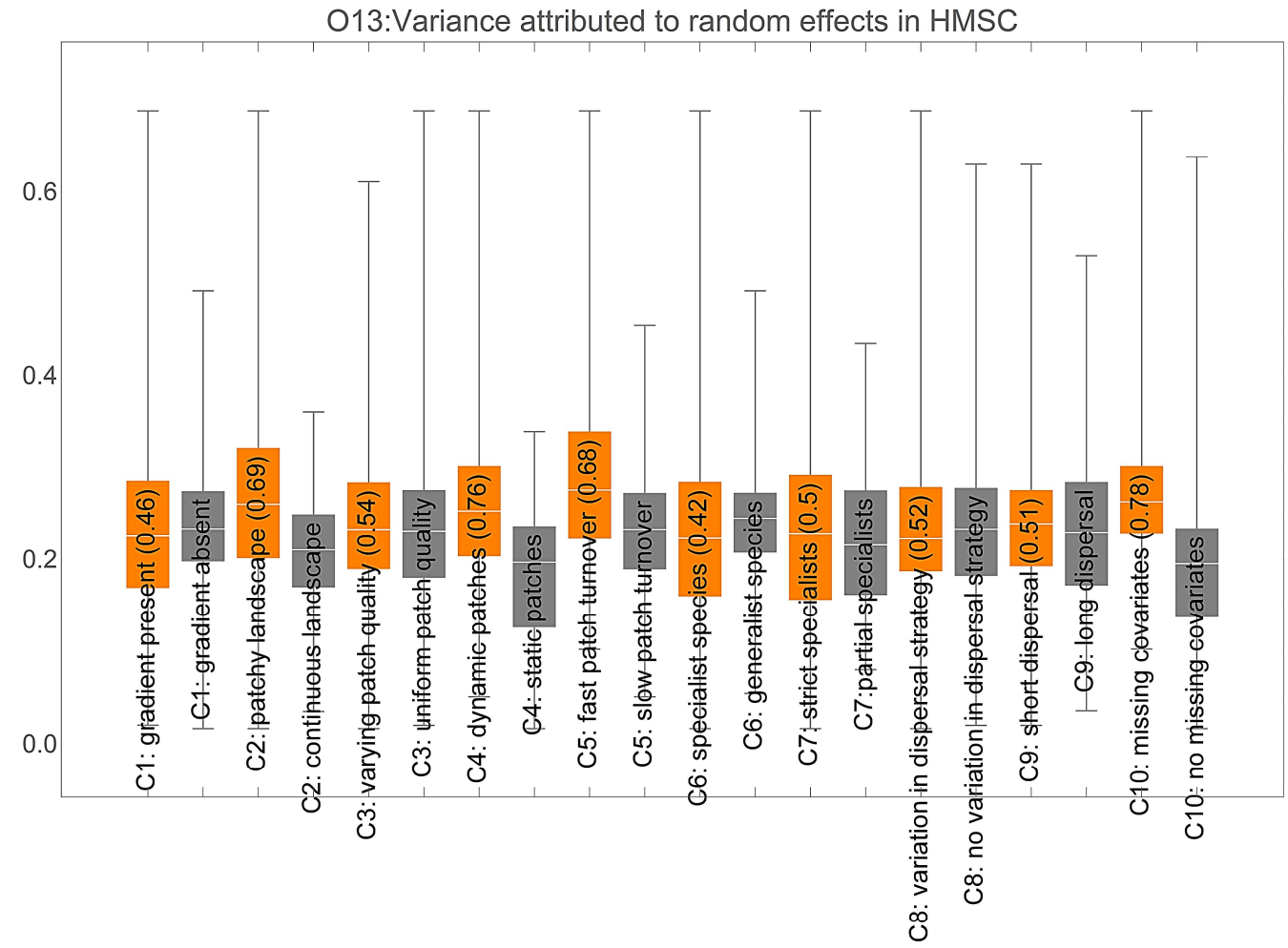

O8:Spatial proportion in db−VP
1.0
0.8
0.6
0.4
0.2
0.0
C1: gradient present (0.67)
C1: gradient absent
C2: patchy landscape (0.34)
C2: continuous landscape
C3: varying patch quality (0.51)
C3: uniform patch quality
C4: dynamic patches (0.52)
C4: static patches
C5: fast patch turnover (0.57)
C5: slow patch turnover
C6: specialist species (0.72)
C6: generalist species
C7: strict specialists (0.36)
C7:partial specialists
C8: variation in dispersal strategy (0.51)
C8: no variation in dispersal strategy
C9: short dispersal (0.6)
C9: long dispersal
C10: missing covariates (0.52)
C10: no missing covariates

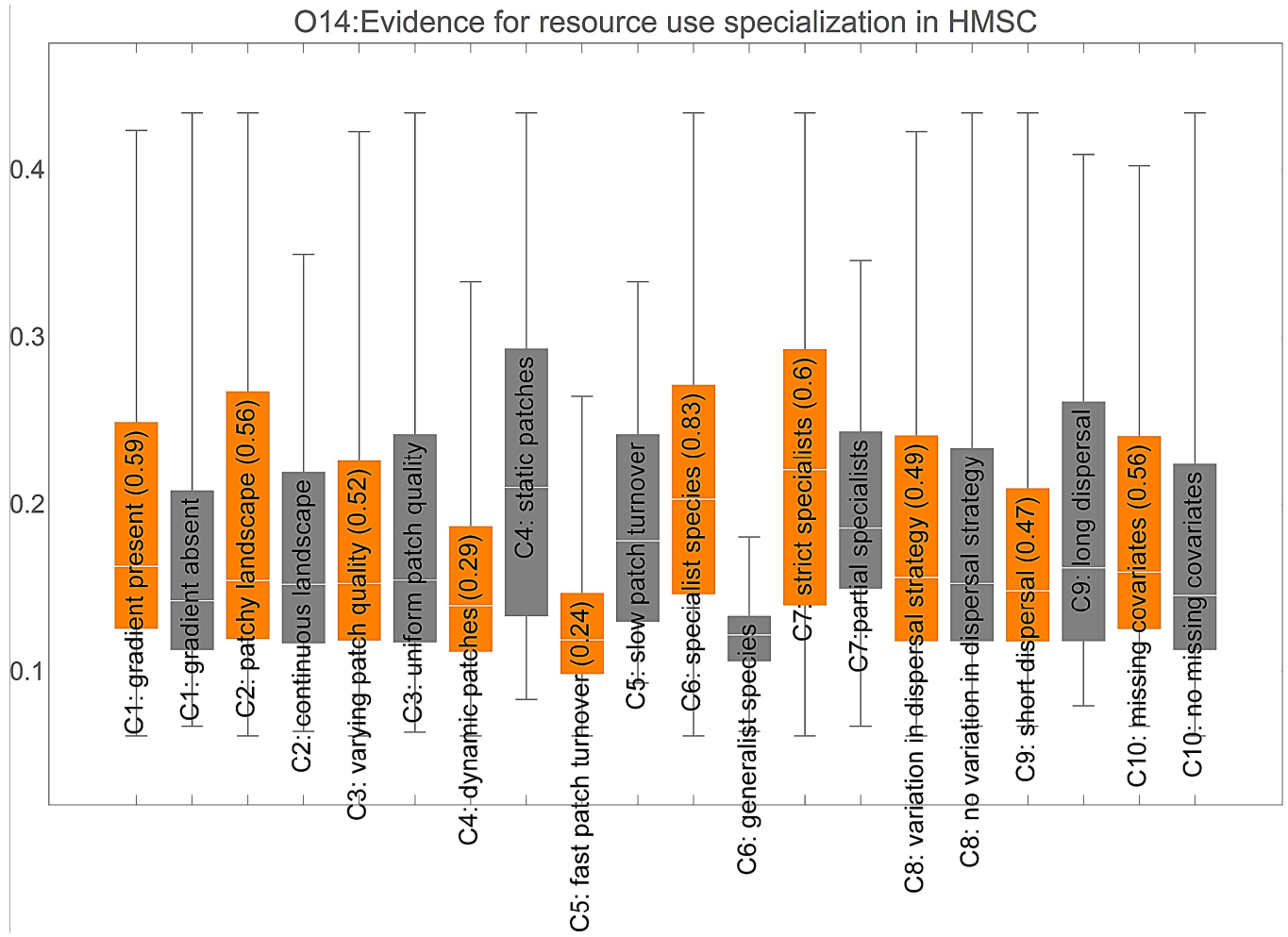

O9:Total explained variance in db−RDA
0.0
0.2
0.4
0.6
0.8
C1: gradient present (0.77)
C1: gradient absent
C2: patchy landscape (0.7)
C2: continuous landscape
C3: varying patch quality (0.5)
C3: uniform patch quality
C4: dynamic patches (0.34)
C4: static patches
C5: fast patch turnover (0.34)
C5: slow patch turnover
C6: specialist species (0.76)
C6: generalist species
C7: strict specialists (0.53)
C7:partial specialists
C8: variation in dispersal strategy (0.5)
C8: no variation in dispersal strategy
C9: short dispersal (0.62)
C9: long dispersal
C10: missing covariates (0.47)
C10: no missing covariates

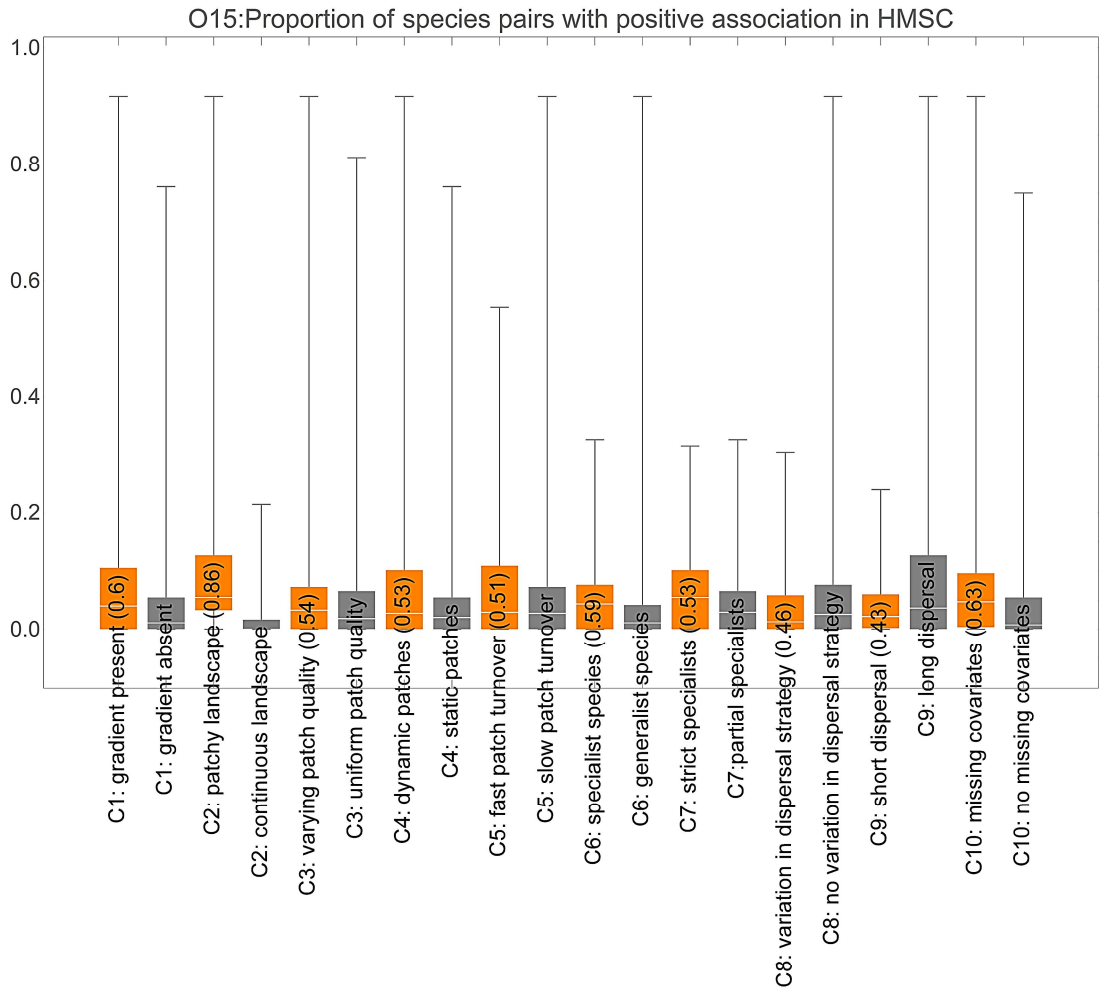

10:Environmental proportion in db−RDA
0.0
0.2
0.4
0.6
0.8
1.0
C1: gradient present (0.4)
C1: gradient absent
C2: patchy landscape (0.84)
C2: continuous landscape
C3: varying patch quality (0.45)
C3: uniform patch quality
C4: dynamic patches (0.35)
C4: static patches
C5: fast patch turnover (0.35)
C5: slow patch turnover
C6: specialist species (0.5)
C6: generalist species
C7: strict specialists (0.65)
C7:partial specialists
C8: variation in dispersal strategy (0.49)
C8: no variation in dispersal strategy
C9: short dispersal (0.28)
C9: long dispersal
C10: missing covariates (0.43)
C10: no missing covariates

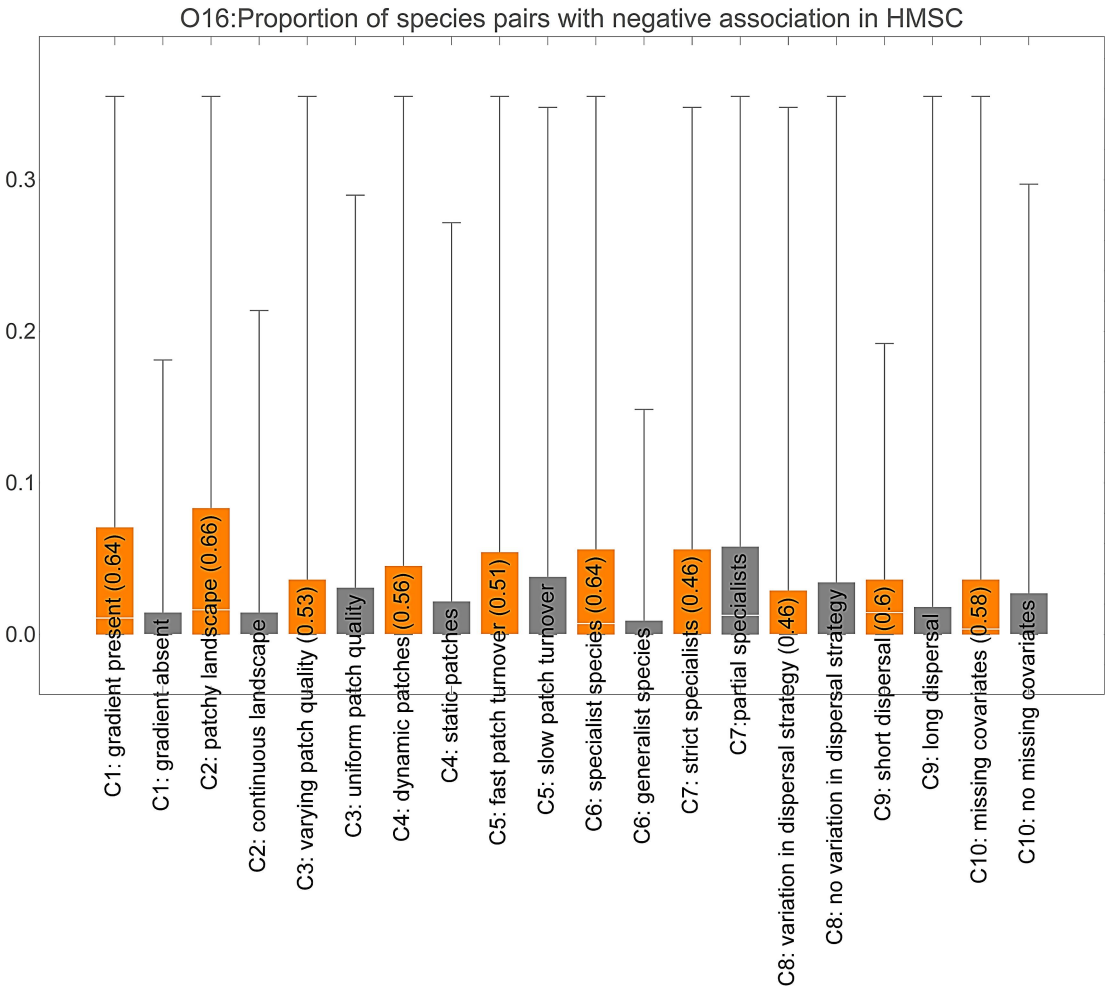

0.6
0.7
0.8
0.9
C1: gradient present (0.74)
C1: gradient absent
C2: patchy landscape (0.64)
C2: continuous landscape
C3: varying patch quality (0.51)
C3: uniform patch quality
C4: dynamic patches (0.27)
C4: static patches
C5: fast patch turnover (0.27)
C5: slow patch turnover
C6: specialist species (0.78)
C6: generalist species
C7: strict specialists (0.58)
C7:partial specialists
C8: variation in dispersal strategy (0.53)
C8: no variation in dispersal strategy
C9: short dispersal (0.52)
C9: long dispersal
C10: missing covariates (0.46)
C10: no missing covariates
O12:Predictive power of HMSC

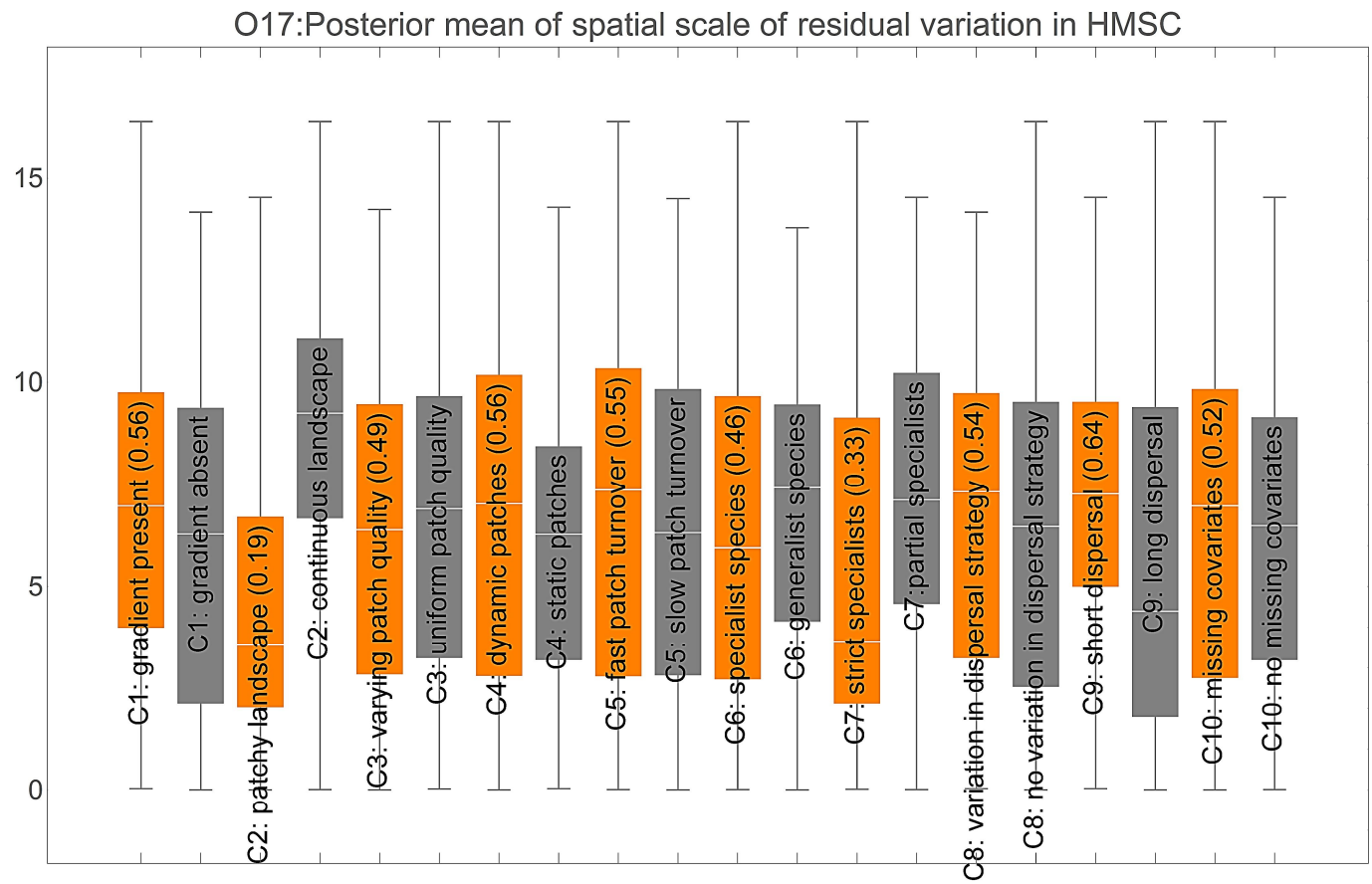

0.0
0.2
0.4
0.6
0.8
1.0
C1: gradient present (0.43)
C1: gradient absent
C2: patchy landscape (0.6)
C2: continuous landscape
C3: varying patch quality (0.53)
C3: uniform patch quality
C4: dynamic patches (0.72)
C4: static patches
C5: fast patch turnover (0.74)
C5: slow patch turnover
C6: specialist species (0.35)
C6: generalist species
C7: strict specialists (0.33)
C7:partial specialists
C8: variation in dispersal strategy (0.54)
C8: no variation in dispersal strategy
C9: short dispersal (0.73)
C9: long dispersal
C10: missing covariates (0.64)
C10: no missing covariates
11:Spatial proportion in db−RDA

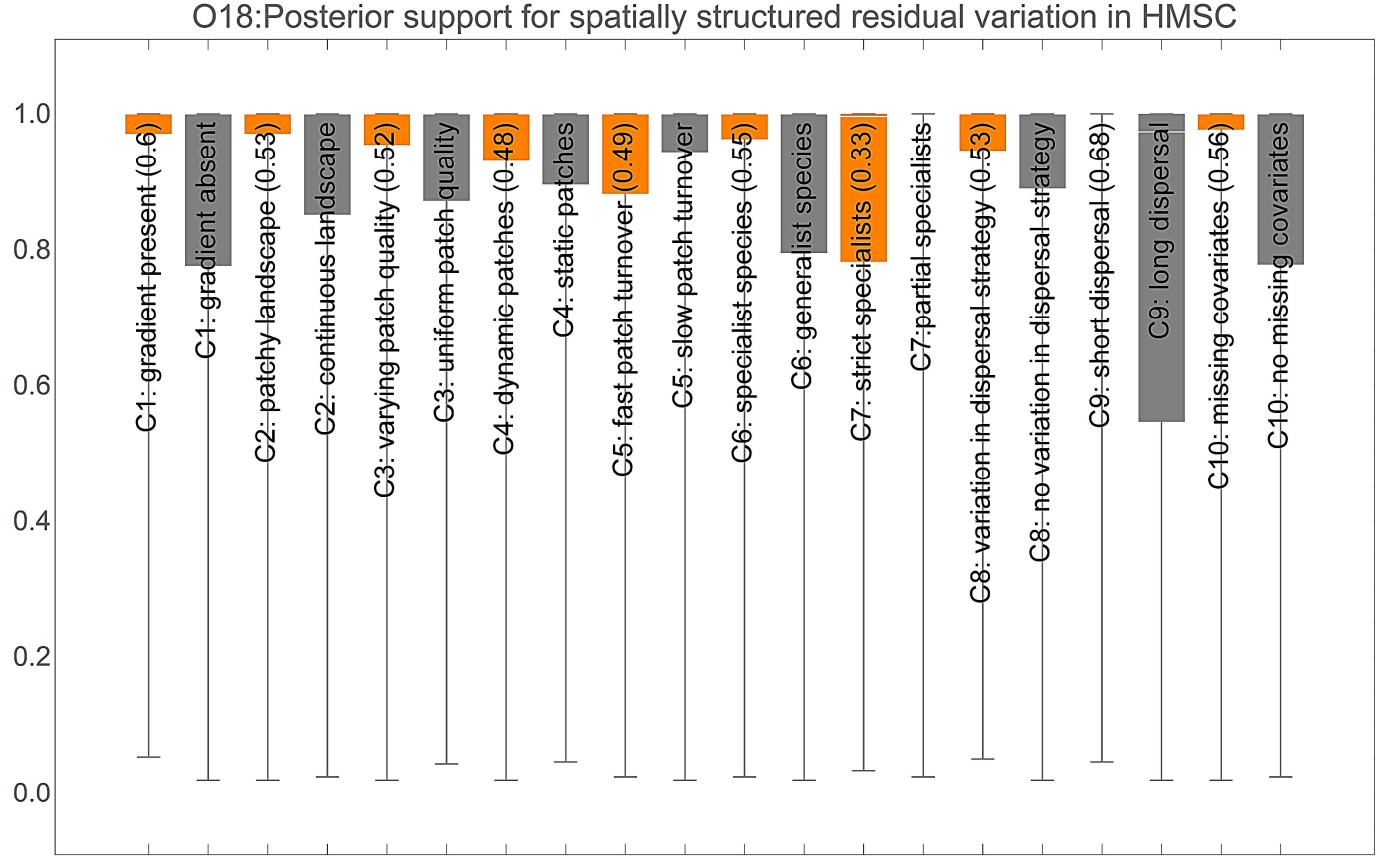

O13:Variance attributed to random effects in HMSC
0.0
0.2
0.4
0.6
C1: gradient present (0.46)
C1: gradient absent
C2: patchy landscape (0.69)
C2: continuous landscape
C3: varying patch quality (0.54)
C3: uniform patch quality
C4: dynamic patches (0.76)
C4: static patches
C5: fast patch turnover (0.68)
C5: slow patch turnover
C6: specialist species (0.42)
C6: generalist species
C7: strict specialists (0.5)
C7:partial specialists
C8: variation in dispersal strategy (0.52)
C8: no variation in dispersal strategy
C9: short dispersal (0.51)
C9: long dispersal
C10: missing covariates (0.78)
C10: no missing covariates


O14:Evidence for resource use specialization in HMSC
0.1
0.2
0.3
0.4
C1: gradient present (0.59)
C1: gradient absent
C2: patchy landscape (0.56)
C2: continuous landscape
C3: varying patch quality (0.52)
C3: uniform patch quality
C4: dynamic patches (0.29)
C4: static patches
C5: fast patch turnover (0.24)
C5: slow patch turnover
C6: specialist species (0.83)
C6: generalist species
C7: strict specialists (0.6)
C7:partial specialists
C8: variation in dispersal strategy (0.49)
C8: no variation in dispersal strategy
C9: short dispersal (0.47)
C9: long dispersal
C10: missing covariates (0.56)
C10: no missing covariates


O15:Proportion of species pairs with positive association in HMSC
1.0
0.8
0.6
0.4
0.2
0.0
C1: gradient present (0.6)
C1: gradient absent
C2: patchy landscape (0.86)
C2: continuous landscape
C3: varying patch quality (0.54)
C3: uniform patch quality
C4: dynamic patches (0.53)
C4: static patches
C5: fast patch turnover (0.51)
C5: slow patch turnover
C6: specialist species (0.59)
C6: generalist species
C7: strict specialists (0.53)
C7:partial specialists
C8: variation in dispersal strategy (0.46)
C8: no variation in dispersal strategy
C9: short dispersal (0.43)
C9: long dispersal
C10: missing covariates (0.63)
C10: no missing covariates


O16:Proportion of species pairs with negative association in HMSC
0.3
0.2
0.1
0.0
C1: gradient present (0.64)
C1: gradient absent
C2: patchy landscape (0.66)
C2: continuous landscape
C3: varying patch quality (0.53)
C3: uniform patch quality
C4: dynamic patches (0.56)
C4: static patches
C5: fast patch turnover (0.51)
C5: slow patch turnover
C6: specialist species (0.64)
C6: generalist species
C7: strict specialists (0.46)
C7:partial specialists
C8: variation in dispersal strategy (0.46)
C8: no variation in dispersal strategy
C9: short dispersal (0.6)
C9: long dispersal
C10: missing covariates (0.58)
C10: no missing covariates


O17:Posterior mean of spatial scale of residual variation in HMSC
15
10
5
0
C1: gradient present (0.56)
C1: gradient absent
C2: patchy landscape (0.19)
C2: continuous landscape
C3: varying patch quality (0.49)
C3: uniform patch quality
C4: dynamic patches (0.56)
C4: static patches
C5: fast patch turnover (0.55)
C5: slow patch turnover
C6: specialist species (0.46)
C6: generalist species
C7: strict specialists (0.33)
C7:partial specialists
C8: variation in dispersal strategy (0.54)
C8: no variation in dispersal strategy
C9: short dispersal (0.64)
C9: long dispersal
C10: missing covariates (0.52)
C10: no missing covariates


O18:Posterior support for spatially structured residual variation in HMSC
1.0
0.8
0.6
0.4
0.2
0.0
C1: gradient present (0.6)
C1: gradient absent
C2: patchy landscape (0.53)
C2: continuous landscape
C3: varying patch quality (0.52)
C3: uniform patch quality
C4: dynamic patches (0.48)
C4: static patches
C5: fast patch turnover (0.49)
C5: slow patch turnover
C6: specialist species (0.55)
C6: generalist species
C7: strict specialists (0.33)
C7:partial specialists
C8: variation in dispersal strategy (0.53)
C8: no variation in dispersal strategy
C9: short dispersal (0.68)
C9: long dispersal
C10: missing covariates (0.56)
C10: no missing covariates