Temporally robust spatial structure in ecosystems explained by local biodiversity regulation

Running title: Temporally robust spatial structure in ecosystems

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

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

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

Keywords: biodiversity — macroecology — spatial ecology — metacommunity — occupancy frequency distribution

Competing interests: The authors declare no competing interests.

Data/code availability: Data and R scripts needed to reproduce results are available at `https://github.com/jacobosullivan/EA-OFD`

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

The distribution in species' site occupancy is a fundamental pattern in spatial ecology. Despite decades of research, the mechanisms responsible for the shape of the distribution remain incompletely understood. Here, we simultaneously examining both spatial and temporal patterns of site occupancy in communities of macroinvertebrates, macrophytes and diatoms. We show that a simple patch dynamics model explains both the spatial structure and temporal dynamics in English rivers. The key mechanism responsible for the emergent spatial structure in the model are intrinsic regulation of biodiversity at the local scale which, when combined with inter-site dispersal and regional-scale invasion, drives local colonisation and extinction. Our analysis supports the notion that metacommunities exist in a kind of dynamic steady state arising from local ecological constraints and that this intrinsic regulation can drives the emergence of a wide variety of macroecological patterns.

1 Introduction

The question of what determines commonness and rarity in ecological assemblages is nothing if not enduring (Preston, 1948). The question of commonness and rarity *in space*, represented by the occupancy frequency distribution (OFD), has been engaging ecologists for even longer (Raunkiaer et al., 1934). Yet despite many decades of research, the drivers of these fundamental ecological distributions remains incompletely understood.

In most ecological assemblages, the majority of species are observed in a minority of sites. Put another way, the OFD of well sampled assemblages is typically found to be right skewed (e.g. Tokeshi 1992; McGeoch and Gaston 2002; van Rensburg et al. 2000; Bossuyt et al. 2004; Heatherly et al. 2007; Astorga et al. 2014; Heino 2015; Boieiro et al. 2018; Renner et al. 2020). A variety of biological processes have been posited to impact the shape of the OFD (McGeoch and Gaston, 2002). Species spatial distributions may be predominantly determined by interspecific variation in environmental niche widths, with habitat specialists dominating low occupancy classes and higher occupancy classes filled by generalists (Brown, 1984; Brown et al., 1995). Indeed, a positive relationship between niche breadth and species' total range size—generally assumed to correlate non-trivially with occupancy patterns—has been reported across taxonomic groups and spatial scales (Slatyer et al., 2013). Likewise, the landscape scale occupancy of environmental generalists can exceed that of specialist competitors (Denelle et al., 2020). It is important to note, however, that niche width and range size estimates are frequently made by inspection of the same data (Carscadden et al., 2020) and as such, the association may frequently be tautological.

An apparent relationship between niche breadth and site occupancy cannot, however, sufficiently explain the shape of the OFD: it must also be the case that narrow niche widths are substantially over-represented in nature. By pooling data at the global scale Belmaker et al. (2012) report a numerical dominance of diet and habitat specialists in birds. However, given the apparent relationship between niche breadth and latitude (Stevens, 1989; Pagel et al., 1991; Ruggiero, 1994), it is likely that the frequency of narrow niche width species is exaggerated relative to single assemblages. Weak right skew in estimated environmental tolerance range has been found in riverine macroinvertebrates (Mykrä and Heino, 2017). However, this contrasts with a symmetric distribution in niche breadths in freshwater macrophytes (Kennedy et al., 2017) and the apparent over-representation of generalist species within a regional-scale plant assemblage (Boulangeat et al., 2012). Consistent evidence of strong bias in niche widths, sufficient to explain the frequently observed right-skewed OFD, has not been found.

Interspecific variation in dispersal traits could also impact the shape of the OFD, with high occupancy classes filled by the best dispersers (Collins and Glenn, 1997). While evidence of a mechanistic link between dispersal

ability and species' total range size is inconclusive (Sheard et al., 2020), positive relationships have been shown (Capurucho et al. 2020; Böhning-Gaese et al. 2006; Luo et al. 2019; Laube et al. 2013; Sheard et al. 2020 but see Lester et al. 2007). Right-skew, sometimes strong, has been found in the distribution of proxies for dispersal ability or estimated dispersal rates (Schloss et al., 2012; Tamme et al., 2014; Sarremejane et al., 2020; Sheard et al., 2020). However, commonly used proxies for dispersal ability based on body- or propagule-size measurements may also be correlates of other ecologically relevant capacities (Beckman et al., 2018). Furthermore, difference in typical dispersal abilities between biomes (Chen et al., 2019; Álvarez-Noriega et al., 2020; McCaslin et al., 2020) mean that global-scale trait databases are again likely to miss-represent landscape-scale distributions.

Crucially, a uniquely phenotypic explanation for the shape of the OFD would imply species tend to occupy roughly consistent numbers of sites in time in the absence of strong environmental change. Species-level site occupancy in contemporary England, for example, is extremely variable in time (see below). Classic patch dynamics type models, which describe stochastic population-level fluctuations in occupancy, have been proposed to explain the shape of the OFD. In this case, local extinction and dispersal limitation, which delays re-colonisation, mean species tend to occupy a minority of the available habitat. This can lead to a right skewed OFD, but requires the assumption that extinction rates decrease with overall occupancy (Hanski, 1982), which may not be empirically justified (Wilfahrt et al., 2021) and is arguably mechanistically circular.

Here we show how the OFD in both time and space can be explained in a straight-forward way by representing an ecological landscape as a dynamical system with the intrinsic constraints that are a well known property of such systems (May, 1972). Using a highly replicated metacommunity dataset describing the distribution of riverine taxa throughout England, we show how the OFD is both consistently right-skewed and temporally robust despite turnover in local community composition and species rank occupancy though time. These observations are readily explained by a simple metacommunity model incorporating only local limits on species richness—the outcome of an implicitly modelled network of ecological interactions—stochastic range expansion and immigration into the assemblage, but without interspecific differences in niche width, dispersal ability, or local extinction probability.

2 Methods

2.1 Compiling the database of temporal OFDs

Taking data from the Environment Agency (EA) river surveys[1] covering the whole of England (Fig. 1), we

[1] https://environment.data.gov.uk/ecology-fish/downloads/

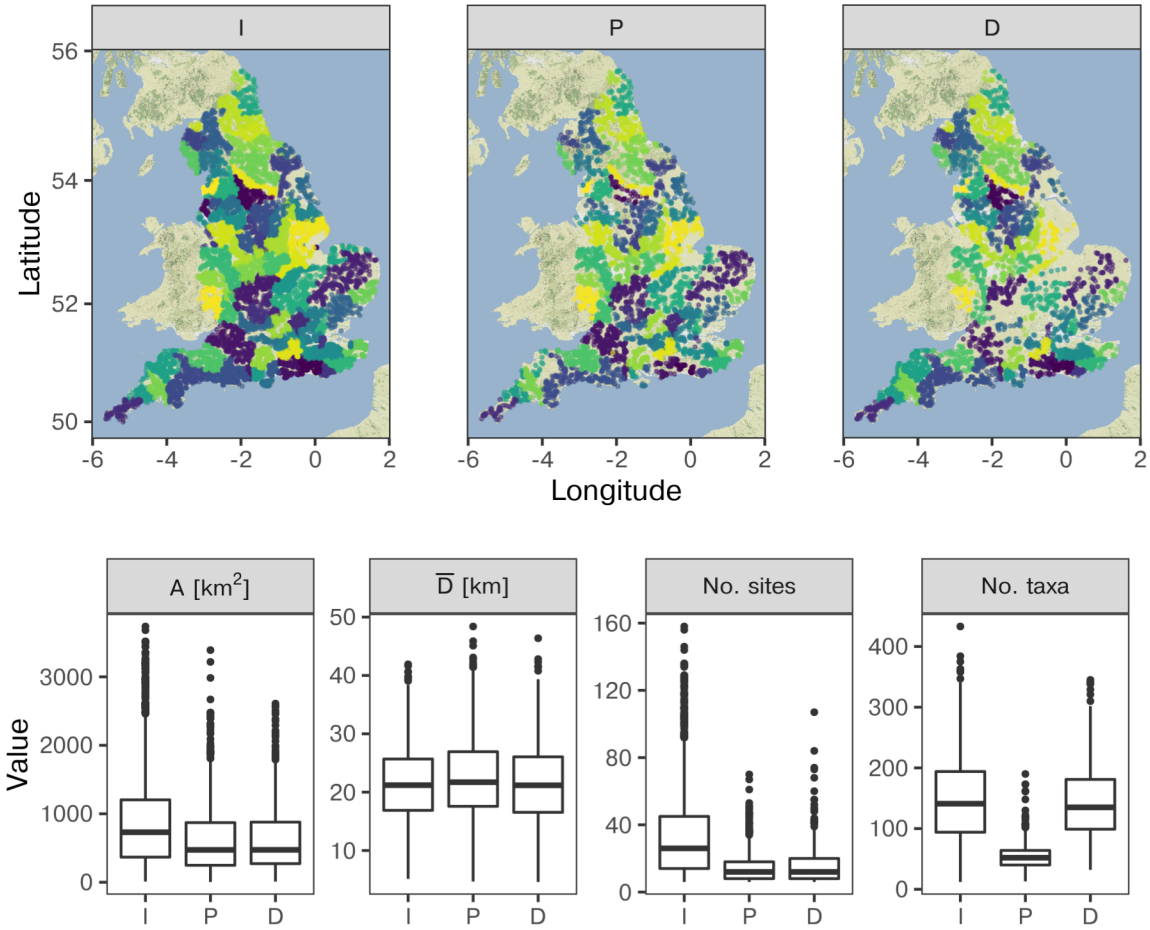

Figure 1: **English riverine community surveys.** Top row: The locations of sites in the three nationwide surveys. Points correspond to sites, colours to the management catchment. Bottom row: The area of the convex hull bounding sites, mean inter-site distance, number of sites, and number of taxa observed in individual catchment years for each dataset. In all cases, figures represent the datasets after filtering as described in the text.

compiled a database of observations of macroinvertebrates (I), macrophytes (P) and diatoms (D) identified to species or genus level. Taxa identified to genus level made up 31%, 30%, and 11% of the I, P and D datasets respectively, thus while our analysis posits local biodiversity regulation via interactions between species as a key organising mechanism, limits to taxonomic resolution must be taken into account.

The EA divides freshwater ecosystems in England into 10 River Basin Districts (RBDs), each of which contains between 2 and 20 Management Catchments (MCs) (average 11.7 per RBD). Waterbody MC allocation was

Box 1: The steady state OFD of the Locally Saturated Patch Dynamics model

The Locally Saturated Patch Dynamics model: The spatially implicit metacommunty is represented by a species by site occupancy matrix describing an assemblage of S species occupying N identical patches, each capable of hosting n populations. At each time step of the model a new species invades the metacommunity from outside. Its initial site occupancy is set to 1, i.e. each new species colonises a single site in the metacommunity. The fact that the initial occupancy is fixed for all invaders ensures the neutrality of the model set up. Simultaneously, each extant local population colonises an additional site with a probability $m/(nN)$, where m is the mixing rate (see text for more details). Each colonisation event, either from within or from outside the metacommunity, produces a corresponding extirpation event of a randomly selected occupant of the colonised site in accordance with the assumption of local ecological saturation (Rossberg, 2013; O'Sullivan et al., 2019). In Fig. 2 we show the average occupancy frequency of the end state of 1000 metacommunity models assembled according to these basic principles. The dashed line represents the predicted OFD after fitting the mixing rate (with some error in the estimate). R script used to generate the simulation and for the fitting procedure are available in the accompanying github repository.

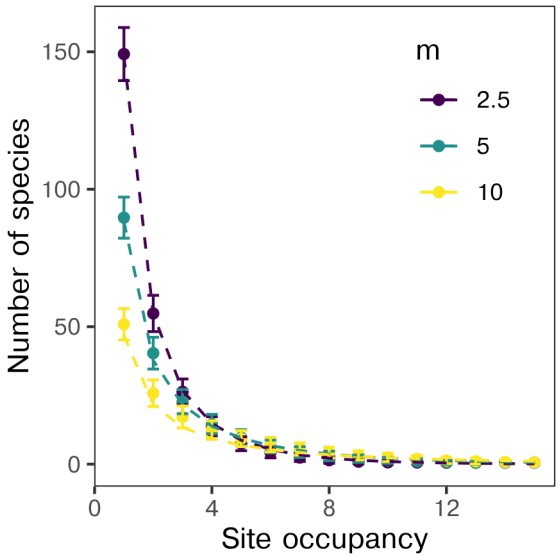

Figure 2: The average site occupancy in 1000 realisations of the probabilistic LSPDM following 1e4 simulated invasions. Here $N = 50$, $n = 10$ and the mixing rate was set to $m = 2.5$, 5, and 10. The dashed lines represent the fitted curves for which m was estimated as 2.8, 5.9 and 12.5 respectively, giving an indication of the measurement error of the fitting procedure.

extracted from the EA Catchment data explorer[2]. MCs are an arbitrary classification based on internal EA decision making. We assume that the allocation of water bodies to MCs involved some analysis of habitat type and river basin structure, and that therefore these catchments can be considered as distinct metacommunities. This is, of course, an approximation.

[2]https://environment.data.gov.uk/catchment-planning/

Box 2: Estimating metacommunity-scale turnover rate

In order to study the mechanisms driving the distribution of species' lifetime, Pigolotti et al. (2005) derive the following equation for the probability of a randomly selected species detected at time $t0 = 0$, persisting until time t for simple neutral models:

$$G(0,t) = \frac{1}{\log(1-\alpha)} \log\left[\frac{(1-)\exp(1-\alpha)\,dt}{\exp[(1-\alpha)\,dt]-\alpha}\right]. \tag{1}$$

Note that we rescaled t to dt to match the time scales of the LSPDM model. Here we translate this result to the patch dynamics metacommunity case such that it describes the probability of persistence of a local population. The parameter d is estimated from the rate of decay in local community similarity, while, after solving for α equation 3 can be fitted to the observed turnover at the catchment scale. Further details including R code for estimating α from metacommunity time-series are included in the Supplementary information and supplementary information.

To account for the fact that the number of sites and their location was not necessarily fixed through time for a given catchment, we focus on the proportional site occupancy distribution. We deal with more pathological variation in sampling effort by applying a suite of filters to the data (summarised in the Supplementary information, Fig. S1). Following filtering, we ended up 83, 52 and 48 catchments respectively sampled over a period of around 30 years leaving 1804, 731 and 588 unique OFDs. The area of the convex hull bounding the sites, the mean distance between sites, the number sites surveyed and the number of taxa detected in independent catchment-years after filtering are shown in Fig. 1. All data and R scripts to run the analyses are available in the accompanying github repository.

2.2 Testing for bimodality in the OFDs

We tested for bimodality in the OFD for each catchment-year using Tokeshi's (1992) method in which the probability of obtaining the observed number of species in the left- and right-most occupancy classes is computed assuming as a null hypothesis a uniform OFD. Significant bimodality was detected in 4%, 15% and 26% of the catchment-years for the I, P and D datasets respectively prior to applying the filtering procedure described above. After filtering, however, the percentages dropped to 0.2% 0.1% and 6% respectively (Fig. S2). This suggests that bimodality in the OFD is predominantly a result of limited sampling rather than a true biological signal. Substantial heterogeneity in the overall shape of the OFDs would invalidate the comparative analysis of the shape of the OFDs described below. Since the majority of OFDs in the dataset were demonstrably unimodal the first three moments of the distributions could be directly compared between catchments and years.

2.3 Analysing the temporal dynamics of the empirical OFDs

We analysed the temporal trends in the mean, variance and skewness of the OFD for each catchment. The existence of a significant temporal trend was tested for using the Mann-Kendall test (Mann, 1945; Kendall, 1948) and the slope of the trend assessed using Sen's estimator (Theil, 1950; Sen, 1968). We also tested for stationarity of the time-series using Augmented Dickey Fuller (Fuller, 2009) and KPSS tests (Kwiatkowski et al., 1992). The ADF test attempts to fit an autoregressive model of arbitrary order to a time-series and tests for the presence of a unit root, indicative of non-stationarity. In the ADF, the presence of a unit root (non-stationarity) is the null hypothesis. The KPSS is analogous, however the absence of a unit root (stationarity) null hypothesis.

We then applied the same time-series analyses to two measures of temporal turnover. The first was the average change in Jaccard similarity (Jaccard, 1912) in catchment composition (i.e. summing over all sites) over time. This is a standard measure of the difference between binary presence-absence vectors describing species composition in communities. The second was the average change in Bray-Curtis similarity (Bray and Curtis, 1957) in the (non-binary) vectors describing species' site occupancy over time. In both cases these were studied as a function of time lag irrespective of directionality (Fig. S3).

2.4 A locally saturated patch dynamics model

We model the OFD at the catchment scale using a locally saturated patch dynamics model (LSPDM). The LSPDM is an extension of the unified neutral theory of biogeography (UNTB; Hubbell 2001) applied at a higher organisational scale. The UNTB describes the temporal dynamics in the number of individuals in a multi-species assemblage under the assumption that the total number of individuals is strictly constrained. By contrast, here we model the temporal dynamics of site occupancy, with the constraint that the number of species at each site is strictly limited. This intrinsic limit to local diversity is understood as the outcome of an implicitly modelled network of ecological interactions. The translation of the neutral theoretical approach to a patch dynamics (PD) framework allows us to explore the impact of local diversity regulation in PD metacommunities using the mathematics already developed for the study of assemblages of individuals.

The steady state of the LSPDM model is a unimodal, right skewed OFD which remains robust in time despite continued invasion into the assemblage and between sites (Box 1). The shape of this OFD is controlled by a single parameter, which we term the 'mixing rate' m, that can be understood as the internal colonisation rate—the rate at which local populations colonise additional sites within the metacommunity. Since we assume local communities

are saturated, this equivalently describes the component of the local extinction rate caused by local colonisation events. The LSPDM is an open system subject to invasion from outside with invasion rate λ. Again due to the assumption of local saturation, λ equivalently describes the rate of local extirpation due to regional invasion.

The proportion of colonisation events due to species range expansion, as opposed to regional invasion, is

$$\frac{m}{m+\lambda} = \frac{b}{d} = \alpha. \tag{2}$$

In the individual based UNTB-type models, b and d are individual birth and death rates. In the translated LSPDM, b is equivalent to m, the rate at which new *populations* are 'born'. Similarly, when translated to the PD scale, d can be understood as the rate of extirpation, which arises due to both internal colonisation and arrival into the network from outside such that $d = \lambda + b$. For convenience we set $\lambda = 1$, thereby modelling patch dynamics in units of average regional invasion times.

Since the emergent OFD of the LSPDM is controlled by a single parameter, m, it can be readily fitted to observed OFDs. A detailed description of the LSPDM model, its steady state solution and the procedure for fitting it to data are given in the Supplementary information. R code to perform the fitting procedure is available in the accompanying github repository.

An alternative method to estimating α, the ratio b/d, was developed by Pigolotti et al. (2005). There, the authors show how the probability of detecting a given occupant species decays in time since the initial observation as a function of the demographic rates b and d (Box 2). These results can be directly translated into the metacommunity case as follows: d is estimated as the average rate of exponential decay in species persistence at the local scale. This is essentially a measure of the rate of temporal turnover. Having estimated d from the local turnover observed in the dataset and solving for α, Eqs. (2) and (1) give an alternative estimate of m based not on the spatial structure of the catchment, but instead based on analysis of the community time-series. For each of the catchments by fitting the LSPDM to the observed OFD, we estimated the mixing rate m_{OFD}, and the corresponding α_{OFD}. From analysis of the metacommunity time-series and using Eq. (1), we compute alternative estimates which we denote m_{TS} and α_{TS}. We then ask whether the m and α estimated using the spatial structure only (fitting to the LSPDM) quantitatively matches those estimated by reference to the observed rate of turnover. So doing we test whether the LSPDM captures both the spatial and temporal rates of ecological turnover. This is an essential step in the analysis since it is possible to fit an vast array of different models to a skewed distribution (McGill et al., 2007). We probe the biological realism of the LSPDM representation by testing its ability to reproduce both the spatial and temporal

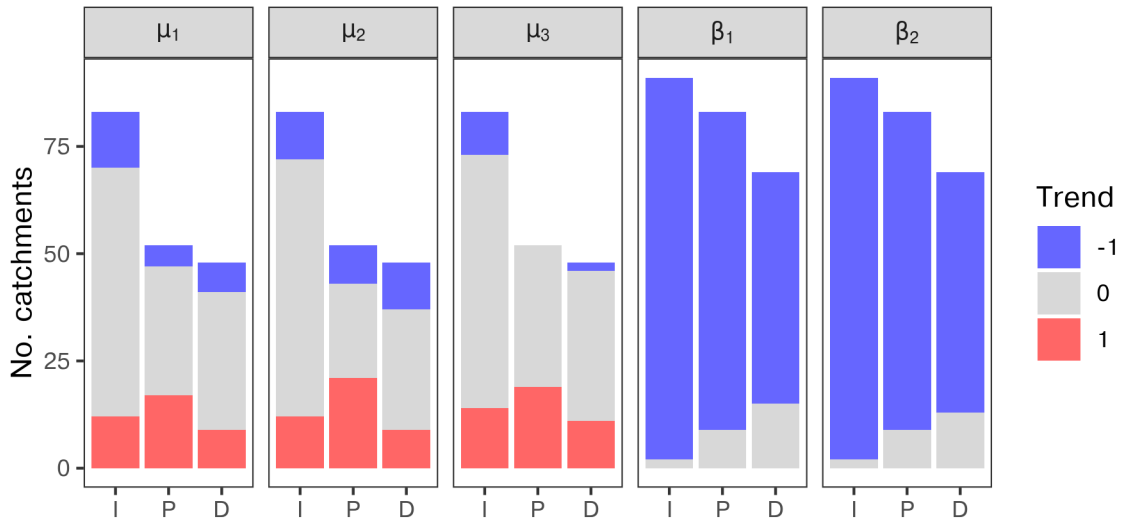

Figure 3: **Temporal trends in metacommunity spatial structure (OFD) and composition.** The sign and statistical significance of the temporal trend in mean (μ_1), variance (μ_2), skewness (μ_3) of the catchment-scale OFDs and in the compositional similarity (Jaccard, β_1) and site occupancy similarity (Bray-Curtis, β_2) of the three surveys. Colour represents the sign of the Sen's trend, with grey indicating no trend detected using Mann-Kendall.

structure of the empirical observation.

3 Results

The proportion of sites occupied by species varied over the time-series. This species-level fluctuation in site occupancy is sufficient to drive substantial change in the species rank occupancy, frequently including the loss of species from the record (Fig. S4). The OFD at the catchment scale was remarkably consistent–almost all catchments had unimodal right-skewed OFDs for all three datasets (Figs. S7-S9).

We tested for time invariance of the OFD by studying the time-series of the first three moments of the distribution, its mean, variance, and skewness. In Fig. S5 we show the linear trend of the skewness estimates over time for each of the high level taxa. We find in the majority of the catchments studied and for each of the large-scale taxa that no significant temporal trend in the moment expansion is the norm in these data. There are catchments for which both positive and negative trends were detected, and the number of significant trends is somewhat more than would be expected by random chance, however these are exceptional cases nonetheless. In contrast, we find that almost all the catchments show significant turnover in composition (presence-absence) as well as in species'

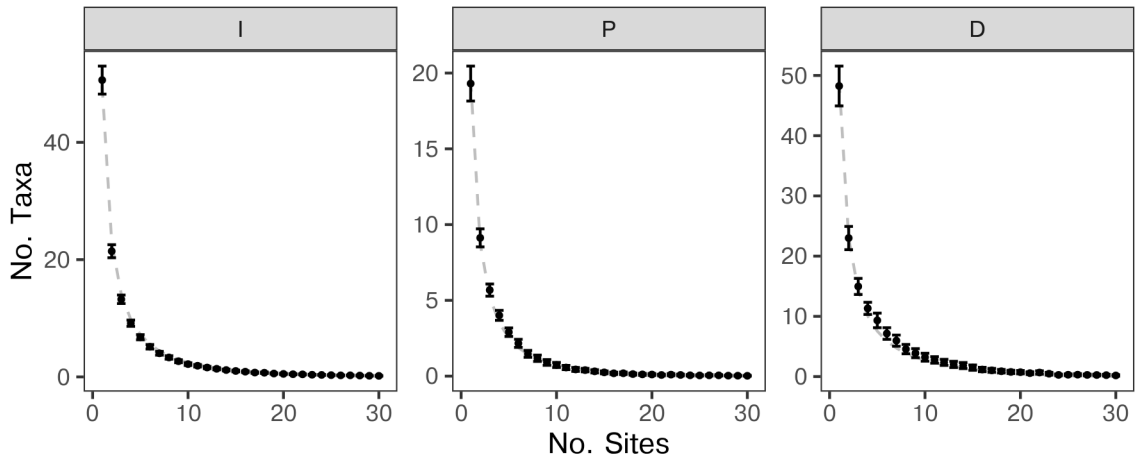

Figure 4: **Fitting the LSPDM to the observed OFDs.** Catchment-time-average site occupancy distributions for the three large taxonomic groupings with the fitted LSPDM. Following standard propagation of uncertainty procedure, error bars represent $e(N) = (1/n)\sqrt{\sum_n(\Delta SO_n(N)^2)}$ where n is a catchment index, N a given site occupancy class, and $\Delta SO_n(N)$ is the standard deviation of the number of species in the Nth occupancy class for a given catchment over the time-series.

site occupancy (Fig. 3).

We tested for stationarity in the time-series of the moment expansion using the ADF and KPSS tests (Fig. S6). For the majority of catchments and in all three datasets, using ADF we find that the time-series of the mean, variance and skewness of the catchment ODFs were non-stationary ($p > 0.05$), despite an strong tendency to detect no temporal trend (Fig. 3). In contrast, using the KPSS test we found that stationarity cannot be ruled out in almost all cases ($p > 0.05$). Thus, we cannot conclusively confirm whether the typically unbiased fluctuations in the shape of the OFD can in fact be modelled as stationary processes.

While not demonstrably stationary, given the spatio-temporal consistency in the shape of the OFD, we chose to fit the LSPD model to the space-time-averaged OFD for each dataset (Fig. 4). The lines in Fig. 4 represent the equilibrium OFD of the LSPDM with the mixing rate parameter m fitted to the observation. The LSPDM accurately reproduces the spatial structure of the observation.

By fitting the LSPDM to the time-averaged OFD for each catchments we compute estimates for the internal colonisation rate m_{OFD}. These ranged across multiple orders of magnitude but had a mode of around 30 for the I and D datasets, and slightly lower for the P dataset (Fig. 5). The corresponding α_{OFD} differed only slightly from 1 implying invasion into the catchment is rare compared to internal colonisation. Computing m_{TS} and α_{TS} by fitting Eq. (1) to the catchment timeseries, we find comparable distributions in the estimates (Fig. 5).The

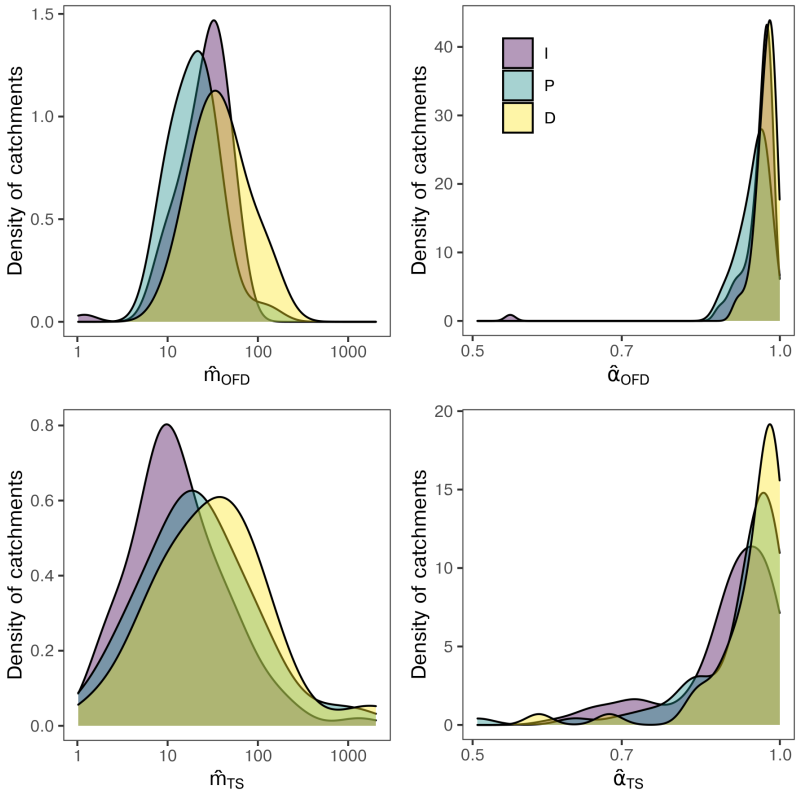

Figure 5: **Comparing time scales estimated from spatial and temporal analyses.** The mixing rates, m, and the α, estimated by analysis of the OFD or alternatively, the observed rate of turnover. Top row: estimates based on the time-averaged spatial structure of each catchment only. Bottom row: estimated by analysis of the observed turnover in species composition.

ratio between the median estimates computed via the spatial and temporal analyses $\hat{m}_{OFD}/\hat{m}_{TS}$ was in the range 1.2-2.7 depending on taxon, while $\hat{\alpha}_{OFD}/\hat{\alpha}_{TS}$ ranged from 1.0-1.1. The non-parametric Mann-Whitney U test (Mann, 1945) detected a significant difference between the distribution of the estimates for the macroinvertebrates ($p < 0.05$), but not for either the macrophytes or the diatoms (Tab. 1).

The differences in the estimated local colonisation rate based on the pure spatial and pure temporal analyses are small: the spatial analysis correctly inferred the colonisation rate to within an order of magnitude. Thus we find that the LSPDM accurately represents the observed spatial structure (Figs. 2, S7-S9) but crucially that the demographic rates implied by this spatial structure are broadly consistent with the local and metacommunity scale turnover, differing by at most a factor of 2.7 on average (Fig. 5).

Finally, to probe the applicability of these results to spatial scales greater than catchments, we ask how representative the proportional occupancy of a given taxon is of the corresponding proportional occupancy at the national

	$\hat{m}_{OFD}/\hat{m}_{TS}$	MW p-value	$\hat{\alpha}_{OFD}/\hat{\alpha}_{TS}$	MW p-value
I	2.67	< 0.05	1.05	< 0.05
P	1.16	> 0.05	1.01	> 0.05
D	1.21	> 0.05	1.01	> 0.05

Table 1: The ratio of the median m and α computed by reference to the time-averaged OFD (using the LSPDM model) to the equivalent quantities estimated by analysis of the time-series (Pigolotti et al., 2005). The significance of the differences in the distribution of estimated parameters was tested using the Mann-Whitney U test.

scale. This is a fairly weak test of scale (in)dependence, however the lack of replication at the national scale makes more rigorous analysis impossible. We find that species' catchment-average proportional occupancy is typically correlated to their national proportional occupancy in a given year (Pearson's $r = 0.94$, 0.93 and 0.96, $p \ll 0.05$, for I, P and D respectively, Fig S10). Further, we find no obvious difference between the catchment-national-occupancy relation for taxa identified to species or genus level, suggesting this association is weakly impacted by taxonomic resolution. The correlation between catchment and national scale, despite substantial variation between years in species' national occupancy (NEED TO SHOW?), suggests analogous organising processes may help to explain spatial distributions at the national scale.

4 Discussion

In recent years, evidence has been found of intrinsic biodiversity regulation, detectable as temporal stationarity in local community species richness (e.g. Brown et al. 2000; Parody et al. 2001; Magurran and Henderson 2010; Magurran et al. 2015). No systematic change in local richness has been found at the global scale (Dornelas et al., 2014; Gotelli et al., 2017; Vellend et al., 2017), though latitudinal and between-biome differences in trend have been found (Blowes et al., 2019). This temporal robustness can also be observed in higher-order community properties such as the distribution of abundances between species, despite substantial changes in the identities of species in each abundance class (Magurran and Henderson, 2010). Here we build upon this work by showing that such intrinsic constraints propagate up to the metacommunity scale and ultimately determine macroecological structure represented in the OFD. Furthermore, we have shown that these instrinsic constraints are readily amenable to analysis using simple dynamical models.

An important preliminary result of this analysis is the demonstration that bimodality is extremely rare in the OFDs for these taxa and that when demonstrably under-sampled catchment-years are removed from the dataset, the proportional incidence of bimodality goes down substantially. Bimodality in the OFD is likely, therefore, to reflect

sampling effort and not underlying biology (van Rensburg et al., 2000; Bossuyt et al., 2004; Heatherly et al., 2007). Recently other studies have reported temporal consistency in the qualitative shape of bimodal OFDs (Suhonen and Jokimäki, 2019; Suhonen, 2021)). Our study suggest such results may reflect not time invariance in the underlying biology, but instead in the sampling regime applied.

We tested for stationarity in the first thee moments of the catchment ODFs, however the results were inconclusive, probably due to the length of the time-series. Stationarity has previously been used to evidence intrinsic regulation of species richness in communities (Gotelli et al., 2017). While the same exact test of regulation cannot be applied to these data, the fact that the OFD is a distribution rather than a scalar property, and the fact that this distribution is consistently strongly biased, should be seen as *prima facie* evidence of some intrinsic regulation akin to an entropic constraint.

The distribution of abundance between species in single communities is often considered to reflect the underlying network of ecological interactions in some sense (McGill et al., 2007). Regional occupancy has repeatedly been shown to correlate with average local density (Brown, 1984; Gaston and Blackburn, 2000; Freckleton et al., 2005; Blackburn et al., 2006; Zuckerberg et al., 2009), though exceptions have also been found (Gaston, 1996; Blackburn et al., 2006; Reif et al., 2006; Symonds and Johnson, 2006; Novosolov et al., 2017; Bijleveld et al., 2018). It is important to note, however, that the OFD and the species abundance distribution are not functionally equivalent phenomena. In addition to the niche width and dispersal length heterogeneities described above, trade-offs between competitive and dispersive abilities can plausibly lead to locally inferior competitors with high regional occupancy though empirical evidence for such a trade-off varies (Levine and Rees, 2002; Cadotte et al., 2006; Limberger and Wickham, 2011; Livingston et al., 2012; Bin et al., 2019). Thus, the intrinsic regulation of the OFD via local ecological saturation should be understood as a distinct macroecological phenomenon.

The model used to explain the shape of the OFD is essentially the simplest possible dynamical-systems representation of a multi-species, multi-site assemblage subject to local biodiversity regulation. The LSPDM builds upon previous work on a related model, the Lotka-Volterra Metacommunity Model (LVMCM). The LVMCM is a simulation based approach that has been shown to reproduce a variety of empirically observed patterns in macroecology (O'Sullivan et al., 2019, 2021). However, at present the our understanding of the precise relationships between local processes and regional patterns remains predominantly experimental and for limited to specific combinations of parameters. Furthermore, the complexity of the parameter space of the LVMCM makes directly fitting the model to empirical data impossible; instead a qualitative pattern matching procedure must be used to test the biological realism of its emergent properties. Key emergent properties of the LVMCM can, however, be implicitly

integrated into the LSPDM.

In the LVMCM local species richness is intrinsically regulated due to the onset of ecological structural instability at the local scale: once this condition is reached, each new local colonisation event produced on average one local extirpation (Rossberg, 2013). The LVMCM has also highlighted how processes entirely intrinsic to the metacommunity can lead to local colonisation-extinction dynamics (O'Sullivan et al., 2021) in a way that is closely analogous to the stochastic colonisations and extinctions that are incorporated in PD type models models. Thus, the LSPDM should be seen as an empirically falsifiable abstraction of the more complex LVMCM.

We demonstrate the efficacy of this representation by fitting the model to the spatial structure of the time averaged OFD. Crucially, the LSPDM seems to capture both the spatial structure of the empirical metacommunities and the key demographic rates the spatial structure reflects. The simple LSPDM model can, therefore, be shown to reproduce both the spatial *and* temporal extensions of ecological heterogeneity in the empirical ecosystems. Thus, we conclude that steady state dynamics in species site occupancy derived from local ecological limits play a crucial role in determining the macroecological structure of natural ecosystems.

References

Mariana Álvarez-Noriega, Scott C Burgess, James E Byers, James M Pringle, John P Wares, and Dustin J Marshall. Global biogeography of marine dispersal potential. *Nature Ecology & Evolution*, 4(9):1196–1203, 2020.

Anna Astorga, Russell Death, Fiona Death, Riku Paavola, Manas Chakraborty, and Timo Muotka. Habitat heterogeneity drives the geographical distribution of beta diversity: the case of new zealand stream invertebrates. *Ecology and Evolution*, 4(13):2693–2702, 2014.

Noelle G Beckman, James M Bullock, and Roberto Salguero-Gómez. High dispersal ability is related to fast life-history strategies. *Journal of Ecology*, 106(4):1349–1362, 2018.

Jonathan Belmaker, Cagan H Sekercioglu, and Walter Jetz. Global patterns of specialization and coexistence in bird assemblages. *Journal of Biogeography*, 39(1):193–203, 2012.

Allert I Bijleveld, Tanya J Compton, Lise Klunder, Sander Holthuijsen, Job Ten Horn, Anita Koolhaas, Anne Dekinga, Jaap van der Meer, and Henk W van der Veer. Presence-absence of marine macrozoobenthos does not generally predict abundance and biomass. *Scientific reports*, 8(1):1–12, 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.

Tim M Blackburn, Phillip Cassey, and Kevin J Gaston. Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution. *Journal of Animal Ecology*, 75(6): 1426–1439, 2006.

Shane A Blowes, Sarah R Supp, Laura H Antão, Amanda Bates, Helge Bruelheide, Jonathan M Chase, Faye Moyes, Anne Magurran, Brian McGill, Isla H Myers-Smith, et al. The geography of biodiversity change in marine and terrestrial assemblages. *Science*, 366(6463):339–345, 2019.

Katrin Böhning-Gaese, Tanja Caprano, Karin van Ewijk, and Michael Veith. Range size: disentangling current traits and phylogenetic and biogeographic factors. *The American Naturalist*, 167(4):555–567, 2006.

M Boieiro, TJ Matthews, C Rego, L Crespo, CA Aguiar, P Cardoso, F Rigal, I Silva, F Pereira, PA Borges, and AR Serrano. A comparative analysis of terrestrial arthropod assemblages from a relict forest unveils historical extinctions and colonization differences between two oceanic islands. *PloS one*, 13(4):e0195492, 2018.

B Bossuyt, O Honnay, and M Hermy. Scale-dependent frequency distributions of plant species in dune slacks: Dispersal and niche limitation. *Journal of Vegetation Science*, 15(3):323–330, 2004.

Isabelle Boulangeat, Sébastien Lavergne, Jérémie Van Es, Luc Garraud, and Wilfried Thuiller. Niche breadth, rarity and ecological characteristics within a regional flora spanning large environmental gradients. *Journal of Biogeography*, 39(1):204–214, 2012.

J Roger Bray and John T Curtis. An ordination of the upland forest communities of southern wisconsin. *Ecological monographs*, 27(4):326–349, 1957.

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

James H Brown, David W Mehlman, and George C Stevens. Spatial variation in abundance. *Ecology*, 76(7): 2028–2043, 1995.

James H Brown, SK Morgan Ernest, Jennifer M Parody, and John P Haskell. Regulation of diversity: maintenance of species richness in changing environments. *Oecologia*, 126(3):321–332, 2000.

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.

João MG Capurucho, Mary V Ashley, Brian R Tsuru, Jacob C Cooper, and John M Bates. Dispersal ability correlates with range size in amazonian habitat-restricted birds. *Proceedings of the Royal Society B*, 287(1939): 20201450, 2020.

Kelly A Carscadden, Nancy C Emery, Carlos A Arnillas, Marc W Cadotte, Michelle E Afkhami, Dominique Gravel, Stuart W Livingstone, and John J Wiens. Niche breadth: causes and consequences for ecology, evolution, and conservation. *The Quarterly Review of Biology*, 95(3):179–214, 2020.

Si-Chong Chen, Riin Tamme, Fiona J Thomson, and Angela T Moles. Seeds tend to disperse further in the tropics. *Ecology letters*, 22(6):954–961, 2019.

Scott L Collins and Susan M Glenn. Effects of organismal and distance scaling on analysis of species distribution and abundance. *Ecological applications*, 7(2):543–551, 1997.

William Jay Conover. *Practical nonparametric statistics*, volume 350. john wiley & sons, 1999.

Pierre Denelle, Cyrille Violle, DivGrass Consortium, and François Munoz. Generalist plants are more competitive and more functionally similar to each other than specialist plants: insights from network analyses. *Journal of Biogeography*, 47(9):1922–1933, 2020.

Maria Dornelas, Nicholas J Gotelli, Brian McGill, Hideyasu Shimadzu, Faye Moyes, Caya Sievers, and Anne E Magurran. Assemblage time series reveal biodiversity change but not systematic loss. *Science*, 344(6181): 296–299, 2014.

RP Freckleton, JA Gill, D Noble, and AR Watkinson. Large-scale population dynamics, abundance–occupancy relationships and the scaling from local to regional population size. *Journal of Animal Ecology*, 74(2):353–364, 2005.

Wayne A Fuller. *Introduction to statistical time series*, volume 428. John Wiley & Sons, 2009.

Kevin J Gaston. The multiple forms of the interspecific abundance-distribution relationship. *Oikos*, pages 211–220, 1996.

Kevin J Gaston and Tim M Blackburn. *Pattern and process in macroecology*, volume 414. Wiley Online Library, 2000.

Nicholas J Gotelli, Hideyasu Shimadzu, Maria Dornelas, Brian McGill, Faye Moyes, and Anne E Magurran. Community-level regulation of temporal trends in biodiversity. *Science advances*, 3(7):e1700315, 2017.

Ilkka Hanski. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos*, pages 210–221, 1982.

T Heatherly, MR Whiles, DJ Gibson, SL Collins, AD Huryn, JK Jackson, and MA Palmer. Stream insect occupancy-frequency patterns and metapopulation structure. *Oecologia*, 151(2):313–321, 2007.

Jani Heino. Deconstructing occupancy frequency distributions in stream insects: effects of body size and niche characteristics in different geographical regions. *Ecological Entomology*, 40(5):491–499, 2015.

SP Hubbell. *The Unified Neutral Theory of Biodiversity and Biogeography (Prin*. Princeton University Press, Princeton, NJ, 2001.

Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.

Maurice George Kendall. Rank correlation methods. 1948.

Michael P Kennedy, Pauline Lang, Julissa Tapia Grimaldo, Sara Varandas Martins, Alannah Bruce, Isabel Moore, Rebecca Taubert, Chantal Macleod-Nolan, Stephanie McWaters, John Briggs, et al. Niche-breadth of freshwater macrophytes occurring in tropical southern african rivers predicts species global latitudinal range. *Aquatic Botany*, 136:21–30, 2017.

Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178, 1992.

Irina Laube, Heiko Korntheuer, Monika Schwager, Sven Trautmann, Carsten Rahbek, and Katrin Böhning-Gaese. Towards a more mechanistic understanding of traits and range sizes. *Global Ecology and Biogeography*, 22(2): 233–241, 2013.

Sarah E Lester, Benjamin I Ruttenberg, Steven D Gaines, and Brian P Kinlan. The relationship between dispersal ability and geographic range size. *Ecology Letters*, 10(8):745–758, 2007.

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.

Romana Limberger and Stephen A Wickham. Competition–colonization trade-offs in a ciliate model community. *Oecologia*, 167(3):723–732, 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.

Bo Luo, Sharlene E Santana, Yulan Pang, Man Wang, Yanhong Xiao, and Jiang Feng. Wing morphology predicts geographic range size in vespertilionid bats. *Scientific reports*, 9(1):1–6, 2019.

Anne E Magurran and Peter A Henderson. Temporal turnover and the maintenance of diversity in ecological assemblages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1558):3611–3620, 2010.

Anne E Magurran, Maria Dornelas, Faye Moyes, Nicholas J Gotelli, and Brian McGill. Rapid biotic homogenization of marine fish assemblages. *Nature communications*, 6(1):1–5, 2015.

Henry B Mann. Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, pages 245–259, 1945.

Robert M May. Will a large complex system be stable? *Nature*, 238(5364):413–414, 1972.

Hanna M McCaslin, T Trevor Caughlin, and Julie A Heath. Long-distance natal dispersal is relatively frequent and correlated with environmental factors in a widespread raptor. *Journal of Animal Ecology*, 89(9):2077–2088, 2020.

Melodie A McGeoch and Kevin J Gaston. Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biological Reviews*, 77(3):311–331, 2002.

Brian J McGill, Rampal S Etienne, John S Gray, David Alonso, Marti J Anderson, Habtamu Kassa Benecha, Maria Dornelas, Brian J Enquist, Jessica L Green, Fangliang He, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology letters*, 10(10):995–1015, 2007.

Steffen Moritz and Thomas Bartz-Beielstein. imputets: Time series missing value imputation in r. *The R Journal*, 9(1):207–218, 2017. doi: 10.32614/RJ-2017-009. URL https://doi.org/10.32614/RJ-2017-009.

Heikki Mykrä and Jani Heino. Decreased habitat specialization in macroinvertebrate assemblages in anthropogenically disturbed streams. *Ecological Complexity*, 31:181–188, 2017.

Maria Novosolov, Gordon H Rodda, Alexandra C North, Stuart HM Butchart, Oliver JS Tallowin, Alison M Gainsbury, and Shai Meiri. Population density–range size relationship revisited. *Global Ecology and Biogeography*, 26(10):1088–1097, 2017.

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.

Jacob D O'Sullivan, J Christopher D Terry, and Axel G Rossberg. Intrinsic ecological dynamics drive biodiversity turnover in model metacommunities. *Nature Communications*, 12(1):1–11, 2021.

Mark D Pagel, Robert M May, and Annie R Collie. Ecological aspects of the geographical distribution and diversity of mammalian species. *The American Naturalist*, 137(6):791–815, 1991.

Jennifer M Parody, Francesca J Cuthbert, and Ethan H Decker. The effect of 50 years of landscape change on species richness and community composition. *Global Ecology and Biogeography*, 10(3):305–313, 2001.

Simone Pigolotti, Alessandro Flammini, Matteo Marsili, and Amos Maritan. Species lifetime distribution for simple models of ecologies. *Proceedings of the National Academy of Sciences*, 102(44):15747–15751, 2005.

Frank W Preston. The commonness, and rarity, of species. *Ecology*, 29(3):254–283, 1948.

Christen Raunkiaer et al. The life forms of plants and statistical plant geography; being the collected papers of c. raunkiaer. *The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer.*, 1934.

Jiří Reif, David Hořák, Ondřej Sedláček, Jan Riegert, Michal Pešata, Záboj Hrázskỳ, Štěpán Janeček, and David Storch. Unusual abundance–range size relationship in an afromontane bird community: the effect of geographical isolation? *Journal of Biogeography*, 33(11):1959–1968, 2006.

Samuel Renner, Marina Schmidt Dalzochio, Eduardo Périco, Göran Sahlén, and Jukka Suhonen. Odonate species occupancy frequency distribution and abundance–occupancy relationship patterns in temporal and permanent water bodies in a subtropical area. *Ecology and Evolution*, 10(14):7525–7536, 2020.

Axel G Rossberg. *Food webs and biodiversity: foundations, models, data*. John Wiley & Sons, 2013.

Adriana Ruggiero. Latitudinal correlates of the sizes of mammalian geographical ranges in south america. *Journal of Biogeography*, pages 545–559, 1994.

Romain Sarremejane, Núria Cid, Rachel Stubbington, Thibault Datry, Maria Alp, Miguel Cañedo-Argüelles, Adolfo Cordero-Rivera, Zoltán Csabai, Cayetano Gutiérrez-Cánovas, Jani Heino, et al. Disperse, a trait database to assess the dispersal potential of european aquatic macroinvertebrates. *Scientific data*, 7(1):1–9, 2020.

Carrie A Schloss, Tristan A Nuñez, and Joshua J Lawler. Dispersal will limit ability of mammals to track climate change in the western hemisphere. *Proceedings of the National Academy of Sciences*, 109(22):8606–8611, 2012.

Pranab Kumar Sen. Estimates of the regression coefficient based on kendall's tau. *Journal of the American statistical association*, 63(324):1379–1389, 1968.

Catherine Sheard, Montague HC Neate-Clegg, Nico Alioravainen, Samuel EI Jones, Claire Vincent, Hannah EA MacGregor, Tom P Bregman, Santiago Claramunt, and Joseph A Tobias. Ecological drivers of global gradients in avian dispersal inferred from wing morphology. *Nature communications*, 11(1):1–9, 2020.

Rachel A Slatyer, Megan Hirst, and Jason P Sexton. Niche breadth predicts geographical range size: a general ecological pattern. *Ecology letters*, 16(8):1104–1114, 2013.

George C Stevens. The latitudinal gradient in geographical range: how so many species coexist in the tropics. *The American Naturalist*, 133(2):240–256, 1989.

Jukka Suhonen. Spatial and temporal changes in occupancy frequency distribution patterns of freshwater macrophytes in finland. *Ecology and Evolution*, 2021.

Jukka Suhonen and Jukka Jokimäki. Temporally stable species occupancy frequency distribution and abundance–occupancy relationship patterns in urban wintering bird assemblages. *Frontiers in Ecology and Evolution*, 7: 129, 2019.

Matthew RE Symonds and Christopher N Johnson. Range size-abundance relationships in australian passerines. *Global Ecology and Biogeography*, 15(2):143–152, 2006.

Riin Tamme, Lars Götzenberger, Martin Zobel, James M Bullock, Danny AP Hooftman, Ants Kaasik, and Meelis Pärtel. Predicting species' maximum dispersal distances from simple plant traits. *Ecology*, 95(2):505–513, 2014.

Henri Theil. A rank-invariant method of linear and polynomial regression analysis. *Indagationes mathematicae*, 12(85):173, 1950.

Mutsunori Tokeshi. Dynamics of distribution in animal communities: theory and analysis. *Researches on Population Ecology*, 34(2):249–273, 1992.

Berndt J van Rensburg, Melodie A McGeoch, Wayne Matthews, Steven L Chown, and Albert S van Jaarsveld. Testing generalities in the shape of patch occupancy frequency distributions. *Ecology*, 81(11):3163–3177, 2000.

Mark Vellend, Maria Dornelas, Lander Baeten, Robin Beauséjour, Carissa D Brown, Pieter De Frenne, Sarah C Elmendorf, Nicholas J Gotelli, Faye Moyes, Isla H Myers-Smith, et al. Estimates of local biodiversity change over time stand up to scrutiny. *Ecology*, 98(2):583–590, 2017.

Peter A Wilfahrt, Ashley L Asmus, Eric W Seabloom, Jeremiah A Henning, Peter Adler, Carlos A Arnillas, Jonathan D Bakker, Lori Biederman, Lars A Brudvig, Marc Cadotte, et al. Temporal rarity is a better predictor of local extinction risk than spatial rarity. *Ecology*, page e03504, 2021.

Benjamin Zuckerberg, William F Porter, and Kimberley Corwin. The consistency and stability of abundance–occupancy relationships in large-scale population dynamics. *Journal of Animal Ecology*, 78(1):172–181, 2009.

S1 Supplementary information

S1.1 Filtering the EA databases

The survey data included 4138, 2952, and 2816 unique water body identifiers in the macroinvertebrate (I), macrophyte (P) and diatom (D) datasets respectively. Of these, however, 597, 297, and 339 waterbodies respectively had no available management catchment identifiers and these were dropped. The three datasets represent distinct observation programs and the sites surveyed in a given catchment year don't necessarily correspond between programs. As such the following filtering steps were applied to each of the datasets independently. Since a clear trend in sites surveyed and taxa observed was found in the earliest years of the surveys, we excluded all observations prior to 1990 from the analysis, leaving catchment OFD time-series spanning around 30 years. We constrained our analysis to catchment-years in which a minimum of 10 taxa were detected, and a minimum of 5 sites were sampled. This helped to reduce small-sample-size effects in the estimated moments.

Since we are particularly interested in the macroscopic shape of the OFD best represented by its skewness, we used a rarefaction method to filter the dataset of catchment-years for which the estimated skewness of the OFD was unlikely to be biologically representative. For each catchment-year, we constructed a binary taxon-by-site table representing the presence-absence of taxa at each site in that year's survey. We then randomly subsampled the rows of the taxon-by-site table 1000 times and computed the skewness of the OFD of the random subsample. We then numerically estimated the first derivative of the skewness estimate as a function of the size of the random subsample. Catchment-years for which the absolute value of the first derivative of the skewness estimate evaluated within 10% of the actual observed richness, was greater than 10^{-2} were dropped from the analysis (Fig. S1), ensuring the accuracy of the skewness estimate despite unavoidable type II errors in the survey.

Finally, once these catchment-years were excluded, we dropped any catchment from the analysis for which the remaining time-series included less than 8 years of data. For catchments retained following filtering but for which single catchment-years were missing, we used a linear interpolation method to handle missing data points in the time-series (Moritz and Bartz-Beielstein, 2017). Prior to filtering, the database included 2708, 1625, 1120 catchment-years split between 97, 92, and 90 catchments for the I, P and D datasets respectively. Following these various filtering steps, we ended up with 1804, 731 and 588 catchment-years in 83, 52 and 48 catchments respectively.

S1.2 Fitting the LSPDM to data

The steady state of the simple locally saturated patch dynamics model can be solved analytically, and this analytic site occupancy distribution can be fitted to observation by tuning the mixing rate m. R code demonstrating the fitting procedure is available in the accompanying github repository.

S1.3 Fitting Pigolotti result to data

R code demonstrating the how to estimate α and m from metacommunity time-series is available in the accompanying github repository.

S2 Supplementary figures

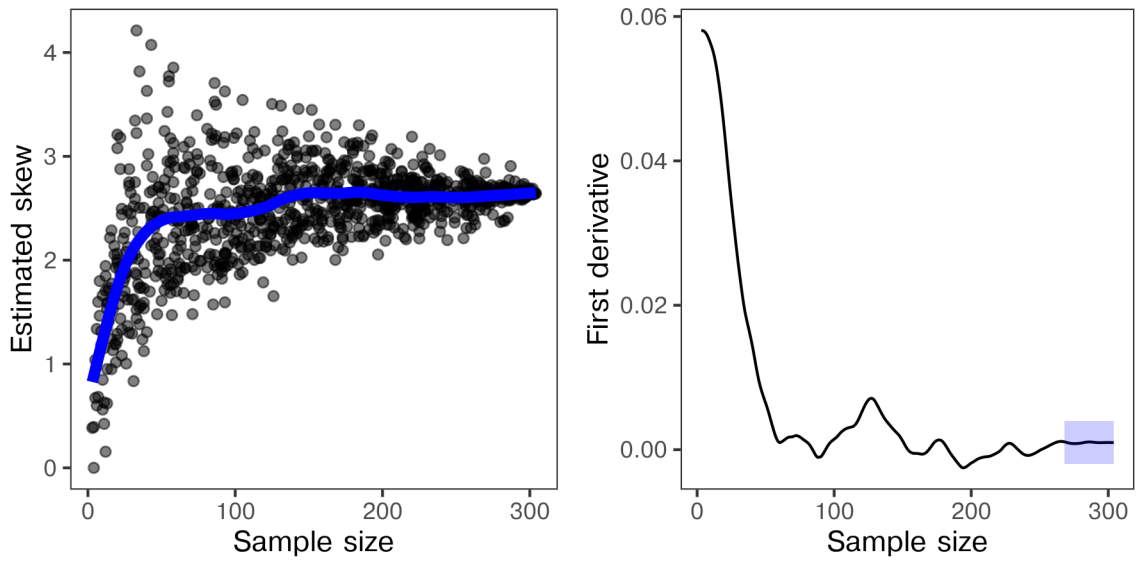

Figure S1: **Testing the accuracy of the skewness estimate for a given catchment-year.** Skewness of the OFD computed after randomly subsampling species to various 'depths' (number of species). The average of the first derivative evaluated over the last 10% of the sub-sample depth (blue box) is taken to represent the degree of convergence in the estimate.

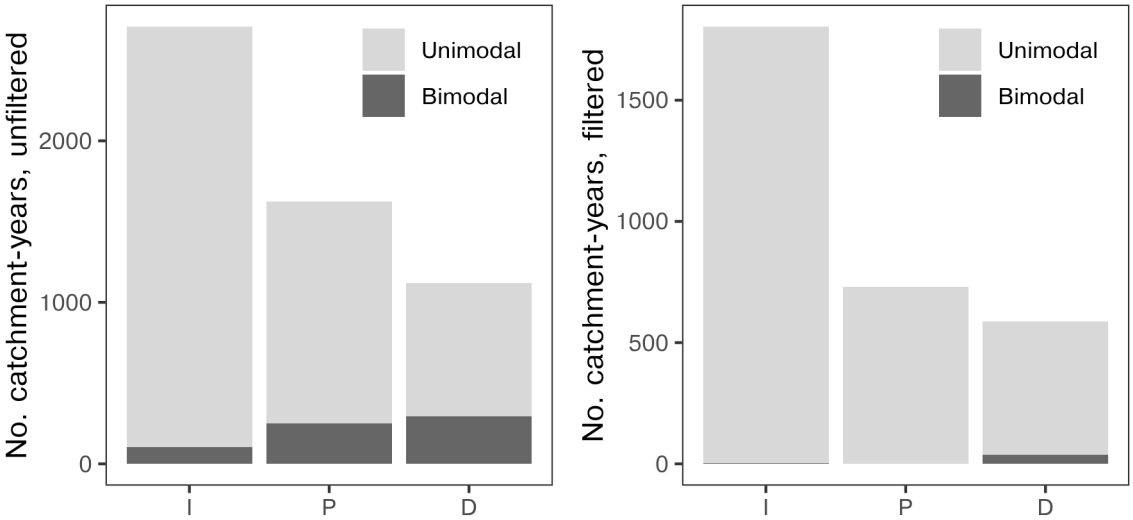

Figure S2: **Catchment-years for which unimodal and bimodal OFDs were detected.** The number of catchment-years for which significant bimodality in the OFD was detected (Tokeshi method, $p < 0.05$) with and without filtering the data.

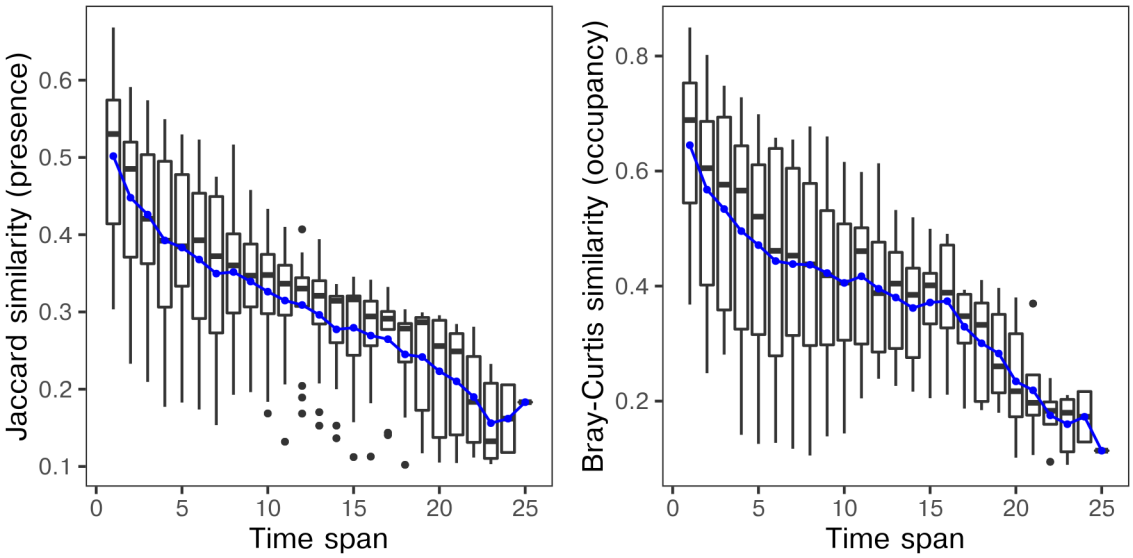

Figure S3: **The decay in compositional similarity with temporal distance.** The decay in macroinvertebrate similarity with regards to species presence-absence and site occupancy at the catchment scale as a function of temporal distance (time span) for a representative catchment (Ouse Upper and Bedford). We test for a significant trend by averaging over temporal distance (blue points represent the mean dissimilarity, lines represent the reconstructed timeseries)

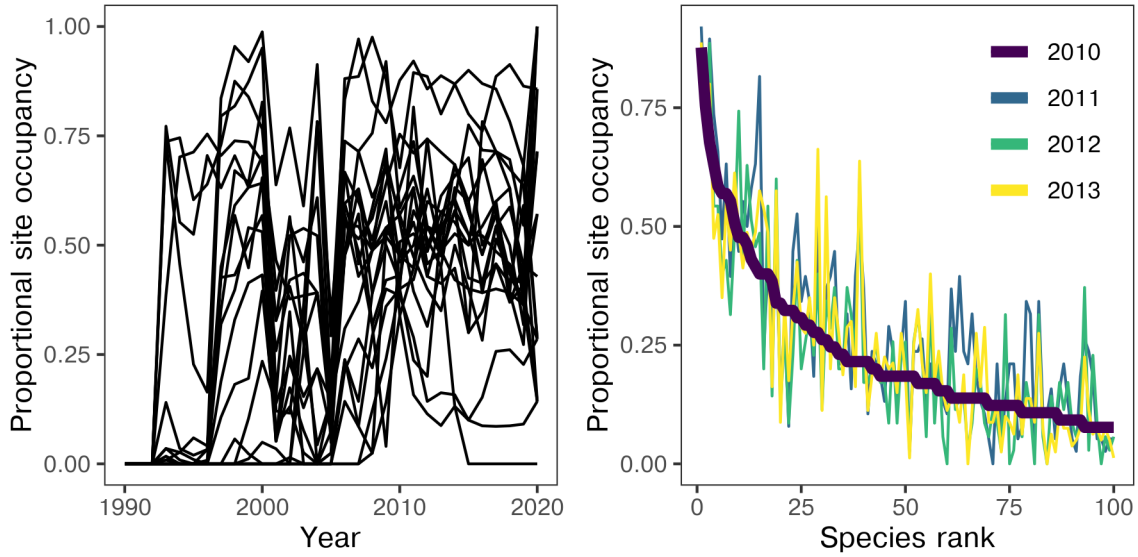

Figure S4: **Species level trends in site occupancy.** A: Proportional site occupancy of the 20 macroinvertebrate species with highest site occupancy in 2010 plotted over the entire time-series (Ouse Upper and Bedford). B: Rank occupancy curve of the 100 species with highest site occupancy in 2010 (thick black line) and the rank occupancy of the same 100 species in the following four years indicating changes in the identities of the most occupant species up to and including species losses from the record.

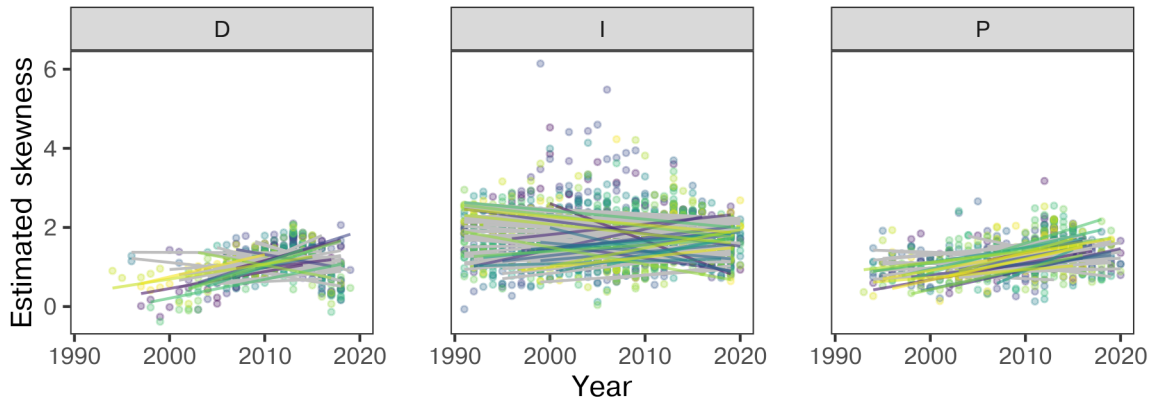

Figure S5: **Species level trends in site occupancy.** Estimated skew for each catchment plotted against sample year. Color represents catchment; lines are coloured grey in the case of a non-significant temporal trend. The slopes are the Sen's slope estimates while the intercepts are defined as the difference between the median skewness and the median year scaled by the estimated slope (Conover, 1999).

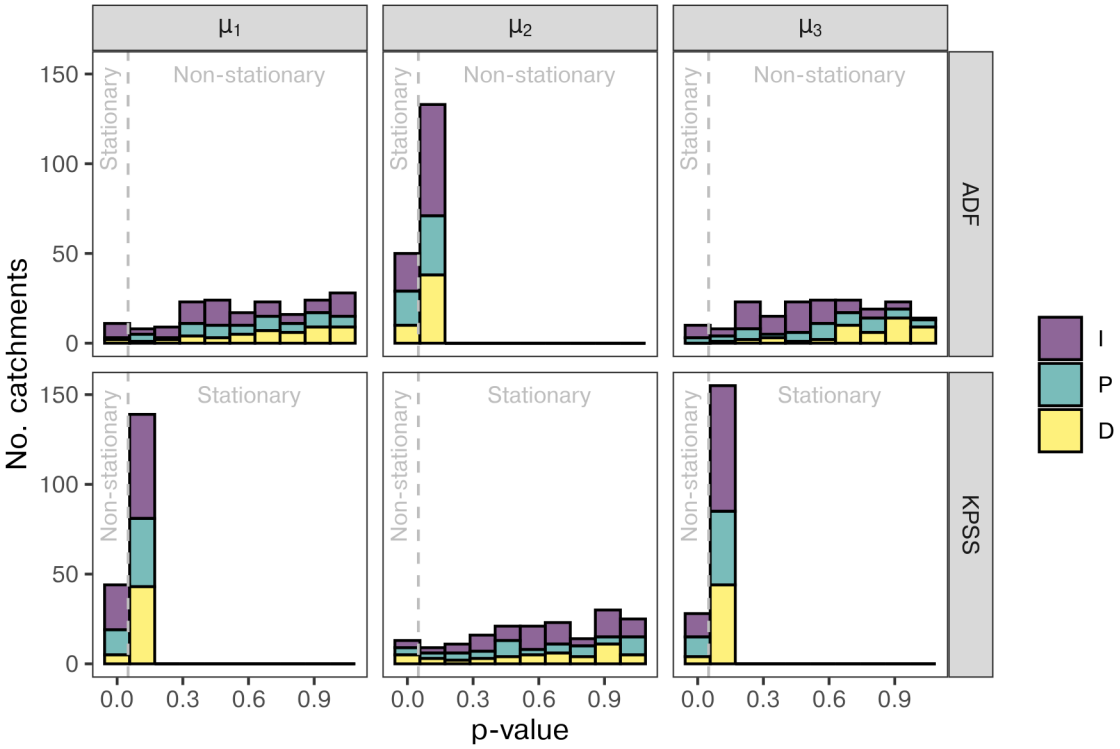

Figure S6: **The significance estimates of the ADF and KPSS tests.** The ADF and KPSS tests applied to the time-series of the mean (μ_1), variance (μ_2) and skewness (μ_3) of the catchment OFDs. Shown here are the p-values of the tests with 0.05 indicated by the dashed vertical line. In the ADF test $p < 0.05$ indicates stationarity, while in the KPSS test $p > 0.05$ suggests stationarity cannot be ruled out. For the majority of the catchments $p > 0.05$ for both tests and as such the stationarity assessment was inconclusive. Interestingly, though still most often non-significant, the strongest indication of stationarity was in the case of the variance (μ_2) of the OFD.

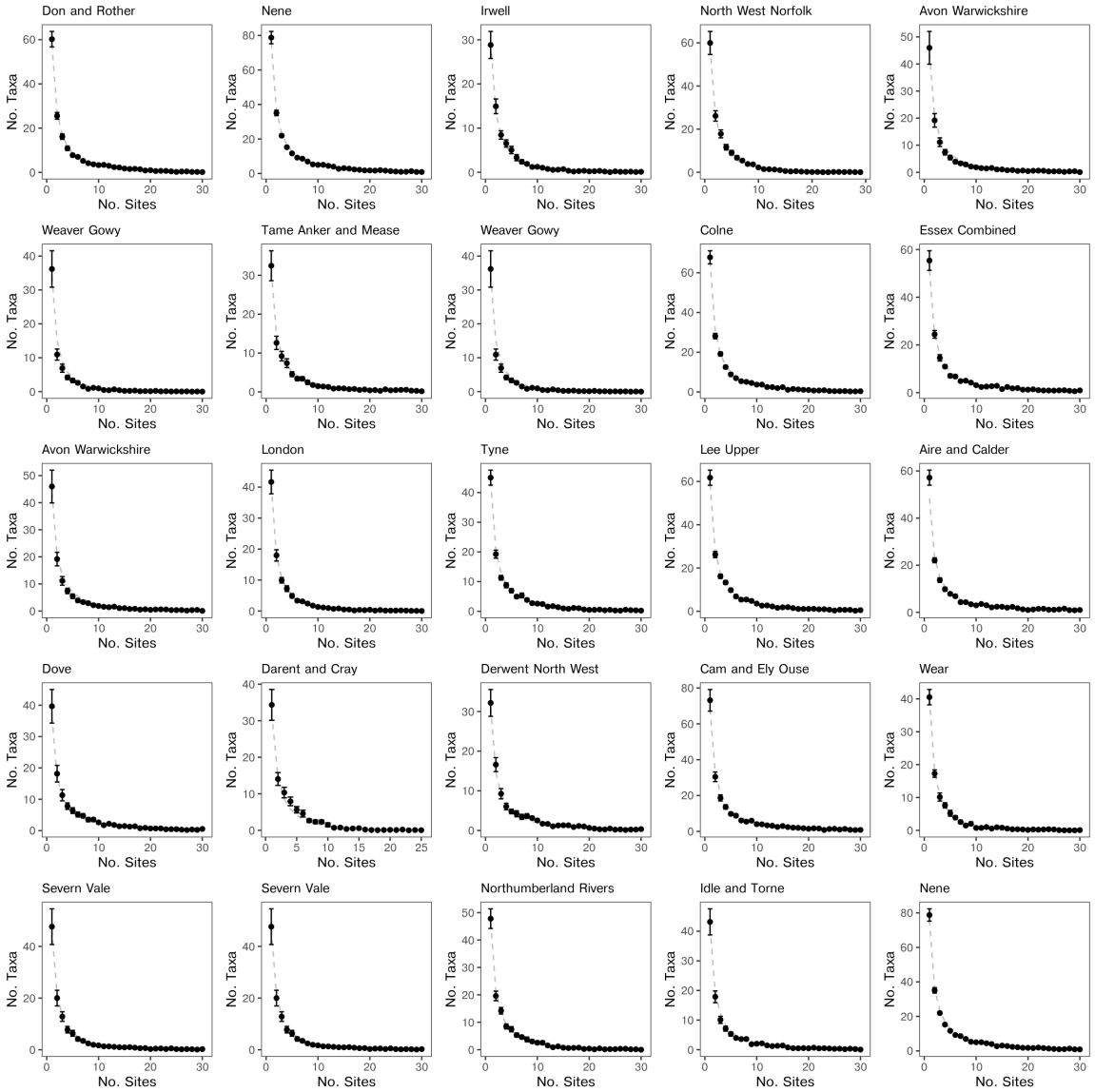

Figure S7: **Time averaged macroinvertebrate site occupancy distributions for 25 random catchments.** The dashed lines represent the fitted LSPDM.

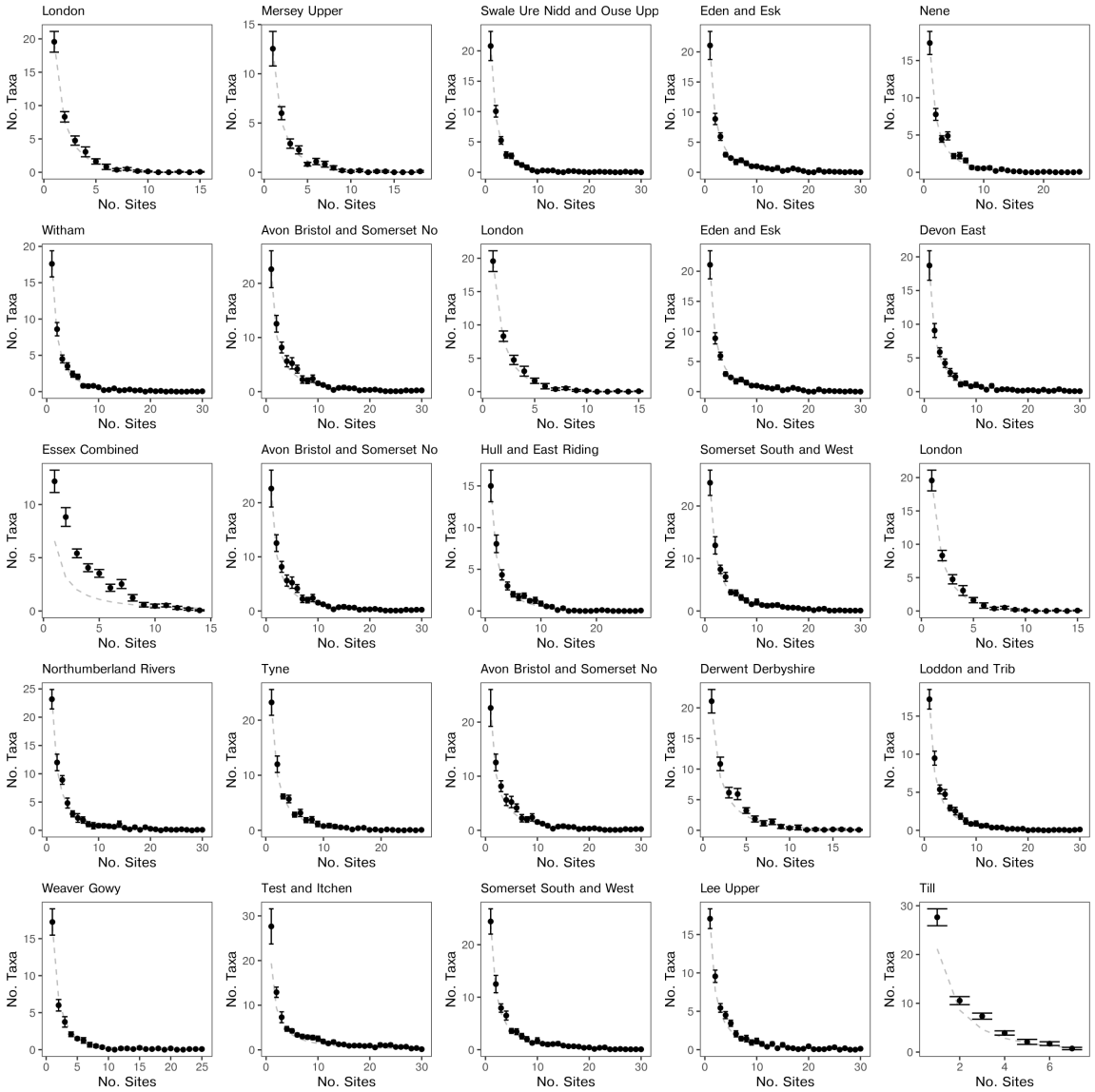

Figure S8: **Time averaged macrophyte site occupancy distributions for** 25 **random catchments.** The dashed lines represent the fitted LSPDM.

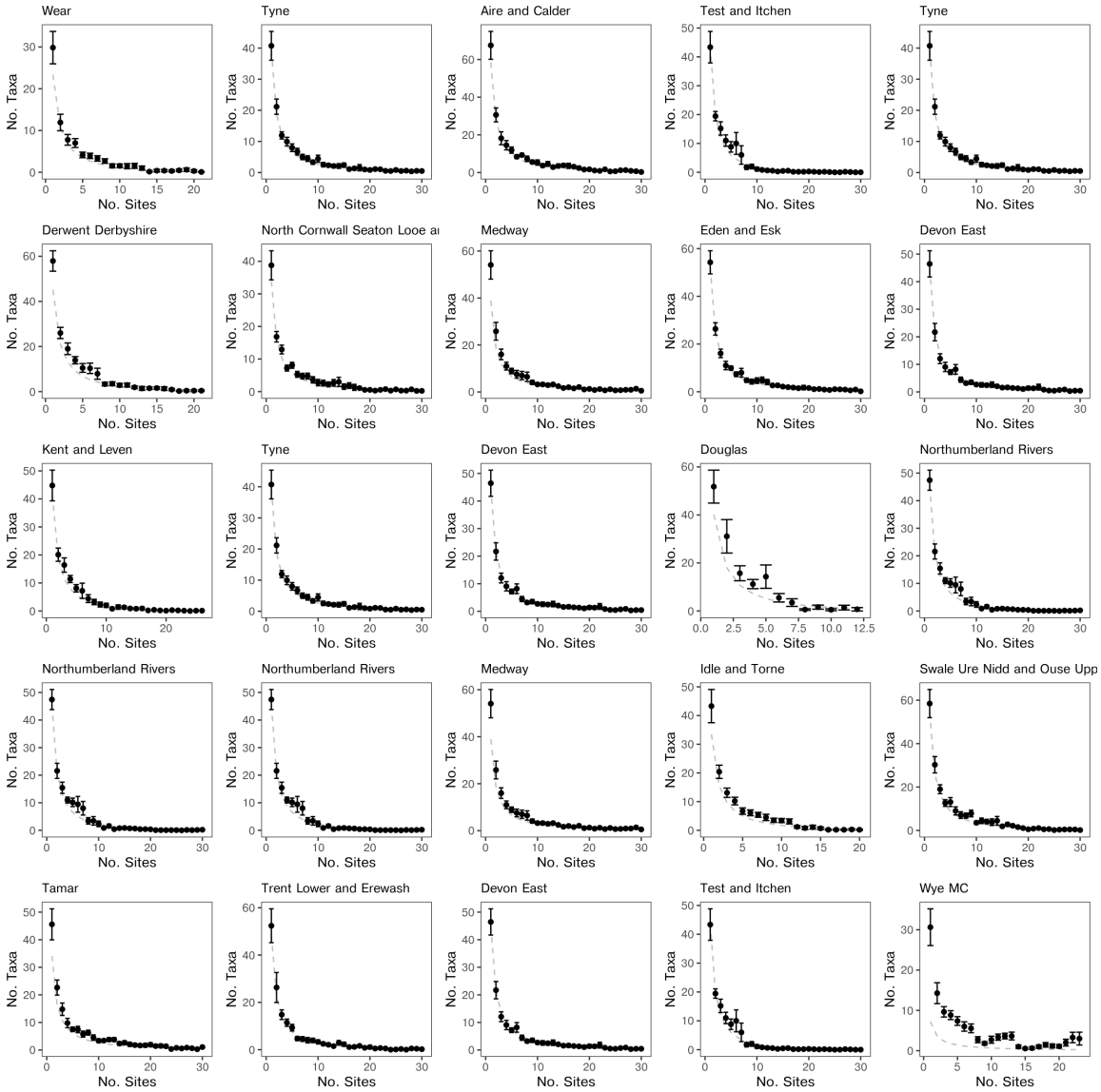

Figure S9: **Time averaged diatom site occupancy distributions for** 25 **random catchments.** The dashed lines represent the fitted LSPDM.

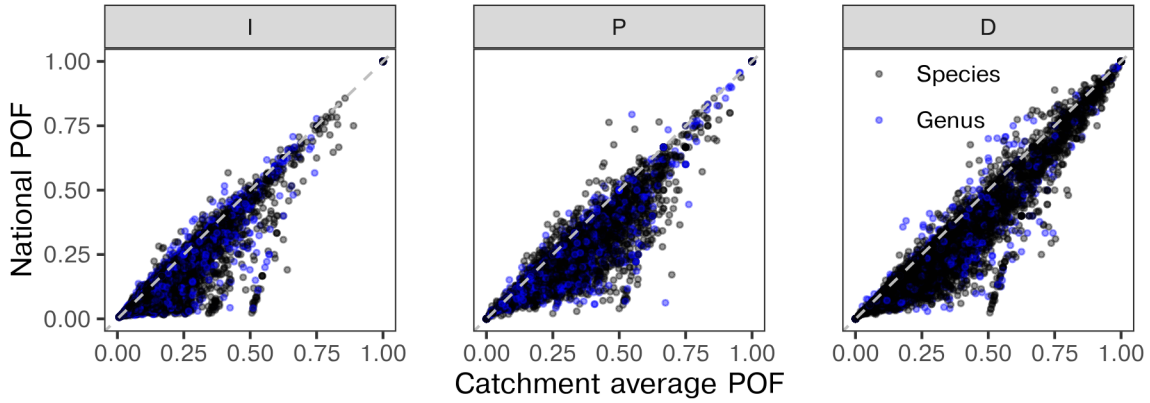

Figure S10: **Comparing time scales estimated from spatial and temporal analyses.** The mixing rates, m, and the α, estimated by analysis of the OFD or alternatively, the observed rate of turnover. Top row: estimates based on the time-averaged spatial structure of each catchment only. Bottom row: estimated by analysis of the observed turnover in species composition.