Predictors of the long-term conservation value of sites in a mechanistic metacommunity model

Authors: Jacob D. O'Sullivan[1], J. Christopher D. Terry[1], Ramesh Wilson[1], Axel G. Rossberg[1]

Affiliations: [1]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, United Kingdom

Corresponding author: Jacob Dinner O'Sullivan (j.osullivan@qmul.ac.uk), School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom

Document statistics: Abstract - 225 words; Main Text - 6576 words, 11 figures; Appendix 1058 words, 2 figures.

Keywords: biodiversity — macroecology — spatial ecology — metacommunity — conservation

Competing interests: The authors declare no competing interests.

Data/code availability: Simulation data required to reproduce the results summarised in this study will be made publicly available upon publication. Previous versions of the software used to generate results are already available at `https://github.com/jacobosullivan/LVMCM_src/`. Study specific updates will be made publicly available upon publication.

Funding: This work forms part of the project 'Mechanisms and prediction of large-scale ecological responses to environmental change' funded by the Natural Environment Research Council (NE/T003510/1).

Abstract

Conserving biodiversity often requires assessment of which sites should be prioritised for protection. Sites are often selected based on area or connectivity, with the assumption that a site's long-term conservation value, as defined by the number of regional species extinctions its removal causes, is smallest for small, disconnected sites. In a simulation study of a mechanistic metacommunity model we find across the parameter range studied that site area is a good predictor of biomass loss following site removal but an insufficient predictor of the long-term species losses incurred as a result. We show that, out of five conceptually distinct predictors tested, including biodiversity, area and connectivity measures, the strongest predictor of long-term species loss (conservation value) is compositional distinctness (average between-site Bray-Curtis dissimilarity) of the impacted community. In extreme cases, small sites located in highly distinct habitats can lead to more species loss when removed than large sites located in more common habitats. Fitting our model to observation data on Andean diatoms and Brazilian lichen-fungi, we show that compositional distinctness exceeds area (total biomass) as a predictor of long-term species losses in the empirically relevant parameter range. Since conservation is primarily concerned with maintaining biodiversity, as opposed to undifferentiated biomass, our results robustly demonstrate that site area alone is not sufficient to gauge conservation priorities; comparative assessment of the community composition of sites is essential.

1 Introduction

Habitat loss due to conversion of natural landscapes for exploitation by human societies is the leading cause of biodiversity loss today (Millenium Ecosystem assessment, Reid et al. 2005; IPBES, Díaz et al. 2019). Immediate species losses, such as occur when the habitats of endemic species are destroyed, only represent part of the impact of land use conversion. At longer time scales, losses may occur as 'extinction debts' (Tilman et al., 1994) are paid. These additional extinctions can arise due to the crossing of critical population thresholds (Swift and Hannon, 2010), the disruption of source-sink dynamics (Mouquet et al., 2011) which can lead to losses among pioneer species (Nee and May, 1992; Livingston et al., 2012), or as a result of changes in the network of ecological interactions (Terborgh et al., 2001) as species distributions shift in response to climate and land-use change (Taheri et al., 2021). The complexity of the ecological response to habitat loss makes predicting long term impacts a major challenge since it requires understanding of how changes in the abiotic and biotic structure of the landscape are likely to interact (Chase et al., 2020).

Attempts to predict the impacts of habitat loss frequently rely on phenomenological models of species-area relationships (SAR) (Kuussaari et al., 2009; Figueiredo et al., 2019; Thompson et al., 2019), explicitly assuming that the long-term impacts can be predicted by simple scaling relationships (e.g. Thomas et al. 2004; Foster et al. 2013). Absent from this approach is an explicit consideration of minimum viable population sizes, range shifts and associated changes in biotic filtering or facilitation (Araújo and Luoto, 2007; Montoya and Raffaelli, 2010; Kissling et al., 2010); in short those processes that can produce extinction debts. Indeed, any approach to predicting the impact of habitat loss which ignores the varied effects of abiotic, biotic filtering and dispersal is likely to overlook important biological processes that may impact long-term outcomes (Chase et al., 2020).

The allocation of resources for the conservation of biodiversity necessarily involves assessment of the least worst outcomes of landscape conversion (Wilson et al., 2009). In order to make the best informed decisions on how to prioritise conservation efforts, a mechanistic understanding of the impact of habitat loss on biodiversity is required (Figueiredo et al., 2019). Here, we use a spatially explicit simulation model called the Lotka-Volterra Metacommunity Model (LVMCM), which has previously been found to reproduce a large variety of well-documented patterns in spatio-temporal macroecology (O'Sullivan et al., 2019, 2021), to explore the long-term effects of habitat destruction on biodiversity.

Central to this study is the question of which community properties, if any, predict the long-term impact of local habitat destruction at regional scales. We approach this question from two complimentary angles, comparing

and contrast the long-term impacts measured as changes in biomass and, alternatively, as regional species losses. Of both we ask whether initial biomass loss, reflecting the area of the removed sites, sufficiently predict long-term outcomes. Where this is not the case, we test a variety of local site properties describing the spatio-temporal compositional distinctness and spatial connectivity of the removed site to assess the *dominant* predictors of impact and discuss the mechanisms that these may reveal. We find that total or species-average biomass losses are reasonably well predicted by the biomass immediately removed from the metacommunity, but that predictive power is particularly poor in the case of rare species, which can undergo long-term relative changes in biomass that differ from the initial losses by multiple orders of magnitude. However, conservation biology is concerned not only with preserving the total amount of biological material in an ecosystem, but critically also with the number of species coexisting within that ecosystem. We find that long-term species losses are rather weakly associated with area alone. Instead regional extinctions are best predicted by the compositional distinctness of the removed community. We then show how metacommunity simulation models like the LVMCM can be fit to empirical species-by-site tables in a way that reproduces key aspects of the spatial structure of the real ecosystem. This model fitting procedure highlights the biological plausibility of the parameter space and the potential for mechanistic metacommunity models as a decision support tool for applied conservation (Chase et al., 2020).

2 Materials and methods

2.1 Model description - Overview

The LVMCM extends the Lotka-Volterra competition model into a spatially explicit system of connected sites. Sites differ in their suitability for each species and communities are connected by dispersal (Fig. 1). We allow metacommunity models to self-organise via a regional assembly process through which species invade the metacommunity and distribute across the landscape, responding to the abiotic and biotic conditions in the sites. We then systematically perturb assembled metacommunities by removing sites and measuring the long term impacts on biodiversity in terms of both biomass and regional species richness losses. For random metacommunities (those not fit to empirical survey data) we vary abiotic niche widths to explore the role of spatial scale on long term impacts. We also test how differences in characteristic dispersal length impact model outcomes. For fitted metacommunities, we ask which region of the model parameter space is most biologically relevant and use analogous removal experiments to rank local sites in order of long-term conservation value.

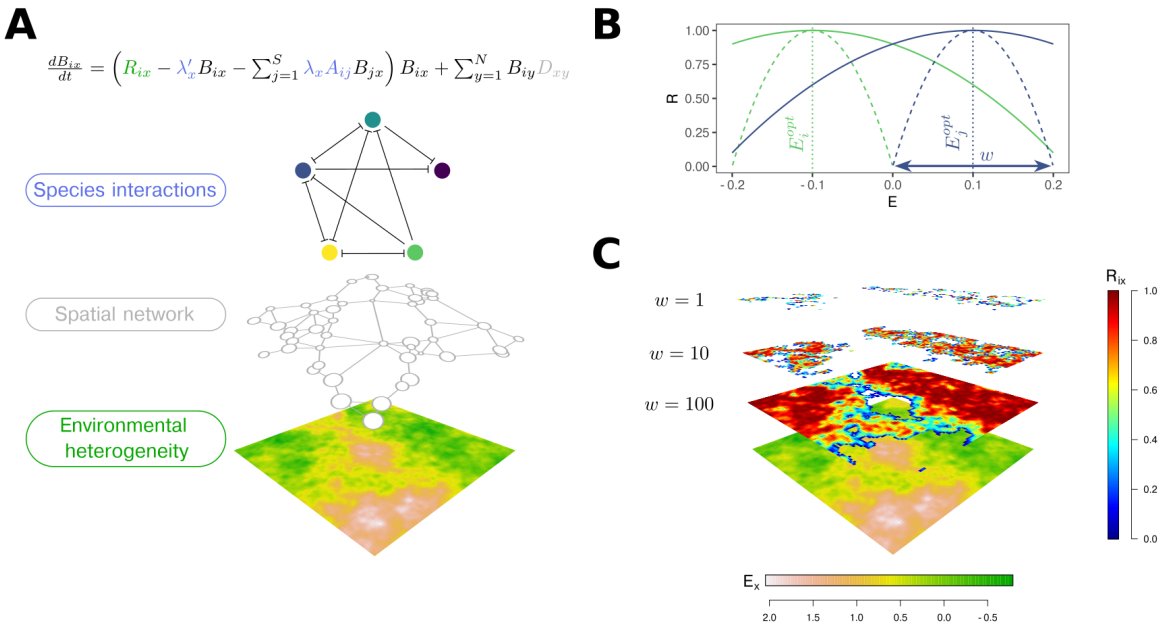

Figure 1: **Elements of the metacommunity model. A**: Model metacommunities include a spatially autocorrelated environment of at least one abiotic variable and a spatial network in which local sites of unequal area are modelled by scaling the local interaction coefficients by λ_x and λ'_x. **B**: Intrinsic growth rates $\mathbf{R}$, representing species' Hutchinsonian niche, are modelled as quadratic functions of the environmental variables. Niche width is controlled by a parameter w. Each species is assigned a randomly sampled environmental optimum E_i^{opt}. **C**: Three examples of $\mathbf{R}_{ix}$ for different values of the w, highlighting the relationship between niche width and effective heterogeneity.

2.2 Model description - Scaling local site area

In order to model local differences in site area (or quality) we developed the following method based on scaling the intensity of local ecological interactions to implicitly model local difference in carrying capacity. The precise functional form of the area dependence of intra- and inter-specific ecological interactions is not currently known. In order to fill this knowledge gap we assembled metacommunities of different total area and measured the decay in effective interaction coefficients. Previous work has shown that regional scale interaction coefficients C_{ij} describing the interaction between species i and j can be estimated for models by systematically perturbing the biomass of each species in turn and summing the impacts on all other species at the landscape scale (O'Sullivan et al., 2019). The resulting interaction matrix $\mathbf{C}$ captures the combined effects of differences in environmental preference, limited dispersal, indirect interactions and, of course, direct interactions modelled explicitly by the local interaction coefficients A_{ij}. Assembling metacommunities of various numbers of sites and computing the mean regional-scale interaction coefficients $\bar{\mathbf{C}}_{ij}$, reveals a decay in mean interaction strength with area (Fig 2a).

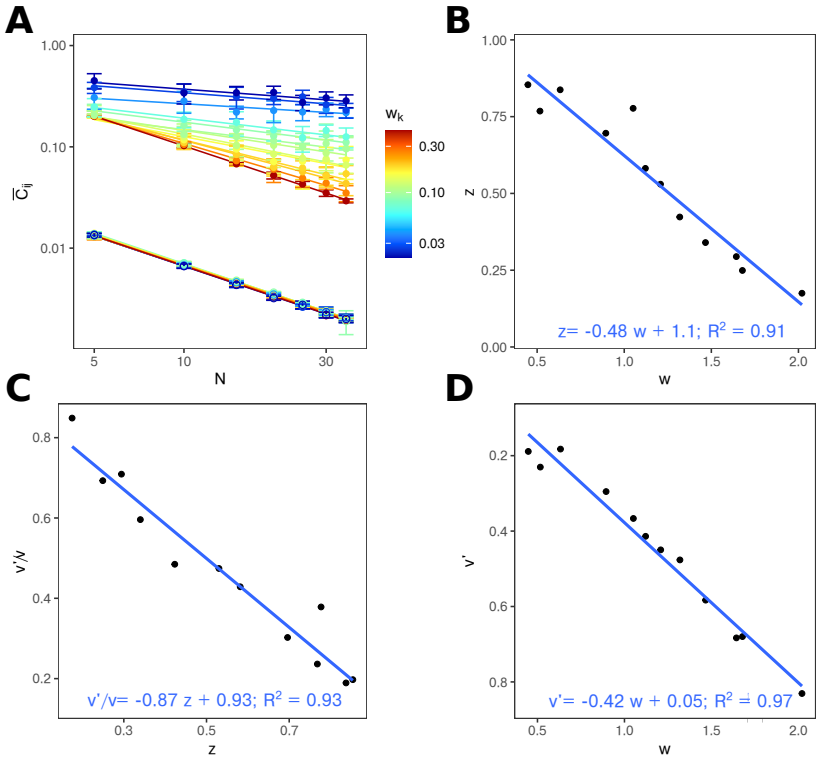

Figure 2: **Competition area relations in the LVMCM.** Assembling metacommunities along a gradient of total area reveals a power-law decay in average regional scale competitive interaction coefficients (A). Here we show the CAR for random metacommunities of $5 \geq N \leq 35$ sites for a range of niche widths (w). The area dependence of mean inter-specific interaction strengths are largely unimpacted by niche width. In contrast there is a substantial effect of w on the exponents v'. Linearly regressing z against w (B), v'/v against z (C) and v' against w (D) offers some indication of the mechanisms through which effective heterogeneity impacts the SAR, and the coefficients of these regressions can be input as constants into Eq. 4. Note z, v and v' were computed by assembling metacommunities of given w in the range $5 \geq N \leq 35$ sites, each in 10-fold replication such that each point in B-D represents an average exponent of 70 independent simulations.

This *competition area relation* (CAR) can be modelled by a power law $\bar{\mathbf{C}}_{ij} \propto a^v$ with exponents $v < 0$. The exponents v depend on model parameters, in particular those which determine species range sizes (in particular the abiotic niche width w) and differ for inter- and intra-specific interaction coefficients. We denote v and v' the exponents of the inter- and intra-specific CAR respectively:

$$\bar{\mathbf{C}}_{ij} = \bar{\mathbf{A}}_{ij} a^v \tag{1a}$$

$$\bar{\mathbf{C}}_{ii} = \bar{\mathbf{A}}_{ii} a^{v'}. \tag{1b}$$

Future developments in the analytic theory of LV metacommunities is likely to reveal key relationships between the CAR, the SAR ($S \propto a^z$) and the effective heterogeneity of the landscape (determined here by w and ℓ). Previous work on neutral models has shown a dependence of z on demographic rates (O'Dwyer and Green, 2010). In lieu of analytic results for non-neutral models, we report the statistical relationships between these key quantities (Fig. 2b-d).

The exponent z decays with abiotic niche width w suggesting greater abiotic heterogeneity (smaller niche width) results a steeper SAR. This is consistent with previous results (O'Sullivan et al., 2019) and qualitatively matches the observation of a steep phase in the empirical SAR at the highest spatial scales (greatest effective heterogeneity) (Allen and White, 2003; Rosindell and Cornell, 2007; Storch, 2016). We also find that the ratio of the exponents of the CAR $v'/v \geq 1$ holds over the examined parameter space and further that $v'/v \approx 1 - z$. This suggests a pivotal role for the ratio between the area dependence of inter- and intra-specific interaction coefficients in determining the accumulation of species with area, and thereby the exponents of the SAR. This should be seen in the context of long-standing theory on coexistence conditions in simple ecological models, which predict a stabilising effect of (relatively) strong self-regulation in competitive and food-web models (May, 1973; De Angelis, 1975) Finally, we find that the rate of decay of the intra-specific interaction coefficient, determined by v', and abiotic niche width w are negatively correlated. When abiotic niche widths are very small, intra-specific interaction coefficients are largely insensitive to total area (v' approaches 0), since species ranges are primarily limited by the environment. In contrast, when niche widths are large and species distributions can encompass much of the landscape, effective intra-specific interactions decay more rapidly with total area (v' approaches v, Fig. 2A). We note that the linear associations revealed in Figs. 2B-D may hold only for a subset of the parameter space of the model, however this subset includes the parameter combinations that best fit the empirical data (see below).

The CAR can be incorporated into the model dynamics so as to control the local biomass (area) of each site by scaling the local interaction matrix $\mathbf{A}_x$ for each site independently.We assume that each site implicitly represents a sub-network of various nodes and scale the interactions to those expected for the corresponding explicitly modelled metacommunity. Formally, if $\mathbf{A}_0$ is a hollow matrix with zeros on the diagonal, the scaled matrix is given by

$$\mathbf{A}_x = \lambda'_x \mathbf{I} + \lambda_x \mathbf{A}_0. \tag{2}$$

where $\mathbf{I}$ is the identity matrix, and the inter- and intra-specific scaling vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ have elements

$$\lambda_x = \left(\frac{a_x}{a_0}\right)^v \tag{3a}$$

$$\lambda'_x = \left(\frac{a_x}{a_0}\right)^{v'}. \tag{3b}$$

Here a_x is the area of the xth site, while a_0 is the area of a single isolated community with a given interaction matrix $\mathbf{A}$ in the absence of scaling. For simplicity we assume $a_0 = 1$.

We then assume that accumulation of diversity with area within sites is linear, that is, that $z = 1$. This is a simplification, however, it is also biologically reasonable, given the tendency for the initial phase of empirical SARs to be linear. By manipulation of the coefficients of the linear regressions shown in Fig. 2, we compute the exponents of the CAR in this linear phase as $v = -0.59$, $v' = -0.04$. For random metacommunities, a_x were chosen such that total biomass varied within each metacommunity in equally spaced intervals up to 30 biomass units. For fitted metacommunities, a_x were extracted from the empirical species by site tables (see below).

2.3 Model description - Environmental heterogeneity and spatial structure

The abiotic environment was modelled with a set of n vectors, with elements E_{kx}, representing the spatial distribution of a given environmental variable or principal component of environmental variation. E_{kx} were sampled from Gaussian random fields (Adler, 2010) with mean 0, standard deviation 1 (Fig. 1) and as such the abiotic component of the model was, by construction, normally distributed and spatially autocorrelated.

The spatial structure of the landscape was modelled by randomly sampling Cartesian coordinates from a uniform distribution in the range $[0; \sqrt{N}$, where N is the number of sites. We linked sites using a Gabriel graph (Gabriel and Sokal, 1969), which restricts dispersal to the local neighbourhood scale while limiting the average node degree for large metacommunities to 4 (Fig. 1).

2.4 Model description - Metacommunity dynamics

Metacommunity dynamics were modelled using a spatial extension to the classic Lotka-Volterra community model (O'Sullivan et al., 2019, 2021)

$$\frac{dB_{ix}}{dt} = \left(R_{ix} - \lambda'_x B_{ix} - \sum_{j=1}^{S} \lambda_x A_{ij} B_{jx} \right) B_{ix} + \sum_{y=1}^{N} B_{iy} D_{xy}, \tag{4}$$

where B_{ix} represents the biomass of the ith species in the xth site. R_{ix} are the intrinsic growth rates, which vary across the landscape. Each species in the model was allocated environmental optima E_{ik}^{opt} for each environmental variable, sampled from uniform distributions in the range $[1.25 \cdot \min(\mathbf{E}_k); 1.25 \cdot \max(\mathbf{E}_k)]$. The effective heterogeneity of the environment was controlled by varying the width of the quadratic environmental response function, which determines species abiotic niche width. The intrinsic growth rate of the ith species in site x was computed as:

$$R_{ix} = 1 - \sum_{k=1}^{n} \left(\frac{2}{w_k} \right)^2 (E_{kx} - E_{ik}^{opt})^2, \tag{5}$$

where n defines the relevant number of independent components of environmental variation. The abiotic niche width was kept fixed for all species and all environmental variables in a given model ($w_k = w$ for all k).

For simplicity the unscaled local interaction coefficients A_{ij} were sampled from a discrete distribution with $P(A_{ij} = 0.3) = 0.3$, and $A_{ij} = 0$ otherwise. Previous work has shown that for species rich model metacommunities the mean, variance and correlation of the interaction coefficients (here we use uncorrelated interaction matrices) are sufficient to predict macroscopic structures (Barbier et al., 2018; O'Sullivan et al., 2021). Within these bounds, the precise nature of the distribution from which these coefficients are sampled does not, therefore, impact model outcomes.

D_{xy} are the elements of the spatial connectivity matrix describing the rate of emigration away from, and immigration into sites. Emigration and immigration rates are given by $D_{xx} = -e$ and $D_{xy} = \sum_y \delta_{xy} e k_y^{-1} \exp(-d_{xy}/\ell)$ respectively, with e an emigration rate, d_{xy} the distance between sites x, y and ℓ the characteristic dispersal length which was systematically varied. The direct delta function $\delta_{xy} = 1$ in the case x, y are joined by a edge, and 0 otherwise. The normalisation k_y represents the unweighted degree of the yth site. For sites x, y that do not share an edge in the spatial network, $D_{xy} = 0$.

2.5 Model description - Metacommunity assembly and removal experiments

Model metacommunities were assembled by iteratively generating species i with randomly sampled E_{ik}^{opt}, A_{ij}. Invaders were introduced to the model at low biomass and dynamics simulated using Eq. 4. Before each invasion, the metacommunity was scanned for any species whose biomass had fallen below the extinction threshold (10^{-4} biomass units) in all sites. These species were considered regionally extinct and removed from the model. Metacommunity models assembled in this fashion eventually saturated with respect to both average local and regional species richness (O'Sullivan et al., 2019) due to the onset of ecological structural instability (Rossberg, 2013). After this condition occurs each invasion generates, on average, a single extinction.

After assembling model metacommunities of 20 sites to regional diversity limits, each site in turn, and all associated edges, were systematically removed and the simulation advanced 10^4 unit times. By construction (Eq. 5) the time required for a population biomass to recover following local extinction is of $\mathcal{O}(1)\ t$. Hereafter we refer to the unit times, t, as 'biomass recovery time'. For further discussion on the empirical equivalent of the a single biomass recovery time see O'Sullivan et al. 2021. The impact of site removal was assessed as the change in biomass at the metacommunity scale, at the at the metapopulation scale, or as the percent loss in regional (γ) diversity (Table 1).

Response	Notation	Calculation
Change in regional species (metapopulation) biomass	ΔB_i	$\sum_x (B_{ix}^{\text{after}} - B_{ix}^{\text{before}})$
Change in metacommunity species biomass	ΔB	$\sum_i \sum_x (B_{ix}^{\text{after}} - B_{ix}^{\text{before}})$
Percent loss of regional diversity	$\Delta\gamma$	$100 \cdot (\gamma^{\text{before}} - \gamma^{\text{after}})/\gamma^{\text{before}}$

Table 1: Alternative measures of the impact of site removal.

2.6 Predictors of long-term impact

We tested five empirically measurable properties as possible predictors of long-term impacts in the case that biomass alone was demonstrably insufficient: total site biomass, which we consider a proxy for site area, community 'distinctness' calculated as the average between site Bray-Curtis (BC) (Bray and Curtis, 1957) dissimilarity, local turnover rate calculated as the average BC dissimilarity between time points in a time-series of length 10^4 biomass recovery times, and two spatial network properties, the exponential-distance weighted degree, and the betweenness centrality of the distance weighted spatial network of the focal site (Table 2). Turnover the the model occurs not due to environmental fluctuation, or due to perturbation following regional invasion. Instead, metacommunity models can exhibit autonomous population dynamics due to a combination of mass effects and ecological redund-

ancy (more that one community composition possible at a given point in environment space, O'Sullivan et al. 2021). Betweenness centrality is a mesoscale spatial network property describing the connectedness of the subgraph in which a focal site is embedded and has previously hypothesised to be predictive of sites that contribute disproportionately to regional diversity (Mouquet et al., 2013). The inverse weighted degree is a representation of the connectedness of a given site to its immediate neighbours and thereby of the influence that site has on them via mass effects.

We modelled the impact of the various predictor variables using non-parametric random forest regression which can be used to assess the importance of various (non-linear) 'features' in driving difference in a response variable. Alternative predictors, correlated with or proximate to those described above, were also considered. These included compositional complementarity (Pressey and Nicholls, 1989; Margules and Pressey, 2000), understood as the immediate species loss upon site removal (Economo, 2011), initial richness of the focal site, deviation in local environmental conditions relative to regional averages and various additional centrality measures. However, variable importance analysis in the case that multiple correlated predictor variables is not advised (Archer and Kimes, 2008; Strobl et al., 2008; Toloşi and Lengauer, 2011). Therefore we selected five, largely independent predictors that could be empirically measured.

Predictor	Notation	Calculation
Community distinctness	β_x^s	Mean between site BC dissimilarity
Area/biomass of removed site as regional percentage	B_x	$100 \cdot \left(\sum_i B_{ix}^{\text{before}}\right) / \left(\sum_i \sum_x B_{ix}^{\text{before}}\right)$
Distance weighted degree	κ_x	$\sum_y \delta_{xy} \exp\left(-d_{xy}\right)$
Community turnover	$\bar{\beta}_x^t$	log (Mean between time point BC dissimilarity + 10^{-3})
Betweenness centrality	C_x^B	Standard computation

Table 2: Predictors of the impact of site removal. We use the logarithm of the turnover because we find a rapid drop-off in regional species loss with the untransformed rate.

Random forest models were run in the R statistical software environment (Version $4.0.3$, R Core Team 2020) using the package 'randomForest' (Liaw et al., 2002). In all cases 1000 trees were assessed, with 1 variable tested at each branch. We ran the analysis for each value of the abiotic niche width w independently to ask how conclusions depend on the ratio between niche width and total landscape side length. We report the estimated importance of each variable measured as the percent increase in mean standard error that results when a given predictor is randomised. We also report the Spearman's rank correlation coefficient representing the strength and direction of the association between impact and a given predictor, since it is a non-parametric measure of statistical dependence. Unless otherwise stated p-values of the correlation tests were much less than the significance threshold however

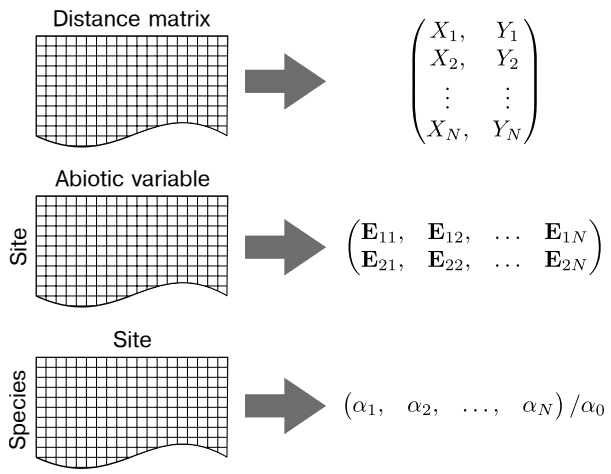

Figure 3: **Extracting data for fitting to empirical observation.** To fit the LVMCM to empirical observation we extract spatial coordinates, scaled to a square of longest side length $\sqrt{20}$ so that parameter values correspond to those used in the random metacommunity simulations. From the species by site presence-absence table, we extract the local site richness, presumed to scale with site area (quality). This is then used to generate λ, and λ' which scale the interaction coefficients $\mathbf{A}_{ij}$ so that the simulated metacommunity reproduces the observed distribution in α-diversity. From the available environmental metadata we extract the first two principle components and input these as the vectors $\mathbf{E}_k$ in the simulation.

given the large sample sizes possible in simulated datasets, random forest regression and correlation coefficients offer more informative summary statistics.

2.7 Fitting to empirical observations

We demonstrate how the LVMCM can be fit to data using two datasets, one describing the distributions of diatoms (D) in lakes straddling the Peruvian-Bolivian border (Benito et al., 2018), the other the distribution of Lichen-Fungi (LF) in Eastern Brazil (da Silva Cáceres et al., 2017). These datasets were chosen for their high diversity and because we assume that these ecological networks are generally simple and can therefore be reasonably represented by a model with a single trophic level. For the D dataset, which covers a large region, we selected a subset of the full database for which the first two principal components of measured abiotic environment formed a distinct cluster. For the LF dataset we reduced the total number of sites (and the computational effort) by pooling observations separated by less than 20km.

The approach is as follows: We extract from each data set a Euclidean distance matrix based on Universal Transverse Mercator coordinates, a multivariate environmental distribution, and the species richness recorded in each site (Fig. 3). The first two are input directly into the model. For better comparison to the parameters used

above, we scale the spatial network such that it occupies an arena for which the longest extension was $\sqrt{20}$. We again use a Gabriel graph to model the landscape connectivity, but note that alternative networks based on topographical data could be readily implemented.

For the environmental distribution we extracted the first two principal components of variation in the measured abiotic environmental variables, and input these as the vectors $\mathbf{E}_k$. For the D dataset, we adopted the approach of the original authors who used the NIPALS (nonlinear estimation by iterative partial least squares) algorithm, which allows ordination of datasets including missing values (Ibáñez et al., 2012), to extract the principal components of 14 environmental variables describing water chemistry, average climate, climate seasonality and variables describing the topography of landscape (for more details see Benito et al. 2018). The LF dataset did not include quantitative environmental measurements. Therefore we extracted elevation, mean annual temperature, and mean annual precipitation data from publicly available remote sensing databases (further details given in the Appendix).

Because the ecological interactions between species are insufficiently known, we randomly sampled the interaction coefficients $\mathbf{A}_{ij}$ from the same discrete distribution used in the random metacommunities. For a community model with no spatial extent for which $P(A_{ij} = 0.3) = 0.3$, and otherwise $A_{ij} = 0$, the onset of ecological structural instability occurs when $S_0 \approx 40$. From the SAR the effective area of a given site can be estimated using the equation

$$a_x = \left(\frac{S_x}{S_0}\right)^{\frac{1}{z}}, \tag{6}$$

were again we assume $z = 1$, that within single sites species richness accumulates linearly with area. Substituting Eq. 6 into Eq. 2 the scaling of the competitive overlap matrix for the fitted metacommunity model can be written

$$\mathbf{A}_x = \left(\frac{S_x}{a_0 S_0}\right)^{\frac{v}{z}} \mathbf{A}_0 + \left(\frac{S_x}{a_0 S_0}\right)^{\frac{v'}{z}} \mathbf{I}, \tag{7}$$

where S_x, the observed richness as a given site, are extracted from the empirical observation.

When assembling fitted metacommunity models, we again made the simplifying assumption that abiotic niche widths, w_k, are identical for both environmental variables and identical for all species. Greater accuracy in model fitting could be achieved by estimating the typical abiotic niche widths of each variable independently, and by incorporting inter-specific variation in niche width.

For simulated LVMCM metacommunities with environmental and topographic distributions matching empir-

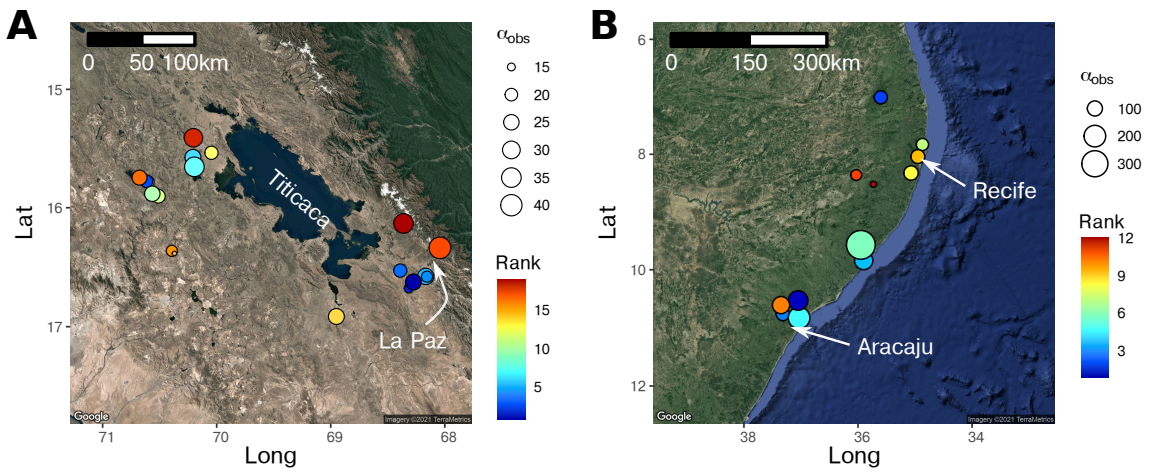

Figure 4: **Datasets used to fit the LVMCM. A**: The subset of the Andean diatom database (Benito et al., 2018) straddling Lake Titicaca and the Peruvian-Bolivian border. **B**: The Brazilian lichen-fungi database (da Silva Cáceres et al., 2017) with nearby samples pooled together. The colour of the points represents the ranked impact of site removal in simulated experiments: the highest ranked site, equal to the number of sites in the survey, was found to have the greatest impact on regional diversity in *in silico* removal experiments. The size of the points represents the observed species richness.

ical observations we fit the parameters w and ℓ based on the match between the observed and simulated site occupancy distribution (SOD), which describes the number of species observed in each occupancy class ranging from 1 to N the number of sites in the dataset. The SOD of the observation was quantitatively compared to the simulation by fitting to a pseudo-mechanistic Patch-Dynamics approximation of the LVMCM which is outlined in the Appendix and will be explored in more detail in future work.

3 Results and discussion

The impact of single site removals in random metacommunities was remarkably varied. To illustrate this, we show in Fig. 5 two site removals that result in dramatically different outcomes despite approximately equal immediate biomass losses. In Fig. 5A, the majority of extinctions occur immediately as 'endemics' (species which occupy only the focal site) are lost. With further relaxation, a handful of secondary extinctions occur, however the impact is largely localised in time and space. In Fig. 5B, few species are lost immediately, however the removal precipitates an extinction 'cascade' as a substantial extinction debt is paid. While a rigorous analysis of this extinction debt is not the primary focus of this study, the fact that extinction cascades occur in competitive metacommunities in the absence of trophic or direct mutualistic interactions, is an important result and a source of concern since it

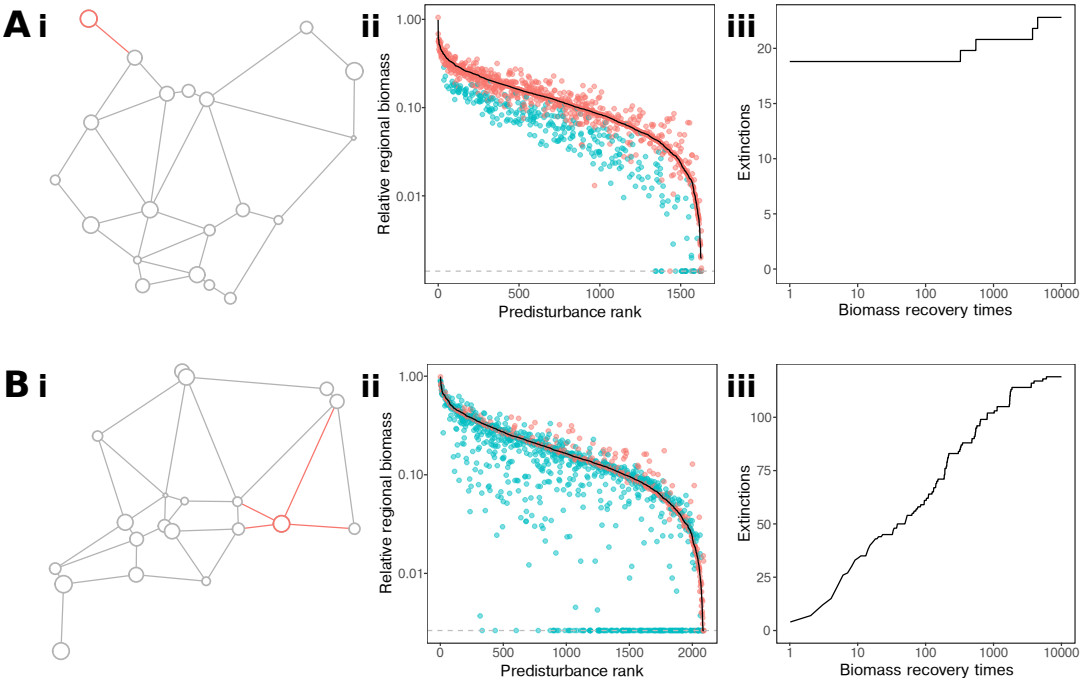

Figure 5: **Simulated site removal experiments.** Each site and all corresponding edges in a 20 patch metacommunity is removed and metacommunity dynamics are advanced 10^4 biomass recovery times. As a result, the distribution of biomass within the assemblage changes with some species driven regionally extinct. Here we show the outcomes of two such removals with the focal site and corresponding edges coloured in red (i). The regional-scale impact of the removal is represented as the change in the relative rank biomass curve, with biomasses summed over the landscape (ii). The rank biomass curve of the pre-disturbance metacommunity is represented by the black line. The points represent the post-disturbance relative biomass plotted with species rank unchanged. The blue points correspond to species with biomass in the removed site greater than the extinction threshold of 10^{-4} biomass units. The red points correspond to species which are considered absent from the focal site and therefore are indirectly impacted by the site removal. The dashed lines represent the regional extinction threshold; biomasses below this value are plotted on the line. The cumulative extinctions reveal very different temporal responses to site removal with some removals resulting in the losses of endemics only while others precipitate extinction cascades (iii).

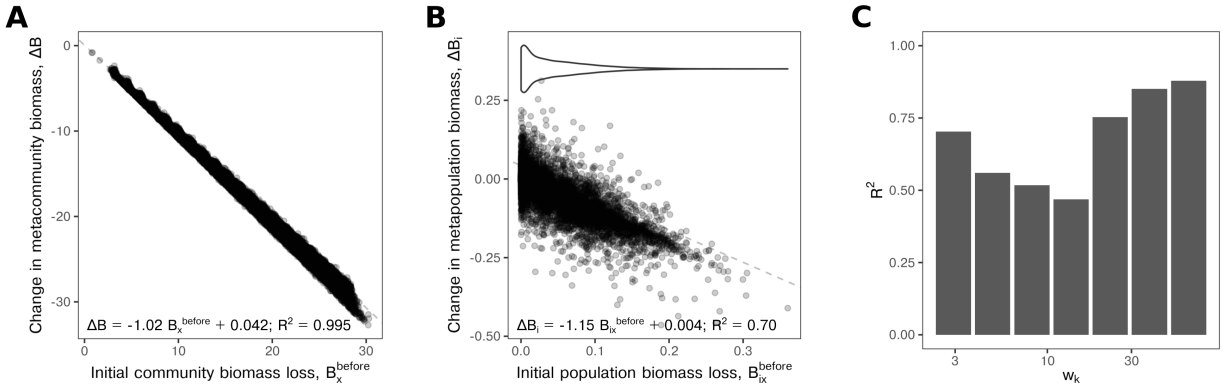

Figure 6: **Change in biomass in response to site removal.** **A**: Immediate community biomass loss as a predictor of long-term metacommunity impact. To a good approximation the total loss of biomass at the metacommunity scale is equivalent to the total biomass (area) of the removed site. That long term biomass losses tend to slightly exceed immediate removals is likely the result of edge effects. **B**: Immediate metapopulation biomass loss as a predictor of long-term species-level impact. Only species with B_{ix}^{before} greater than the extinction threshold in the removed site are included in this analysis. The spread in the data shows that long-term impacts can differ substantially from initial losses. Coloured lines represent the estimated 90% quantile for different abiotic niche widths. The overlaid violin plot, representing the frequency of local populations of a given biomass in the full dataset, highlights the numerical dominance of low biomass species in removed sites. Of the 7×10^6 species included in the full experiment, a random subsample of 10^4 metapopulations are shown here. **C**: Initial biomass loss is least predictive of long-term change in metapopulation at intermediate niche widths shown here in the R^2 of the linear regression.

suggest such secondary extinction cascades can occur in almost any conceivable ecological network. Since we have no evidence to suggest that such cascades occur in competitive models without an explicit spatial component, we consider this an emergent property of metacommunities, driven by the delicate, self-organised spatial structure of the regional assemblage.

3.1 Long-term shifts in biomass

First we ask to what extent the change in biomass can be predicted by the biomass previously detected in the removed site. Later we explore potential predictors of change in regional species richness. The long-term loss of biomass at the regional scale ($\Delta B = \sum_i \sum_x (\mathbf{B}_{ix}^{\text{after}} - \mathbf{B}_{ix}^{\text{before}})$), is well predicted by the total biomass removed initially (B_x^{before}, Fig. 6A). We find a weak tendency for long-term impacts to exceed the biomass of the removed site ($|\Delta B| > B_x^{\text{before}}$). This is likely the result of increased edge effects following site removal. In reality we might expect this effect to be exaggerated in some ecosystems. Strong edge effects can occur, for example, due to regime shifts in light and humidity levels at fragment edges, leading to a shift in the relative abundance of small-bodied

pioneer species, though empirical evidence of such an effect varies (Chen et al., 1992; Laurance et al., 1997; Haddad et al., 2015; Reinmann and Hutyra, 2017). This minor deviation notwithstanding, we find (unsurprisingly) that the total change metacommunity biomass is trivially predicted by the total biomass (area) removed. Inspection of Fig. 6A suggests that if the primary goal of ecological management is the maximisation of biomass then area is indeed a good predictor of long-term conservation value.

The change in regional species biomass ($\Delta B_i = \sum_x (B_{ix}^{\text{after}} - B_{ix}^{\text{before}})$) was largely uncorrelated with the total biomass lost immediately upon site removals due to the fact that site area alone is generally a poor predictor of species level biomass, which is strongly regulated by inter-specific biotic effects. We do find, however, that the long-term impact on metapopulation biomass of the ith species was correlated with the initial *species-level* biomass loss (B_{ix}^{before}, Fig. 6B). Inspection of Fig. 6B suggests that if one is primarily concerned with the average regional impact on single species, then biomass of the focal species immediately lost following site removal can be fairly predictive. However, the residuals of the linear regression were particularly large amongst rare species, which constitute the majority of the biodiversity of any ecological community (density of points represented by violin plot inset Fig. 6B).

Interestingly, initial population biomass in the removed site was least predictive of the long-term impacts on metapopulation biomass at intermediate values of w (Fig. 6C). This can be understood as follows: For homogeneous landscapes (wide abiotic niches) occupied by metacommunities reasonably approximated by a single spatially unstructured community, slow relaxation and spatial restructuring of the metacommunity after site removal are minimal. As a result the long-term impact on the ith species corresponds closely to the total biomass of species i initially removed—minimal additional losses in remaining sites occur. In highly heterogeneous environments (narrow niches) species tend to occupy few sites and spatial restructuring is impeded by strong environmental filtering. Thus we again find that long-term impacts are well predicted by the initial biomass removed. In the intermediate case, metacommunity responses to perturbation are more complex, and so the association between initial losses and long-term impacts is weakened.

We find that some 53% of all species modelled experienced long-term regional biomass losses in excess of immediate loss B_{ix}^{before}, 25% experienced losses less than the original impact implying compensatory gains elsewhere, and 22% of metapopulations actually benefit regionally ($\Delta B_i > 0$). Interestingly, decomposing the data into species which lose and gain biomass regionally, we find that *relative* impacts on metapopulation biomasses ($|\Delta B_i|/B_{ix}^{\text{before}}$) span multiple orders of magnitude. For species which lose biomass, relative losses of $\log_{10}\left(|\Delta B_i|/B_{ix}^{\text{before}}\right) > 0$ indicate metapopulations which experience a regional collapse following site removals

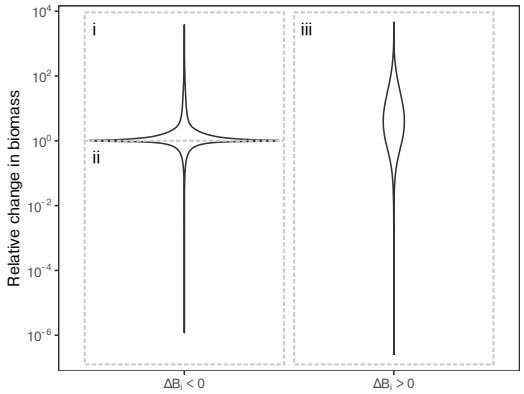

Figure 7: **Logarithmic relative shifts in metapopulation biomass decomposed by species which lose or gain regionally following site removals.** While most species which lose biomass after relaxation $\Delta B < 0$ experience impacts of the same order as the original perturbation, species can either lose orders of magnitude more biomass than lost immediately, implying a species-level collapse (i) or lose a small fraction of that removed initially—i.e. partially compensate loses by gains elsewhere in the metacommunity (ii). Species that gain biomass following site removal, $\Delta B > 0$, can experience a proportional increase of up to multiple orders of magnitude (iii). These violin plots highlight both the complexity and the range of species-level responses to habitat destruction in even simple ecological models.

(Fig. 7i). In contrast, relative losses of $\log_{10}\left(|\Delta B_i|/B_{ix}^{\text{before}}\right) < 0$ imply metapopulations which partially compensate for immediate losses by gains elsewhere in the metacommunity (Fig. 7ii). For species which gain biomass regionally following site removal, proportional gains can reach several orders of magnitude greater than initial losses, likely in cases that rare species that substantially increase their share of the total biomass in response to simultaneous losses for competitor species. Thus, while absolute losses do, on average, correlate with long-term impacts (Fig. 6B), relative impacts can substantially deviate, in particular among rare species, likely to be the primary targets of population scale conservation measures.

3.2 Long-term shifts in regional diversity

We now report the long-term impact of habitat removal on regional biodiversity, and show that this is much less easily predicted based on initial biomass losses alone. We consider the long-term percent species loss as representative of the 'conservation value' of a given site, and therefore assume that preservation of undifferentiated biomass alone is not a sufficient goal of conservation. In 57.9% of the simulated removal experiments we observe the extinction of at least one species at the regional scale. Of these, the highest impact of a single site removal was a loss of $\Delta\gamma = 23\%$ of pre-disturbance species (Fig. 8A). We find a clear positive association between con-

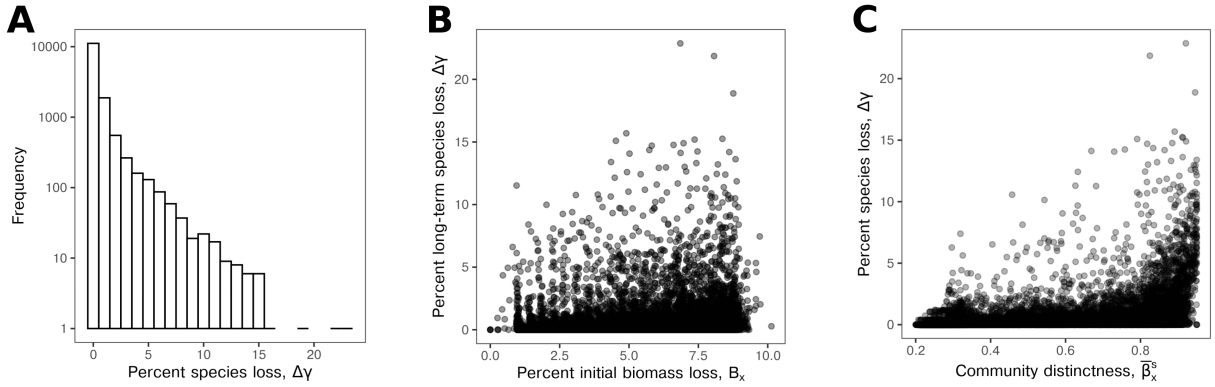

Figure 8: **Change in regional diversity in response to site removal.** **A**: The distribution in percent regional diversity loss $\Delta\gamma$ for the full simulation study ($n = 1$). While most site removals lead to percent species losses at the regional scale of less that 5%, catastrophic species losses of over 20% are possible, though rare. The two dominant predictors of $\Delta\gamma$ were immediate biomass loss (**B**), and the compositional distinctness of the removed site (**C**).

servation value and site biomass (area) as a percentage of the regional whole, but with substantial spread (Fig. 8B, $\rho = 0.32$, full dataset). The strongest association was with compositional distinctness, represented by the mean spatial BC dissimilarity of the removed site (Fig. 8C, $\rho = 0.41$, full dataset). Of the remaining predictors, percent species loss was weakly positively associated with distance weighted degree ($\rho = 0.08$, full dataset), negatively associated with temporal turnover ($\rho = -0.15$, full dataset), and weakly negatively associated with betweenness centrality ($\rho = -0.04$, full dataset). Random forest regression quantitatively ranks predictors according to their 'importance' based on the percent increase in mean squared error detected when a given predictor is randomised. The most important predictor of conservation value was invariably the compositional distinctness of the removed site, followed by its biomass as a percent of the full metacommunity, distance weighted degree, temporal turnover rate and finally its betweenness centrality (Fig. 9).

While the ordering of the predictive power of the five predictors does not appear to differ greatly under variation in abiotic niche width (Fig 9), the effect sizes vary substantially and accordingly so does the ability to quantitatively predict the impact. To show this we ran the random forest regression for each value w independently. Predicting $\Delta\gamma$ in metacommunities of very low abiotic heterogeneity is particularly challenging. When niche widths are large, and species relatively insensitive to the underlying environmental distribution, biotic interactions (not considered in the statistical modelling) and dispersal (indirectly included via spatial metrics) play a much greater role in determining the metacommunity response to perturbation. For high w, the proportion of variance explained by

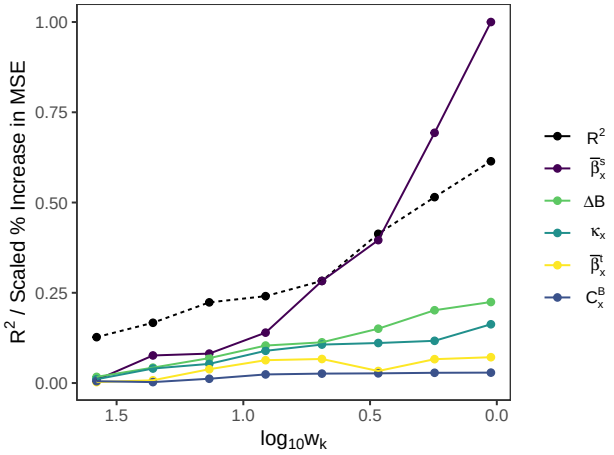

Figure 9: **The importance of the predictors estimated using random forest regression.** The variable importance estimated by random forest regression of the compositional distinctness, $\bar{\beta}^s_x$, total biomass, B_x, temporal turnover, $\bar{\beta}^t_x$, and betweenness centrality, C^B_x, of the removed site. We decompose the simulation data by the environmental niche width, w in order to ask how the importance estimates and overall model fit vary with spatial scale (effective heterogeneity).

the random forest regression was low, typically $15 - 25\%$. When niches are narrow compared to environmental variation, these unmodelled effects, represented by the random forest regression model as unexplained variance, are outweighed by the impact of abiotic filtering. For small w, the variance explained reached a maximum of 63%.

We consider spatial predictors of conservation value at three distinct scales, local connectivity represented by distance weighted degree, mesoscale connectivity represented by betweenness centrality, and regional scale connectivity via the characteristic dispersal length ℓ. That site degree weighted arbitrarily by an exponential decay with distance is positively associated with long-term losses suggests that even in simple models, greater connectivity denotes greater conservation value, though the effect is weak and the mechanism responsible is not immediately clear. It is possible that mass effects from highly connected sites can regulate resident biomass in adjacent sites and thereby permit the unstable coexistence of competitors, increasing the conservation value of the more connected sites. Betweenness centrality was the least predictive of long-term species loss, though we do find a weak negative association. Betweenness centrality is minimised at the exterior of the network suggesting a tendency for species losses to be greater when sites at the edges of the network are removed. This could be due to the fact that metapopulations tend to spatially redistribute following site removals, and this spatial relaxation process is restricted at the edge of the network. Whether this is effect translates to natural ecosystems depends strongly on the particulars of the spatial structure, since not all natural systems are likely to be subject to strong spatial edge effects. We might,

however, expect sites near strong barriers to dispersal such as mountain ranges or sharp transitions in habitat type to have particularly high conservation value for this reason. Interestingly we find no evidence of a relationship between $\Delta\gamma$ and dispersal length ℓ (Fahrig, 1997). The parameter ℓ represents connectivity at the regional scale and the absence of a relationship between ℓ and $\Delta gamma$ implies that source-sink dynamics have little impact on regional persistence. Connectivity has long been considered vitally important to conserving biodiversity at the landscape scale (Taylor et al., 1993). Biotic corridors are understood to play a key role in recolonisation following stochastic local extinction (Reed, 2004), to promote genetic diversity (Levey et al., 2005) and facilitate community response to climate change (Robillard et al., 2015). It has also been proposed that a competition-colonisation trade off between species could further buffer a regional assemblage against perturbation (Nee and May, 1992; Livingston et al., 2012) though empirical evidence of such an effect varies (Levine and Rees, 2002; Cadotte et al., 2006; Limberger and Wickham, 2011; Livingston et al., 2012; Bin et al., 2019). It is plausible that inclusion of environmental stochasticity or directional change could result in greater dependence of $\Delta\gamma$ on ℓ in metacommunity models, however in the absence of this driving force, regional scale connectivity does affect long-term outcomes.

The negative association between temporal turnover and long-term conservation value likely arises due to the fact that turnover in the model is purely autonomous, driven by ecological interactions and dispersal (O'Sullivan et al., 2021) as opposed to external processes such as abiotic fluctuations or regional-scale invasions. Autonomous turnover requires ecological redundancy as well as spatial connectivity to function; without a pool of species in adjacent sites capable of passing through local abiotic filters, communities converge on fixed points in the absence of external driving forces (Roy et al., 2020; Pearce et al., 2020; O'Sullivan et al., 2021). Those sites undergoing rapid autonomous turnover have least impact on regional diversity since the ecological redundancy required for steady state dynamics means the metacommunity can readily respond to their loss. In a sense, then, those sites for which spatial and environmental conditions permit autonomous turnover can be seen as buffering the regional community against disturbance. Correlation coefficients and rank variable importance estimated by random forest regression are summarised in Table 3.

Broad characterisation of the impact of the various predictors confirms that if the unit of impact is species losses, area alone is an insufficient predictor (Thompson et al., 2019). That compositional distinctness would predict long-term impacts on regional biodiversity may not represent a complete paradigm shift, but that it typically exceeds biomass in driving long-term losses, is an important insight. Thus, we find that while it is plausible that average species- or community-level change in biomass can be predicted based on the initial area or biomass removed alone, this is not the case for impacts measured in terms of species losses, a unit we consider the more

Predictor	Spearman's ρ	Rank importance
β_x^s	$(0.13, 0.56)$	1
B_x	$(0.16, 0.26)$	2
κ_x	$(0.05, 0.19)$	3
$\bar{\beta}_x^t$	$(0.01, -0.24)$	4
C_x^B	$(-0.02, -0.1)$	5

Table 3: Correlation between predictors and long-term species loss. Spearman's ρ (min/max observed correlation under variation of w) describes both the approximate strength of the association and its direction. Absolute correlation coefficients generally increase with decreasing w as abiotic processes become increasingly dominant. Predictors are ranked by the percent increase in MSE upon randomisation.

appropriate when assessing conservation value. In order to estimate the long-term effects of habitat destruction on biodiversity, it is critically important to take the community composition of the affected areas into account. Above we describe the results of the removal experiment for the case that a single environmental distribution was modelled ($n = 1$). We find the conclusions are not impacted by the inclusion of greater dimensionality in the environment/Hutchinsonian niche (Fig. S2). We also find that the summary statistics of the model fits are almost unchanged by subsampling each metacommunity to the regional diversity observed for the largest niche width, suggesting the increase in γ with inverse w does not artificially impact model outcomes.

3.3 Fitting the model to empirical data

Having explored the impact of site removal in arbitrary LVMCM metacommunities, it was critical to ascertain which region of the parameter space is most representative of natural ecosystems at large scales. For this we took the spatial, environmental data, and recorded richness data from two empirical datasets and ask which combination of w and ℓ best reproduces the spatial structure (site occupancy distribution) of the natural system. We find that the D dataset was best fit when $w = 3.46$, $\ell = 7.74/cdot10^{-3}$. In the case of the LF dataset $w = 2.21$, $\ell = 1.29$ gave the closest match to the observation. That the LF assemblage is best fit by a smaller abiotic niche width reflects in part the fact that the total area of the LF dataset was much larger than that of the subset of D fit here. That the best fitting dispersal length of the LF dataset is also much larger partly reflects that fact that the sampling of this metacommnity is likely to be much less complete, again due to the large area surveyed and the contiguous nature of the habitat. When discretized as we do here, the connectivity of 'adjacent' sites is likely to be over-estimated. In both cases the empirical between-site diversity showed little evidence of a distance decay, suggesting sites were samples at sufficient distance in Cartesian and environmental space that little compositional similarity was detected, with the exception of the most ubiquitous species. A similarly rapid drop off in community similarity

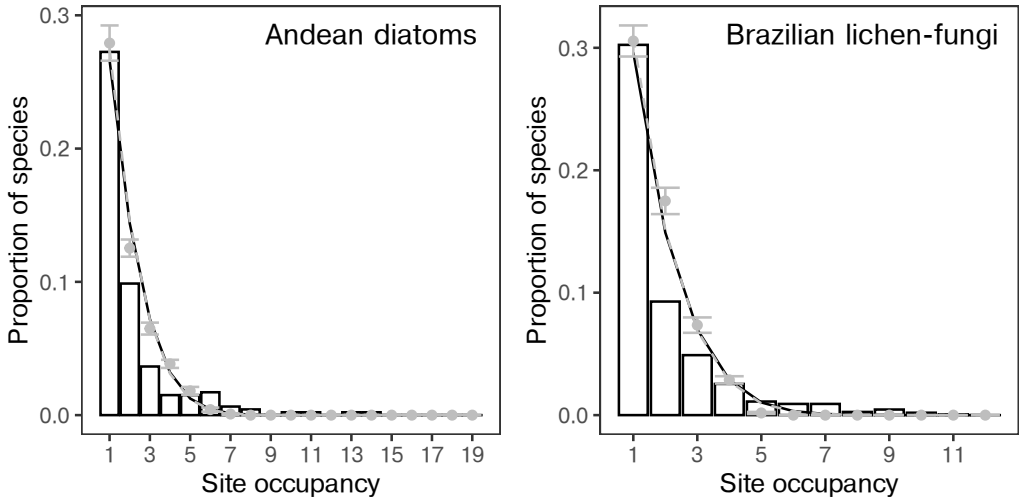

Figure 10: **The proportional site occupancy distributions of the three empirical datasets and the simulated metacommunities which best fit them.** The barplots represent the empirical SODs while the black curves are the occupancies of the patch dynamics models with mixing rate m estimated from the observations. The grey points are the mean occupancies of 10 replicate simulations with parameters that best fit the empirical observations. The grey dashed lines fit the patch dynamics model to the simulations.

with distance was found in the simulation.

The simulations deviate from the empirical data in a number of important respects. At the regional scale the best fitting simulations overestimated the regional diversity by a factor of $5.6 - 6.1$ for the D and $1.7 - 1.8\%$ for the LF dataset, with comparable over-estimates at the local scale. The current method for fitting the local interaction matrices is numerical, and therefore approximate, and it is conceivable that deviations from the power law CAR emerge for large areas. However, since large metacommunities do not reach fixed points but instead manifest persistence steady state dynamics (O'Sullivan et al., 2021), our method for computing regional interaction matrices $\mathbf{C}$ (O'Sullivan et al., 2019) is not directly applicable. We also find that the empirical SODs tend to have a sharper peaks at single site occupancy than the simulations and slightly fatter right tails. This is most likely due to the model's simplifying assumption that all species have the same abiotic niche width. Inspection of the SOD for various niche widths (Fig. S1A,B) suggests that a better fit to the empirical distribution could be achieved by relaxing this assumption. To see this, it seems plausible that metacommunities including species with a mixture of large and small abiotic niche widths would lead to a sharp peak at a site occupancy of 1 while increasing the number of large occupancy species. However, estimating the distribution of generalist and specialist species and include this additional heterogeneity in the model is beyond the scope of the current study.

Despite these differences, this simple model fitting procedure generated model metacommunities with mac-

roecological spatial structure (SOD) approximately matching that of the empirical observation. Critically, both datasets are fit by abiotic niche widths of order $log_{10}\ w = 0$ suggesting the parameter space in which compositional dissimilarity substantially exceeds biomass as a predictor of conservation value is the most biologically relevant.

From these fitted models we then make a mechanistic assessment of the rank conservation priority of each site (Fig. 4) using the systematic removal experiment described above. In simulated removal experiments the percent species lost range from $1.1 - 13.9\%$ for the D and $0.1 - 16.9\%$ for the LF dataset (Fig. 11). In both cases, site area was a rather poor predictor of simulated species losses. On the other hand, a strong non-linear relationship between impact and community distinctness was found, consistent with previous results at these abiotic niche widths. For the D dataset, the rank conservation priority typically increased with altitude as spatial isolation and deviation from mean environments increased. Surprisingly, for the LF dataset the greatest impact occurred when the two smallest sites were removed. This apparently incongruous result is explained by the fact that for this dataset, observed species richness was negatively correlated with environmental rarity measured as the mean Euclidean distance of the xth site to all sites y in environment space (Spearman's $\rho = -0.56$, $p = 0.057$). In the random metacommunities studied above, site area was independent of spatial positional or local environment. In the LF community, the case that small sites occupy the most distinct environments, and therefore host the most distinct communities, means the relationship between long term impact and site area is reversed due to the dominance of community distinctness in driving conservation value.

4 Conclusions

With this study we set out to probe whether the impact of habitat destruction, measured in terms of long-term changes in biomass or biodiversity, can be predicted in simple models. We also aimed to test the assumption that area is a sufficient predictor of long-term impacts, implicit in common area-based approaches to allocating conservation priorities (Brown, 1984; Durrett and Levin, 1996; Thomas et al., 2004; Foster et al., 2013). We found that total metacommunity and average metapopulation biomass change can be predicted directly by the initial biomass removed from the model. However, the predictive power of biomass alone is suppressed for rare species—which can undergo relative shifts in regional biomass that differ from the initial impact by multiple orders of magnitude—and for metacommunities of intermediate environmental heterogeneity.

In effect our study aimed to test the hypothesis that single sites can contribute disproportionately to regional

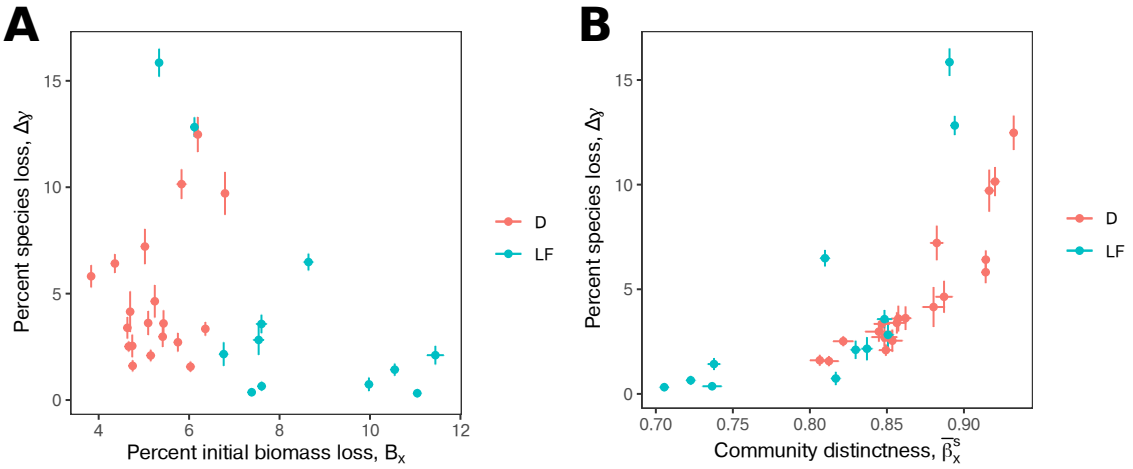

Figure 11: **Long term impact of site removals in fitted LVMCM models.** Simulated removal experiments are used to predict long term species losses on site removals for models fit to empirical datasets. For both datasets considered, percent biomass removed intially (A) is a worse predictor of long term species losses than compositional distinctness (B).

diversity. We find that while it is clear that some communities can be understood as keystones in this sense (Mouquet et al., 2013), even in simple models predicting 'keystoneness' is a major challenge (Resetarits et al., 2018; Yang et al., 2020). We show that area alone is an insufficient predictor of regional species loss, which depends more sensitively on the compositional distinctness of the removed habitat. In extreme cases, due to the evident non-linearity of the response, conservation value can decrease with area if the drivers of compositional change are non-randomly associated with effective area.

We find only weak and scale dependent evidence for a role of connectivity in determining conservation value. This suggests that, while connectivity has a key role to play in local coexistence, at regional scales, in the absence of specific biological processes such as competition-colonisation trade-offs, connectivity does not strongly influence regional coexistence. It's important to note, however, that removal of the most central sites increases the proportion of edge sites and decreases the average centrality of all remaining sites, thereby increasing the likelihood of species losses in the case of secondary perturbation. Furthermore, greater connectivity is likely to buffer against large-scale, directional abiotic change not included in the present study.

By fitting the LVMCM to empirical data we have shown which regions of the parameter space of the model are most directly applicable to studies of landscape-scale ecology. We find that for both datasets explored, the spatial scale of sampling is such that measurable abiotic properties of the landscape dominate the assemblage response to perturbation and therefore that metacommunity processes may indeed be amenable to modelling without substan-

tial knowledge of the network of interactions governing local biomass distributions. However, even at these scales biotic responses to perturbation remain complex and as such we do not advocate for predictive models which completely overlook the potential role for biotic interactions in long-term ecological outcomes. While the procedure for fitting to data demonstrated here is highly simplified, and the model lacks a number of key biological processes, by fitting the LVMCM to data we have shown that bespoke, mechanistic modelling studies such as these could play an important role as a decision support tool when designing conservation strategies (Chase et al., 2020).

The fact that we find reasonably correlation on average between immediate and long-term biomass losses at the single species scale (Fig. 6B), despite a clear tendency for complex restructuring of metapopulation distributions follow site removal is in itself and important outcome of this study. Assuming that, for all species i of the guild in question, estimates of global population biomass $B_{\mathrm{G},i}$ are available, this can be used to compute the expected impact of site removal on the quantity $\mathcal{L} = \sum_i \log(B_{\mathrm{G},i})$, which is closely related to geometric-mean-abundance biodiversity indices (Buckland et al., 2011) such as the Living Planet Index loh2005living. To first order in the local changes in biomass, the expected long-term change in $\mathcal{L}$ resulting from removal of site x can simply be approximated as $\Delta\mathcal{L} \approx \sum_i (B_{x,i}/B_{\mathrm{G},i})$, with the sum running over all members of the guild present at x. This result might prove useful for defining a simple scientific indicator of local biodiversity impact.

Finally, though explored only indirectly in the present study, we have found early evidence that the highly simplified patch dynamics used to fit simulation to observation may indeed capture some important biological processes. While further work is needed to fully understand the limits of this model fitting procedure, there are encouraging signs that the structural complexity (despite the dynamical simplicity) of the LVMCM may be incorporated into an analytically tractable framework.

Appendix

S1 Harvesting experiment and the CAR

Following (O'Sullivan et al., 2019) we computed an $S \times S$ spatially unresolved competition matrix $\mathbf{C}$ that summarised biotic interactions at the regional scale. For this we constructed a spatially unresolved Lotka-Volterra system,

$$\frac{d\mathbf{B}_i}{dt} = \left(\rho_i - \sum_{j=1}^{S} \mathbf{C}_{ij}\mathbf{B}_i \right) \cdot \mathbf{B}_i, \tag{S1}$$

where $\mathbf{B}_i$ represents total biomass of species i, as an approximation of the spatially resolved model. The regional scale interaction coefficients $\mathbf{C}_{ij}$ were then estimated using a computational harvesting experiment, which assumes that interaction strengths can be inferred from the changes in regional abundances that result from controlled changes in the regional abundances of harvested species (Gilbert et al., 2014).

For the unresolved model, Eq. S1, the harvesting of species i at a rate h, represented by substituting $\rho_i - h$ for ρ_i in Eq. S1, produces a shift in the equilibrium biomasses given by

$$\sum_y \left(\mathbf{B}_{jy} - \mathbf{B}^*_{jy} \right) = \Delta \mathbf{B}_j = -\mathbf{C}^{-1}_{ij} h. \tag{S2}$$

where $\mathbf{B}^*$ represents the unharvested equilibrium. Then

$$\frac{d\mathbf{B}_{ix}}{dt} = \left(\sum_{jy} \mathbf{J}_{ixjy} (\mathbf{B}_{jy} - \mathbf{B}^*_{jy}) \right) - h\mathbf{B}^*_{ix}, \quad \text{and} \tag{S3a}$$

$$\frac{d\mathbf{B}_{kx}}{dt} = \left(\sum_{jy} \mathbf{J}_{kxjy} (\mathbf{B}_{jy} - \mathbf{B}^*_{jy}) \right) \quad \text{for } k \neq i\ . \tag{S3b}$$

where h is the harvesting rate, and $\mathbf{J}$ is the numerically estimated $SN \times SN$ Jacobian of the full metacommunity model. We vectorize the matrix $\mathbf{B}$ (denoted $\vec{\mathbf{B}}$) to match the dimensionality of the spatially resolved Jacobian, and the write the equilibrium condition for Eq. S3a as

$$\mathbf{J}(\vec{\mathbf{B}} - \vec{\mathbf{B}}^*) - \vec{\mathbf{H}} = 0, \tag{S4}$$

where the elements of vector $\vec{\mathbf{H}}$ are hb^*_{jy} for $j = i$ and 0 otherwise. From this, we obtain

$$h^{-1}\left(\vec{\mathbf{B}} - \vec{\mathbf{B}}^*\right) = \mathbf{J}^{-1}h^{-1}\vec{\mathbf{H}}. \tag{S5}$$

The left hand side of Eq. S5 represents the local shift in biomasses due to the harvesting of the focal species i per unit h. From Eq. S5 we compute the change in total biomass $\Delta\mathbf{B}_j = \sum_x \mathbf{B}_{jx} - \mathbf{B}^*_{jx}$ for each species j which gives row i of $\mathbf{C}^{-1}$ (Eq. S2). Iterating over all species $i = 1 \ldots S$, we computed $\mathbf{C}^{-1}$ and from this the spatially unresolved interaction matrix $\mathbf{C}$. This method implicitly assumes that the metacommunity as at equilibrium with respect to biomass. In practice this assumption is violated by the onset of autonomous compositional turnover. For this reason we restrict our analysis to a range of N for which turnover is low.

We compute the exponents v of the CAR for a given combination of parameters by assembling random metacommunities spanning a range of N, computing the regional interaction matrices $\mathbf{C}$ as described, and linearly regressing $\log_{10} \bar{\mathbf{C}}$ against $\log_{10} N$ for mean inter and intra-specific coefficients independently (Fig. 2A).

S2 Remote sensing data collection

We extracted temperature, precipitation and elevation data for the 23 sites of the Brazilian Lichen-Fungi survey (da Silva Cáceres et al., 2017). The first two abiotic variable were time averaged over the 20 years from January 2000 to January 2020. We pooled nearby sites in order to reduce the dimensionality and therefore the computational load of the simulations. For these sites we took abiotic values, including spatial coordinates, as the between site averages.

Temperature data were collected from the MODIS satellite database (Hulley and Hook, 2017). For each site, sensor products were filtered by land surface temperature and emissivity to retrieve data from the MODIS thermal infrared bands. Precipitation data were collected from the IMERG satellite database (Huffman et al., 2014). Elevation data were taken from the GTOPO30 digital elevation model database (Danielson and Gesch, 2011) which provides data from collective satellite inputs, compiled by U. S. Geological Survey.

S3 Patch dynamics fitting model

We use a simple patch-occupancy model for metacommunities (called Metacommunity Patch Dynamics, MCPDM) as an intermediary for comparison of the site-occupancy distributions (SOD) of LVMCM and empirical datasets.

This is advantageous because the MCPDM has only a single fitting parameter, called the 'mixing rate', and can be solved analytically. Thus, comparison between SODs reduced to comparison of this single fitting parameter. We will discussion motivation and validity of the MCPDM in detail elsewhere. Here, we only provide a bare-bones statement of the model and its analytic solution.

As common in metacommunity theory, the MCPDM conceptualizes metacommunities in terms of distinct patches. In addition, it makes use of the observation that at each patch the number of co-existing species from a given guild is limited by ecological structural stability constraints. In the approximation that this limit is sharply defined, one can think of each patch as offering a number of 'slots' (given by the structural stability limit to local richness) that can be occupied by different species. The total number of 'slots' offered by a metacommunity is the sum of the number of slots over all patches. Each species present in the metacommunity occupies one or more slots, and the distribution of the numbers occupied is the SOD. If most species occupy much fewer slots than there are patches (as we find here), then one can disregard, in a first instance, the constraint that a species cannot occupy more than one slot in the same patch.

The MCPDM is defined through a Markov process for the SOD that results from iterative invasions into the metacommunty and 'mixing' in the following way: (i) A species newly invading the metacommunity will occupy one or more slots. The total number of slots it occupies is sampled as one plus a Poisson distributed number with mean $\lambda - 1$, where $\lambda - 1$ equals the mean number of slots resident species occupy in the model. The previous occupiers are removed from the slots (but may remain present in other slots). (ii) At a rate $m \geq 0$ per invasion event (the *mixing rate*), the current occupier of a randomly chosen slot will occupy an additional, randomly chosen slot (thus expanding its range). The previous occupier of this slots is then removed from that slot (but may remain present in other slots).

The steady state SOD for this process is such that, in a large metacommunity with a total of Z slots, the mean number of species occupying n slots is Zs_n, where for $m = 0$

$$s_n = \frac{1 - Q(n-1, \lambda - 1)}{n\lambda} \tag{S6}$$

and for $m > 0$

$$s_n = \frac{1}{n\lambda}\left[1 - \left(\frac{m}{\lambda + m}\right)^n - Q(n, \lambda - 1) + \exp\left(\frac{(\lambda - 1)\lambda}{m}\right)\left(\frac{m}{\lambda + m}\right)^{n-1} Q\left(n, \frac{(\lambda - 1)(\lambda + m)}{m}\right)\right], \tag{S7}$$

where $Q(a, x) = \Gamma(a, z)/\Gamma(z)$ is the regularized incomplete gamma function. Noting that the total number of species in this metacommunity is $Z \sum_n s_n$, the definition of λ translates to

$$\lambda - 1 = \frac{Z}{Z \sum_n s_n} = \frac{1}{\sum_n s_n}. \tag{S8}$$

This is solved numerically by adjusting λ.

To find the best fitting combination of w and ℓ for an empirical dataset, we estimated the mixing rate m for the empirical SOD, then, for model metacommunities with abiotic variables extracted from observation data as described above, we scanned assembled metacommunities with various combinations of the focal parameters S1. The best fitting combination of w and ℓ was selected as that with emergent SOD which closest matched the mixing rate m of the observation. Note, for simplicity we define the site occupancy with reference to source populations only. These are the local populations which persist following a long relaxation (10^4 biomass recovery times) with dispersal switched off. In effect this assumes that only source populations are detected in the empirical data due to type II sampling errors. This assumption explains the modest impact of ℓ on the simulated SOD S1. Alternative approaches to defining local site occupancy in simulations are available but typically require some assumed model, or arbitrary numerical threshold.

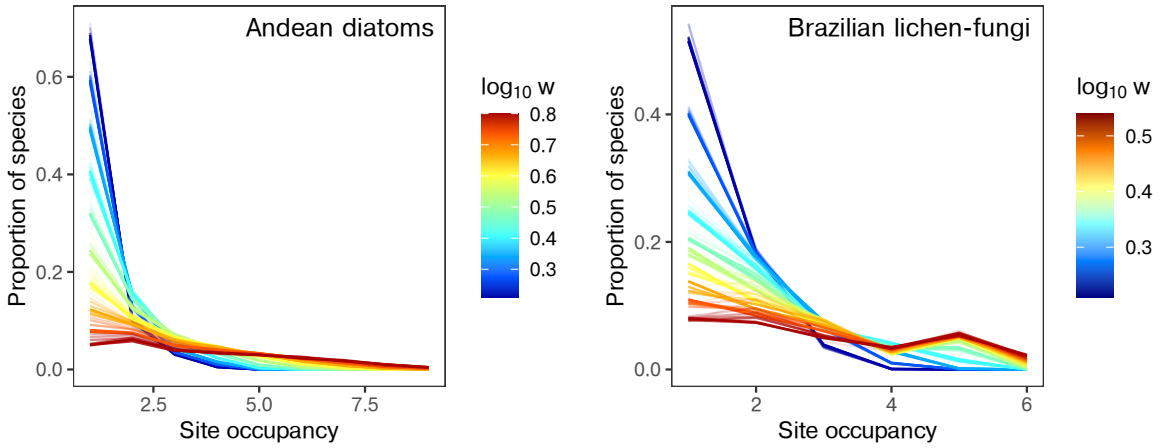

Figure S1: **The impact of parameter variation on the SOD for fitted models.** The average proportional site occupancy for various combinations of w and ℓ with landscape properties extracted from empirical datasets. Colour represents abiotic niche width w, line saturation dispersal length ℓ. We measure site occupancy using source populations only, which accounts for the small effect of dispersal on spatial structure.

S4 Two dimensional environmental variability

When two independent environmental variables are modelled, the outcomes of the model assembly and simulated removal experiments are qualitatively similar. Note that the variable importance of temporal turnover estimated by random forest analysis was greater when two environmental variables were used

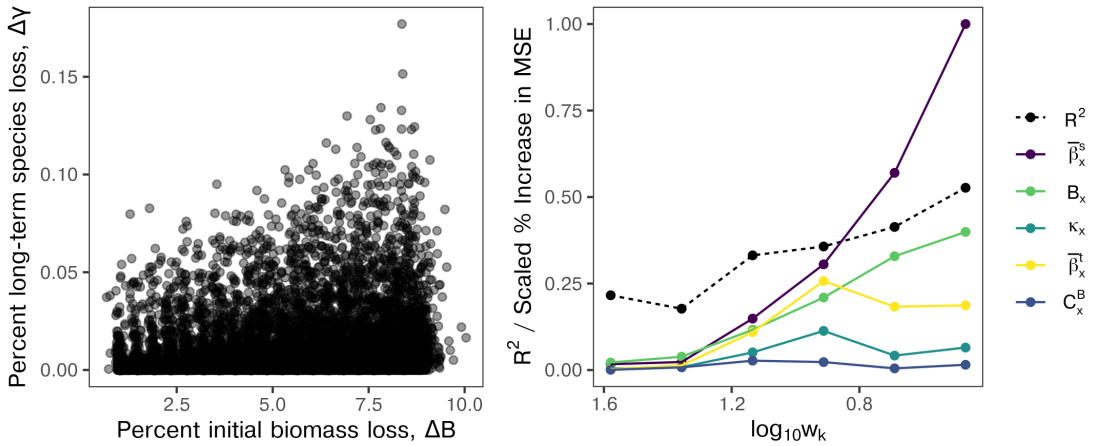

Figure S2: **The outcome of site removal experiments in two dimensional abiotic environments.** The impact $\Delta\gamma$ of single site removals in the case that the abiotic environment is modelled as two independent Gaussian Random Fields and the variable importance estimated using random forest regression.

References

Robert J. Adler. *The Geometry of Random Fields*. SIAM, 2010.

Andrew P Allen and Ethan P White. Effects of range size on species–area relationships. *Evolutionary Ecology Research*, 5(4):493–499, 2003.

Miguel B Araújo and Miska Luoto. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16(6):743–753, 2007.

Kellie J Archer and Ryan V Kimes. Empirical characterization of random forest variable importance measures. *Computational statistics & data analysis*, 52(4):2249–2260, 2008.

Matthieu Barbier, Jean-François Arnoldi, Guy Bunin, and Michel Loreau. Generic assembly patterns in complex ecological communities. *Proceedings of the National Academy of Sciences*, 115(9):2156–2161, 2018.

Xavier Benito, Sherilyn C Fritz, Miriam Steinitz-Kannan, Maria I Vélez, and Michael M McGlue. Lake regionalization and diatom metacommunity structuring in tropical south america. *Ecology and evolution*, 8(16): 7865–7878, 2018.

Yue Bin, Guojun Lin, Sabrina E Russo, Zhongliang Huang, Yong Shen, Honglin Cao, Juyu Lian, and Wanhui Ye. Testing the competition-colonization trade-off and its correlations with functional trait variations among subtropical tree species. *Scientific reports*, 9(1):1–10, 2019.

J. Roger Bray and John T. Curtis. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27(4):325–349, 1957.

James H Brown. On the relationship between abundance and distribution of species. *The american naturalist*, 124 (2):255–279, 1984.

Stephen T Buckland, Angelika C Studeny, Anne E Magurran, Janine B Illian, and Stuart E Newson. The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere*, 2(9):1–15, 2011.

Marc William Cadotte, Donny V Mai, Samuel Jantz, Michael D Collins, Monica Keele, and James A Drake. On testing the competition-colonization trade-off in a multispecies assemblage. *The American Naturalist*, 168(5): 704–709, 2006.

Jonathan M Chase, Alienor Jeliazkov, Emma Ladouceur, and Duarte S Viana. Biodiversity conservation through the lens of metacommunity ecology. *Annals of the New York Academy of Sciences*, 1469(1):86–104, 2020.

Jiquan Chen, Jerry F Franklin, and Thomas A Spies. Vegetation responses to edge environments in old-growth douglas-fir forests. *Ecological applications*, 2(4):387–396, 1992.

Marcela Eugenia da Silva Cáceres, André Aptroot, and Robert Lücking. Lichen fungi in the atlantic rain forest of northeast brazil: the relationship of species richness with habitat diversity and conservation status. *Brazilian Journal of Botany*, 40(1):145–156, 2017.

Jeffrey J Danielson and Dean B Gesch. *Global multi-resolution terrain elevation data 2010 (GMTED2010)*. US Department of the Interior, US Geological Survey, 2011.

Don L De Angelis. Stability and connectance in food web models. *Ecology*, 56(1):238–243, 1975.

Sandra Myrna Díaz, Josef Settele, Eduardo Brondízio, Hien Ngo, Maximilien Guèze, John Agard, Almut Arneth, Patricia Balvanera, Kate Brauman, Stuart Butchart, et al. The global assessment report on biodiversity and ecosystem services: Summary for policy makers. 2019.

Rick Durrett and Simon Levin. Spatial models for species-area curves. *Journal of Theoretical Biology*, 179(2): 119–127, 1996.

Evan P Economo. Biodiversity conservation in metacommunity networks: linking pattern and persistence. *The American Naturalist*, 177(6):E167–E180, 2011.

Lenore Fahrig. Relative effects of habitat loss and fragmentation on population extinction. *The Journal of wildlife management*, pages 603–610, 1997.

Ludmilla Figueiredo, Jochen Krauss, Ingolf Steffan-Dewenter, and Juliano Sarmento Cabral. Understanding extinction debts: spatio–temporal scales, mechanisms and a roadmap for future research. *Ecography*, 42(12): 1973–1990, 2019.

Nicola L Foster, Andrew Foggo, and Kerry L Howell. Using species-area relationships to inform baseline conservation targets for the deep north east atlantic. *PLoS one*, 8(3):e58941, 2013.

K. Ruben Gabriel and Robert R. Sokal. A new statistical approach to geographic variation analysis. *Systematic Zoology*, 18(3):259–278, 1969.

Benjamin Gilbert, Tyler D. Tunney, Kevin S. McCann, John P. DeLong, David A. Vasseur, Van Savage, Jonathan B. Shurin, Anthony I. Dell, Brandon T. Barton, and Christopher DG Harley. A Bioenergetic Framework for the Temperature Dependence of Trophic Interactions. *Ecology Letters*, 17(8):902–914, 2014.

Nick M Haddad, Lars A Brudvig, Jean Clobert, Kendi F Davies, Andrew Gonzalez, Robert D Holt, Thomas E Lovejoy, Joseph O Sexton, Mike P Austin, Cathy D Collins, et al. Habitat fragmentation and its lasting impact on earth's ecosystems. *Science advances*, 1(2):e1500052, 2015.

G Huffman, D Bolvin, D Braithwaite, K Hsu, R Joyce, and P Xie. Integrated multi-satellite retrievals for gpm (imerg), version 4.4. *NASA's Precipitation Processing Center*, 2014. URL `ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/`.

G Hulley and S Hook. Mod21a2 modis/terra land surface temperature/3-band emissivity 8-day l3 global 1km sin grid v006. *NASA EOSDIS Land Processes DAAC*, 2017. doi: https://doi.org/10.5067/MODIS/MOD21A2.006.

Carles Ibáñez, Carles Alcaraz, Nuno Caiola, Albert Rovira, Rosa Trobajo, Miguel Alonso, Concha Duran, Pere J Jiménez, Antoni Munné, and Narcís Prat. Regime shift from phytoplankton to macrophyte dominance in a large river: top-down versus bottom-up effects. *Science of the Total Environment*, 416:314–322, 2012.

W Daniel Kissling, R Field, H Korntheuer, Ursula Heyder, and K Böhning-Gaese. Woody plants and the prediction of climate-change impacts on bird diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1549):2035–2045, 2010.

Mikko Kuussaari, Riccardo Bommarco, Risto K Heikkinen, Aveliina Helm, Jochen Krauss, Regina Lindborg, Erik Öckinger, Meelis Pärtel, Joan Pino, Ferran Rodà, et al. Extinction debt: a challenge for biodiversity conservation. *Trends in ecology & evolution*, 24(10):564–571, 2009.

William F Laurance, Susan G Laurance, Leandro V Ferreira, Judy M Rankin-de Merona, Claude Gascon, and Thomas E Lovejoy. Biomass collapse in amazonian forest fragments. *Science*, 278(5340):1117–1118, 1997.

Douglas J Levey, Benjamin M Bolker, Joshua J Tewksbury, Sarah Sargent, and Nick M Haddad. Effects of landscape corridors on seed dispersal by birds. *Science*, 309(5731):146–148, 2005.

Jonathan M Levine and Mark Rees. Coexistence and relative abundance in annual plant assemblages: the roles of competition and colonization. *The American Naturalist*, 160(4):452–467, 2002.

Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

Romana Limberger and Stephen A Wickham. Competition–colonization trade-offs in a ciliate model community. *Oecologia*, 167(3):723, 2011.

George Livingston, Miguel Matias, Vincent Calcagno, Claire Barbera, Marine Combe, Mathew A Leibold, and Nicolas Mouquet. Competition–colonization dynamics in experimental bacterial metacommunities. *Nature communications*, 3(1):1–8, 2012.

Christopher Robert Margules and Robert L Pressey. Systematic conservation planning. *Nature*, 405(6783):243–253, 2000.

Robert McCredie May. *Stability and Complexity in Model Ecosystems*, volume 6. Princeton University Press, 1973.

José M Montoya and Dave Raffaelli. Climate change, biotic interactions and ecosystem services. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1549):2013–2018, 2010.

Nicolas Mouquet, Birte Matthiessen, Tom Miller, and Andrew Gonzalez. Extinction debt in source-sink metacommunities. *PLoS One*, 6(3):e17567, 2011.

Nicolas Mouquet, Dominique Gravel, François Massol, and Vincent Calcagno. Extending the concept of keystone species to communities and ecosystems. *Ecology Letters*, 16(1):1–8, 2013.

Sean Nee and Robert M May. Dynamics of metapopulations: habitat destruction and competitive coexistence. *Journal of Animal Ecology*, pages 37–40, 1992.

James P O'Dwyer and Jessica L Green. Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecology letters*, 13(1):87–95, 2010.

Jacob D. O'Sullivan, Robert J. Knell, and Axel G. Rossberg. Metacommunity-scale biodiversity regulation and the self-organised emergence of macroecological patterns. *Ecology Letters*, 22(9):1428–1438, 2019. ISSN 1461-0248. doi: 10.1111/ele.13294.

Jacob D. O'Sullivan, J. Christopher D. Terry, and Axel G. Rossberg. Intrinsic ecological dynamics drive biodiversity turnover in model metacommunities. *Nature Communications, in Press*, 2021. doi: 10.1101/2020.05.22.110262.

Michael T Pearce, Atish Agarwala, and Daniel S Fisher. Stabilization of extensive fine-scale diversity by ecologically driven spatiotemporal chaos. *Proceedings of the National Academy of Sciences*, 117(25):14572–14583, 2020.

RL Pressey and AO Nicholls. Efficiency in conservation evaluation: scoring versus iterative approaches. *Biological conservation*, 50(1-4):199–218, 1989.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

David H Reed. Extinction risk in fragmented habitats. *Animal Conservation*, 7(2):181–191, 2004.

Walter V Reid et al. *Millennium ecosystem assessment*. 2005.

Andrew B Reinmann and Lucy R Hutyra. Edge effects enhance carbon uptake and its vulnerability to climate change in temperate broadleaf forests. *Proceedings of the National Academy of Sciences*, 114(1):107–112, 2017.

Emlyn J Resetarits, Sara E Cathey, and Mathew A Leibold. Testing the keystone community concept: effects of landscape, patch removal, and environment on metacommunity structure. *Ecology*, 99(1):57–67, 2018.

Cassandra M Robillard, Laura E Coristine, Rosana N Soares, and Jeremy T Kerr. Facilitating climate-change-induced range shifts across continental land-use barriers. *Conservation Biology*, 29(6):1586–1595, 2015.

James Rosindell and Stephen J Cornell. Species–area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology letters*, 10(7):586–595, 2007.

Axel G. Rossberg. *Food Webs and Biodiversity: Foundations, Models, Data*. John Wiley & Sons, 2013.

Felix Roy, Matthieu Barbier, Giulio Biroli, and Guy Bunin. Complex interactions can create persistent fluctuations in high-diversity ecosystems. *PLoS computational biology*, 16(5):e1007827, 2020.

David Storch. The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. *Journal of Vegetation Science*, 27(5):880–891, 2016.

Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. Conditional variable importance for random forests. *BMC bioinformatics*, 9(1):1–11, 2008.

Trisha L Swift and Susan J Hannon. Critical thresholds associated with habitat loss: a review of the concepts, evidence, and applications. *Biological reviews*, 85(1):35–53, 2010.

Shirin Taheri, David García-Callejas, and Miguel B Araújo. Discriminating climate, land-cover and random effects on species range dynamics. *Global Change Biology*, 27(6):1309–1317, 2021.

Philip D Taylor, Lenore Fahrig, Kringen Henein, and Gray Merriam. Connectivity is a vital element of landscape structure. *Oikos*, pages 571–573, 1993.

John Terborgh, Lawrence Lopez, Percy Nuñez, Madhu Rao, Ghazala Shahabuddin, Gabriela Orihuela, Mailen Riveros, Rafael Ascanio, Greg H Adler, Thomas D Lambert, et al. Ecological meltdown in predator-free forest fragments. *Science*, 294(5548):1923–1926, 2001.

Chris D Thomas, Alison Cameron, Rhys E Green, Michel Bakkenes, Linda J Beaumont, Yvonne C Collingham, Barend FN Erasmus, Marinez Ferreira De Siqueira, Alan Grainger, Lee Hannah, et al. Extinction risk from climate change. *Nature*, 427(6970):145–148, 2004.

Samuel ED Thompson, Ryan A Chisholm, and James Rosindell. Characterising extinction debt following habitat fragmentation using neutral theory. *Ecology letters*, 22(12):2087–2096, 2019.

David Tilman, Robert M May, Clarence L Lehman, and Martin A Nowak. Habitat destruction and the extinction debt. *Nature*, 371(6492):65–66, 1994.

Laura Toloşi and Thomas Lengauer. Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*, 27(14):1986–1994, 2011.

Kerrie A Wilson, Josie Carwardine, and Hugh P Possingham. Setting conservation priorities. *Annals of the New York Academy of Sciences*, 1162(1):237–264, 2009.

Xian Yang, Jiaqi Tan, Kevin Harry Sun, and Lin Jiang. Experimental demonstration of the importance of keystone communities for maintaining metacommunity biodiversity and ecosystem functioning. *Oecologia*, 193(2):437–447, 2020.