# Temporally robust occupancy frequency distributions in riverine ecosystems explained by local biodiversity regulation

Running title: Temporally robust OFDs in rivers

**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 of species' site occupancy—the occupancy frequency distribution (OFD)—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 examine both spatial and temporal patterns of site occupancy in communities of macroinverterbrates, macrophytes and diatoms. We find that the shape of the OFD at the catchment scale is typically time invariant despite substantial turnover in composition and species' proportional occupancy over the $30$ year time series. This can been seen as evidence in favour of an *ecological* mechanism for the shape of the OFD driven by local regulation of species richness and other biodiversity properties. We show that a simple, locally saturated patch occupancy model explains both the shape of the OFD and the temporal turnover rate in English rivers. This model is mathematically analogous to Hubbell's neutral theory however concerns patches occupied by a given species rather its population size. Our analysis supports the view that metacommunities exist in a dynamic steady state arising from local ecological constraints and that this intrinsic regulation can drives the emergence of a wide variety of macroecological patterns. We also show that interspecific differences are unlikely to play a dominant role in driving large scale ecosystem structure.

# 1 Introduction

The question of what determines commonness and rarity in communities of species has motivated ecologists for over seventy years (Preston, 1948). The question of commonness and rarity *in space*, represented by the occupancy frequency distribution (OFD), has been engaging researchers for even longer (Raunkiaer et al., 1934). Despite many decades of research, the mechanisms driving this fundamental ecological distributions remain incompletely understood.

The OFD quantifies the number of sites in occupied by species in a given landscape and the distribution of site occupancy across species. In most ecological assemblages the majority of species are observed in a minority of sites and the OFD is typically 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 proposed to impact the shape of the OFD. The still-influential review into the OFD by McGeoch and Gaston (2002) identified a number of biogical mechanisms not rooted in sampling effects plausibly responsible for the shape of the distribution: interspecific variation in niche breadth, interspecific variation in dispersal ability, and site occupancy dynamics described in Hanski's (1982) classical joint metapopulation model. However, we contest whether these proposed mechanisms can sufficiently explain the overall shape of the distribution.

Environmental niche widths are likely to influence occupancy patterns at the species scale, with habitat specialists and generalists typically of low and high occupancy respectively (Brown, 1984; Brown et al., 1995). Indeed, the occupancy of environmental generalists has been found to exceed that of specialist competitors (Denelle et al., 2020). Likewise, a positive relationship between niche breadth and range size—generally assumed to correlate with occupancy—has been reported across taxonomic groups and spatial scales (Slatyer et al., 2013). 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 that, as such, the association may often be tautological.

Moreover, an apparent relationship between niche breadth and site occupancy alone does not explain the shape of the OFD. This would also require that narrow niche widths are substantially over-represented in nature, however, evidence for such a tendency is weak. By pooling data at the global scale Belmaker et al. (2012) find a numerical dominance of diet and habitat specialists in birds. However, the apparent positive association between latitude and niche breadth (Stevens, 1989; Pagel et al., 1991; Ruggiero, 1994), combined with the extensively documented negative association with richness (Pianka, 1966), mean it is likely that the bias toward narrow niche widths

is exaggerated at the global scale compared to the more narrow latitudinal ranges at which OFD are typically reported. At regional scales, weak right skew in the estimated range of environmental tolerance has been found in riverine macroinvertebrates (Mykrä and Heino, 2017). However, this contrasts with a symmetric distribution found in niche breadths of freshwater macrophytes (Kennedy et al., 2017) and the apparent over-representation of generalist species within a regional-scale plant assemblage (Boulangeat et al., 2012). Thus, consistent strong evidence of right skew in the distribution of niche widths that could explain that ubiquitous shape of the OFD has not been found.

Alternatively, high occupancy could plausibly be associated with good disperseral ability (Collins and Glenn, 1997). While evidence for a mechanistic link between dispersal ability and 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). However, OFD a generally reported for guilds of species for which commonly used proxies for dispersal ability based on body- or propagule-size do not vary substantially. A correlation between such proxies and *within guild* occupancy has, to our knowledge, not been demonstrated.

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). Again, 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 likely to miss-represent landscape-scale distributions. Crucially, a uniquely phenotypic explanation for the shape of the OFD would imply that species tend to occupy roughly consistent numbers of sites in time in the absence of strong environmental change. In fact, as we show below, species' site occupancy in contemporary England, for example, is extremely variable in time.

Hanksi's (1982) classic multi-species metapopulation model, which describes stochastic population-level fluctuations in occupancy, has 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 or bimodal OFD, but requires the assumption that local extinction rates decrease with overall occupancy, which may not be empirically justified (Wilfahrt et al., 2021) and is arguably mechanistically circular.

Here we test an alternative hypothesis, that the shape of the OFD is a fundamental outcome of an ecosystem subject to intrinsic constraints on the structure of biodiversity. This *ecological* explanation, in contrast to the phenotypic hypotheses described elsewhere, does not depend on interspecific heterogeneities, but instead on strong

constraints on site-level species richness reported in both theoretical (Drossel et al., 2001; Yoshida, 2003; Pawar, 2009; Rossberg, 2013; O'Sullivan et al., 2019) and emprical studies of community ecology (Brown et al., 2000; Parody et al., 2001; Magurran and Henderson, 2010; Magurran et al., 2015; Dornelas et al., 2014; Gotelli et al., 2017; Vellend et al., 2017). Such constraints occur at the local scale, the outcome of basic limits on the structure and diversity of local ecological assemblages. Evidence of the propagation of local processes up organisation scales to drive metacommunity patterns in biodiversity would imply a important step toward theoretically unifying the local and regional community scales (Rabosky and Hurlbert, 2015; Harmon and Harrison, 2015). The key signature of such an ecological mechanism would be the detection of time invariance in empirical OFDs despite continuous change in the identities and proportional occupancies of the constituent taxa (O'Sullivan et al., 2021b). We find that indeed the shape of the OFD for riverine communities is typically time invariant despite substantial temporal turnover in composition. We then model the empirical OFDs using a simple patch occupancy model in which colonisation and extinction rates are independent of species identity or global abundance. Our model reproduces both the spatial and temporal properties of the empirical observation, a crucial outcome since many conceivable models can reproduce a right-skewed distribution; rigorous model validation requires testing against ecological properties beyond a "hollow curve" distribution (McGill et al., 2007).

# 2 Methods

## 2.1 Compiling the database of temporal OFDs

We tested for time invariance on the OFD using a database of macroinvertebrates (I), macrophytes (P) and diatoms (D) observations collated from publicly available material recorded by the Environment Agency (EA) of England[1]. The database covers the whole of England at good spatial resolution (Fig. 1), and includes over 30 years of observations. We retained only observations identified to species or genus level. Of these, taxa identified to genus level made up $31\%$, $30\%$, and $11\%$ of the I, P and D datasets respectively.

The EA divides freshwater ecosystems in England into 10 River Basin Districts, each of which contains between 2 and 20 Management Catchments (MCs). Waterbody MC allocation was extracted from the EA Catchment data explorer[2] and this information was merged with the database of ecological observations. MCs are delineated based on internal EA decision making. We plausibly assume that the allocation of water bodies to MCs

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

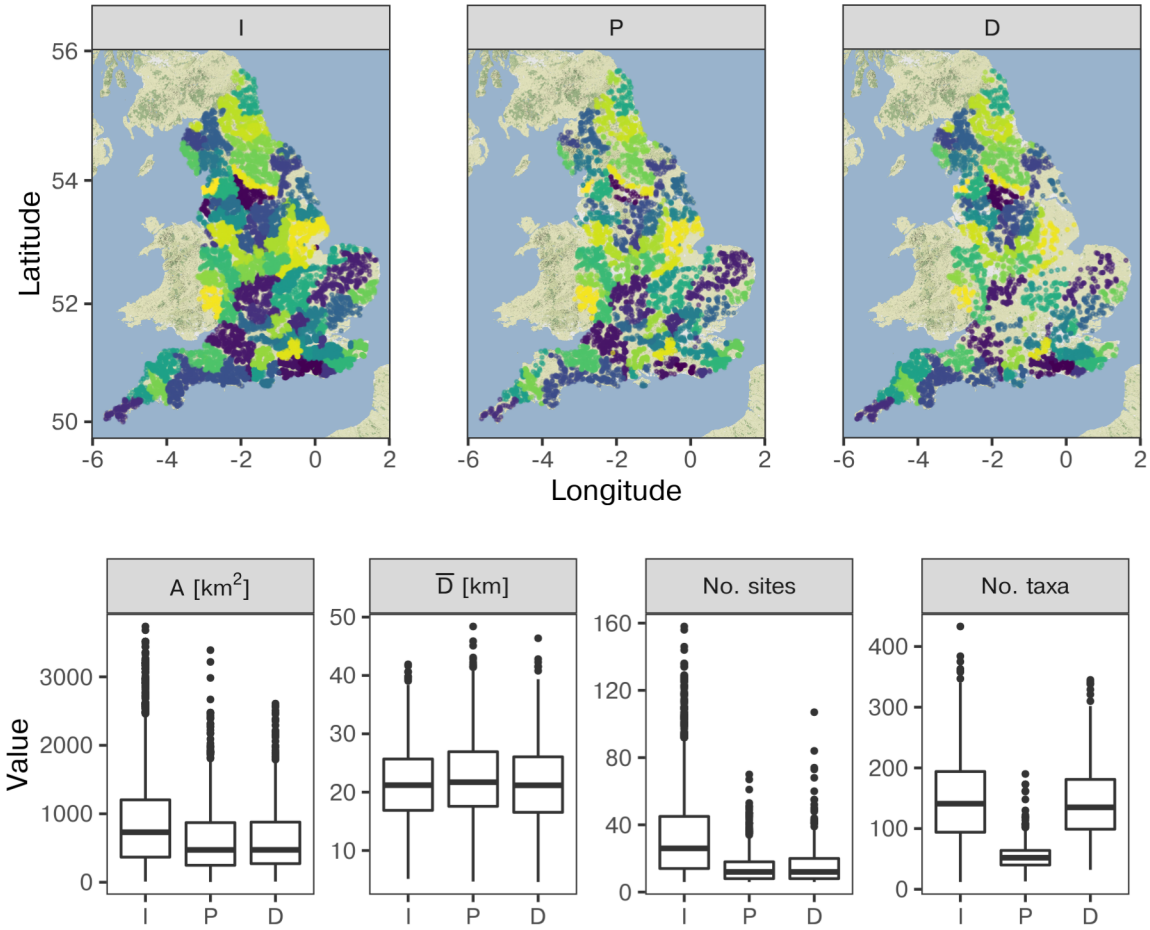


Figure 1: **English riverine community surveys.** Top row: The locations of sites in the three nationwide surveys. Points correspond to sites, colours to the management catchment. Bottom row: The area of the convex hull bounding sites, mean inter-site distance, number of sites, and number of taxa observed in individual catchment years for each dataset, plotting according to the standard Tukey representation (Tukey, 1977). In all cases, figures represent the datasets after filtering the data (see Supplementary information).

involved some analysis of habitat type, spatial cohesion and river basin structure, such that MCs represent distinct metacommunities to a good approximation. In the following, we ask, for each catchment and each of the three taxonomic groups independently, whether shape of the catchment-scale OFD varies over time.

To account for the fact that the number of sites and their locations were not necessarily fixed through time for a given catchment, for the time series analyses we focus on the proportional site occupancy distribution. We deal with cases of strong variation in sampling effort by applying a suite of filters to the data (summarised in

**Box 1: The steady state OFD of the Locally Saturated Patch Occupancy Model**

The Locally Saturated Patch Occupancy model: The spatially implicit metacommunty is represented by an $S$-by-$N$ 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). The R script used to generate the simulation and for the fitting procedure are available in the accompanying Github repository.

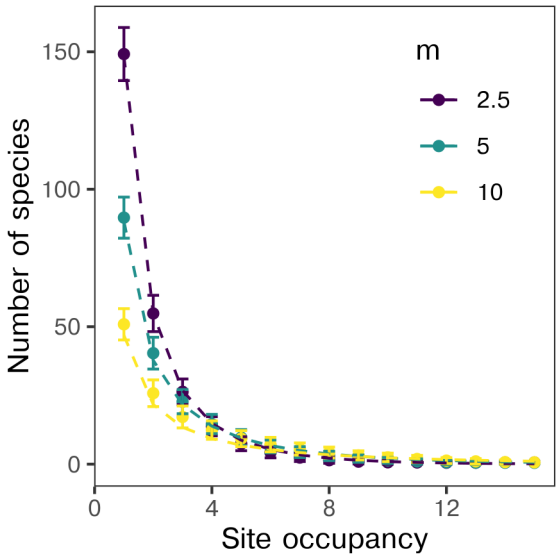


Figure 2: The average site occupancy in 1000 realisations of the probabilistic LSPOM following $10,000$ simulated invasions. Here $N = 50$, $n = 10$ and the mixing rate was set to $m = 2.5$, 5, and 10. The dashed lines represent the fitted curves for which $m$ was estimated as 2.8, 5.9 and 12.5 respectively, giving an indication of the measurement error of the fitting procedure.

the Supplementary information, Fig. S1). Following filtering, we retain 83, 52 and 48 catchments (I, P and D respectively) sampled over a period of around 30 years, resulting in 1804, 731 and 588 unique OFDs. The area of the convex hull bounding the sites, the mean distance between sites, the number of sites surveyed and the number of taxa detected in each catchment-year after filtering are shown in Fig. 1. All data and R scripts required to run

**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 LSPOM model. Here we translate this result to the patch occupancy scale 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 $\alpha$ equation 3 can be fitted to the observed turnover at the catchment scale. Further details including R code for estimating $\alpha$ from metacommunity time series are included in the Supplementary information and supplementary information.

the analyses are available in the accompanying Github repository.

## 2.2 Testing for bimodality in the OFDs

Our analysis assumes broad consistency in the qualitative shape of the OFD across space and time. If this were not the case, a more *ad hoc* analysis would be needed. We assessed the overall shape of the OFDs for each catchment-year by testing for bimodality 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. This broad consistency in the qualitative shape of the distributions meant we could use the first three moments of the distributions to compare OFDs between catchments and years without substantial loss of information.

## 2.3 Catchment-scale turnover in composition and occupancy

In order to verify that time invariance in catchment OFD was not simply because the time series were typically shorter than the timescale of community turnover, we explored the temporal trends in two measures of compositional similarity. The first was the average change in Jaccard similarity (Jaccard, 1912) in catchment composition (i.e. summing over all sites) over time. The second was the average change in Bray-Curtis similarity (Bray and

Curtis, 1957) in the non-binary vectors describing the site occupancy for each species over time. In both cases these were studied as functions of time lag irrespective of directionality (Fig. S3).

## 2.4 Analysing the temporal dynamics of the empirical OFDs

We tested whether the shape of the OFDs varied over time by analysing the temporal trends in the mean, variance and skewness of the distribution for each catchment. The existence of a significant temporal trend in each of these descriptive statistics 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 in 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 null hypothesis is the absence of a unit root (stationarity).

## 2.5 A locally saturated patch occupancy model

We model the OFD at the catchment scale using a locally saturated patch occupancy model (LSPOM). The LSPOM is analogous to the Unified Neutral Theory of Biogeography (UNTB; Hubbell 2001) applied at a higher organisational scale. It can also be understood as an abstraction of a deterministic metacommunity model constructed via the coupling of Lotka-Volterra communities in a spatially explicit landscape. This Lotka-Volterra Metacommunity Model (LVMCM) has been previously shown to reproduce a variety of spatial and temporal macroecological patterns (O'Sullivan et al., 2019, 2021b). Elsewhere we develop an approach to fitting the LVMCM to empirical data using the LSPOM as an intermediate (O'Sullivan et al., 2021a). For completeness, in the Supplementary information we show that the LSPOM as described provides a good fit to both an empirical OFD (Benito et al., 2018) and an LVMCM simulation with matching spatial and environmental structure.

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, the LSPOM models 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 occupancy (PO) framework allows us to explore the impact of local diversity regulation in PO metacommunities using the mathematics already developed

in the context of the UNTB.

The steady state of the LSPOM model is a (time invariant) unimodal, right skewed OFD (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 colonisation from within the assemblage. Invasion of new taxa from outside occurs at a rate $\lambda$. Again due to the assumption of local saturation, $\lambda$ 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 LSPOM, $b$ is equivalent to $m$, the rate at which new *populations* are 'born' via colonisation. Similarly, when translated to the PO scale, $d$ can be understood as the extirpation rate, driven by both internal colonisation and arrival into the network from outside such that $d$ is analogous to $m + \lambda$ in the LSPOM. For convenience we set $\lambda = 1$, thereby modelling patch occupancy dynamics in units of average regional invasion times.

Since the emergent OFD of the LSPOM is controlled by a single parameter, $m$, it can readily be fitted to observed OFDs. A detailed description of the LSPOM 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.

## 2.6 Comparing LSPOM colonisation rate to observed turnover

Key to this analysis is the fact that by fitting the process based LSPOM to a snapshot OFD—a static phenomenon—we can estimate the internal colonisation rate represented by $m$. We then compare this mixing rate to an alternative estimate computed based on comparison of the local- and catchment-scale turnover. By comparing these estimates, we ask if the LSPOM meaningfully captures both the spatial and temporal properties of the empirical observation.

An alternative method for estimating $\alpha$, the ratio $b/d$, can be found by analysis of the temporal turnover rate at the local and metacommunity scales. This method builds upon analytic results derived for individual based neutral models Pigolotti et al. (2005). 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). This analytic probability 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 $\alpha$, Eqs. (2) and (1) give an estimate of $m$ based not on the OFD of the catchment, but instead on analysis of the community time series.

For each of the catchments by fitting the LSPOM to the observed OFD, we estimated the mixing rate $m_{OFD}$, and the corresponding $\alpha_{OFD}$. From analysis of the metacommunity time series and using Eq. (1), we compute alternative estimates which we denote $m_{TS}$ and $\alpha_{TS}$. We then ask whether the $m$ and $\alpha$ estimated using the OFD only (fitting to the LSPOM) quantitatively matches those estimated by reference to the observed rate of turnover.

# 3 Results

## 3.1 Temporal trends in the OFD

The proportion of sites occupied by species varied over the time series. This species-level variability in site occupancy is sufficient to drive substantial change in the occupancy rank of species, frequently including the loss of species from the record (Fig. S5). Studying these fluctuations at the whole commuity scale, we find that both the Jaccard similarity in community composition, and the Bray-Curtis similarity in proportional site occupancy showed clear signs of turnover. In Fig. 3A we show the decay in the latter of these two measures for all catchments and each taxonomic group. Interestingly, the shape of the OFDs were largely consistent across catchments and taxa (Figs. S7-S9).

Despite the university tendency for catchments to turnover with respect to composition and occupancy over the time serires, we detected significant temporal trends in mean, variance or skewness of the catchment OFDs in a small proportion of the catchments. In Figs. S10-S12 we show the linear trend of the moment expansion over time for 12 random catchments and in Fig. 3B we show the number of catchments for which positive, negative or no trend were detected. For macroinvertebrates and diatoms in particular the vast majority of catchments show no temporal trend in the first three moments of the OFD. The effect is perhaps weaker in the macrophytes and indeed the only exception to the numerical dominance of catchments with no temporal trend in shape was in the variance of macrophyte OFDs. The proportion of significant trends overall is higher than would be expected by random chance, however these are exceptional cases nonetheless. It is conceivable that the catchments for which significant trends were detected are in fact relaxing following some anthropogenic perturbation.

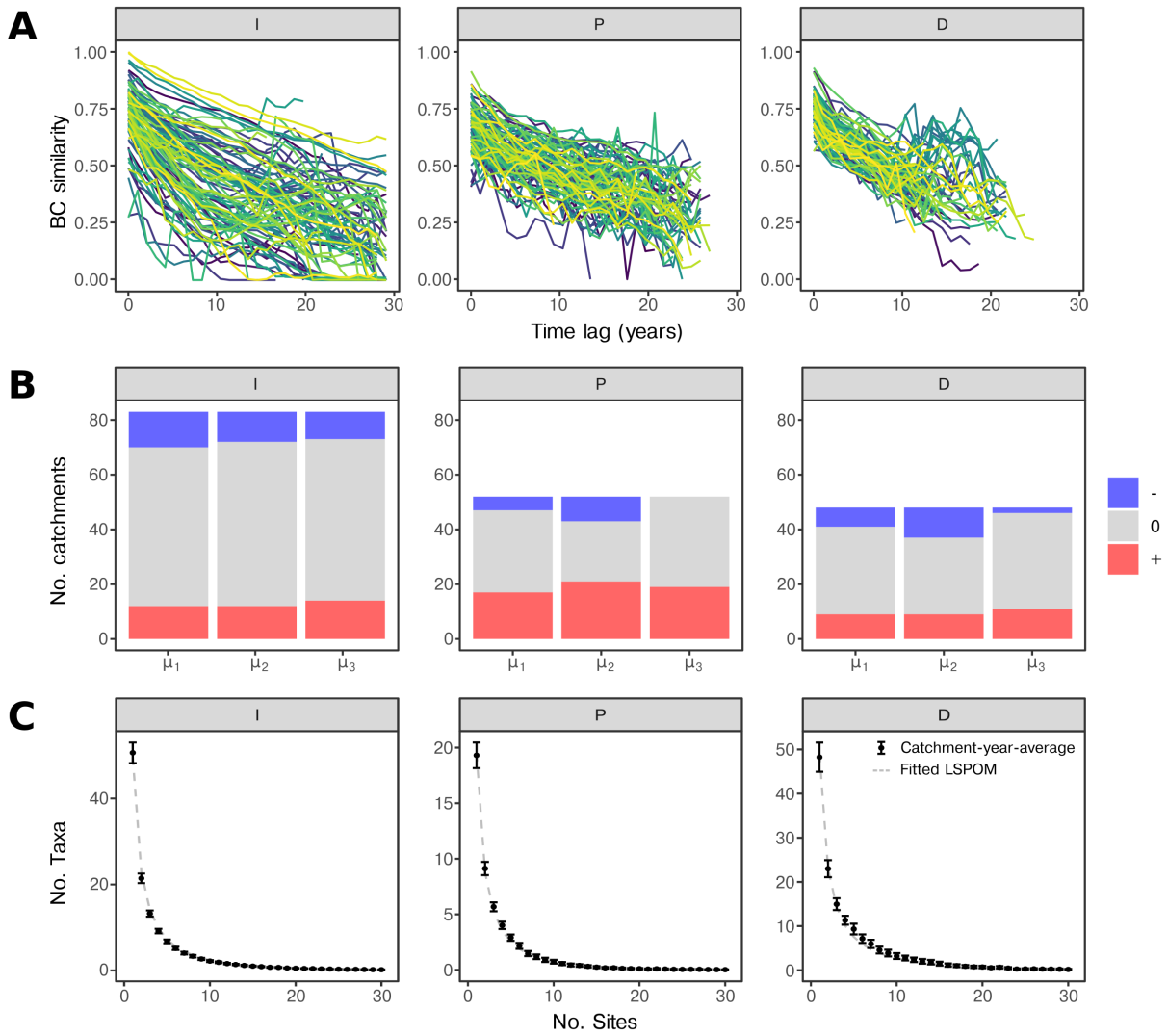


Figure 3: **Temporal trends in metacommunity OFD and composition.** **A**: Turnover in composition and occupancy at the metacommunity scale shown as the average BC similarity of the proportional site occupancy vectors plotted against time lag. Metacommunity-scale turnover in composition and occupancy was observed in all catchments over the time series. **B**: The outcome of time series analysis of the first three moments of catchment OFDs. For each of the mean ($\mu_1$), variance ($\mu_2$) and skewness ($\mu_3$), in all three taxonomic groups, the majority of catchments showed no significant temporal trend, with the single exception of the variance in macrophyte OFDs. Colour represents the sign of the Sen's trend, with grey indicating no trend detected using Mann-Kendall. **C**: The OFDs generated by averaging across catchments and years fitted by the LSPOM. The estimated mixing rates for these composite OFDs were 12.2, 8.6 and 14.7 for the I, P and D datasets respectively.

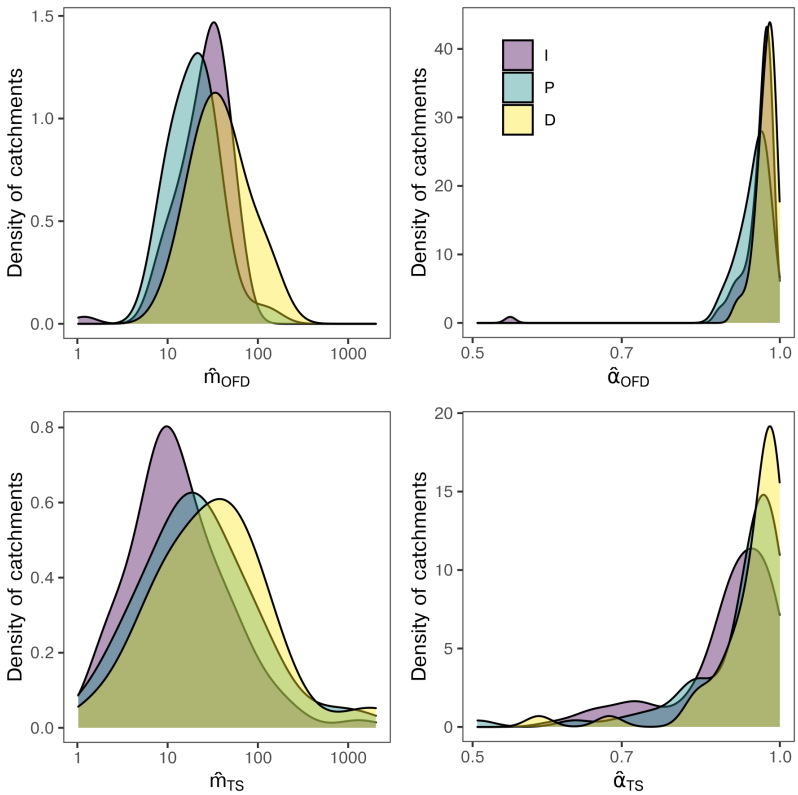


Figure 4: **Comparing time scales estimated from spatial and temporal analyses.** The mixing rates, $m$, and the $\alpha$, estimated by analysis of the OFD or alternatively, the observed rate of turnover. Top row: estimates based on the time-averaged OFD of each catchment only. Bottom row: estimated by analysis of the observed turnover in species composition.

Having confirmed that the shape of the OFDs remained time invariant for most catchments, we then tested for stationarity in the time series of the moment expansion (Fig. S6). For the majority of catchments and in all three datasets, using ADF we did not find evidence of non-stationarity in the time series of the mean, variance and skewness of the catchment ODFs ($p > 0.05$), despite an strong tendency to detect no temporal trend (Fig. **??**). In contrast, using the KPSS test we did not find evidence of stationarity ($p > 0.05$). Thus, we cannot conclusively confirm that fluctuations in the shape of the OFD can be modelled as stationary processes, though we find a weak indication of stationarity of the variance (Fig. S11). Longer time series will be required to decide this question.

## 3.2 Patch-dynamic modelling of the OFD

Given the spatial and temporal consistency in riverine OFDs, we fitted the LSPOD to the composite distribution generated by averaging over both space and time for a given taxonomic group. We find that the simple model gives

a surpisingly accurate fit between the LSPOD and the taxon-average OFD with estimated mixing rates of $12.2$, $8.6$ and $14.7$ for the I, P and D datasets respectively (Fig. 3C). In the supplementary information we show the LSPOM fitted to the time-averaged OFDs for a random subset of the catchments. In the majority of cases, the model fits the OFD of the empirical metacommunities well.

Fitting the LSPOM to the time-averaged OFD for catchments independently, we computed estimates for the internal colonisation rate $m_{OFD}$ of each. 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. 4). The corresponding $\alpha_{OFD}$ differed only slightly from $1$ implying invasion into the catchment is rare compared to internal colonisation. We then computed alternative estimates $m_{TS}$ and $\alpha_{TS}$ by fitting Eq. (1) to the catchment time series. We find comparable estimates by reference to the observed turnover rate (Fig. 4). The ratio between the median estimates computed via the spatial and temporal analyses $\hat{m}_{OFD}/\hat{m}_{TS}$ was in the range $1.07$-$2.86$ 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 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. This implies that the spatial analysis correctly inferred the timescales of catchment turnover to within an order of magnitude. Thus, we find that the LSPOM accurately represents the observed OFD (Figs. 2, S7-S9), but crucially that the demographic rates implied by the occupancy patterns are broadly consistent with the local and metacommunity scale turnover, differing by at most a factor of $2.7$ on average (Fig. 4).

| | $\hat{m}_{OFD}/\hat{m}_{TS}$ | MW $p$-value |
|---|---|---|
| I | $2.86$ | $1.9 \times 10^{-6}$ |
| P | $1.07$ | $0.92$ |
| D | $1.10$ | $0.34$ |

Table 1: The ratio of the median mixing rate $m$ computed by reference to the time-averaged OFD (using the LSPOM model) to the equivalent quantity estimated by analysis of the time series (Pigolotti et al., 2005). The significance of the difference in the distribution was tested using the Mann-Whitney U test.

# 4 Discussion

Despite decades of research, the mechanisms responsible for the shape of the OFD remain poorly understood. Researchers often rely on incomplete explanations targeted at the single species level—niche width or dispersal

heterogeneities—but ignore the fact that on their own such heterogeneity sheds little light on the cause of the particular *shape* of observed OFDs. With the present study we provided compelling evidence in favour of a fundamental *macroecological* mechanism underlying this pattern. We have successfully modelled the OFD as the outcome of a stochastic colonisation and extinction process with equal probabilities for all species and assumed locally saturated assemblages. Thus, we have highlighted the fact that interspecific differences are unlikely to play a dominant role in driving this fundamental macroecological pattern. Our proposed model framework improves on previous attempts to distil the basic biology structuring macroecological assemblages in its simplicity, the readiness with which it can be fitted to empirical data, and the good fit to both spatial and temporal macroecology. Moreover, we have shown that interspecific *interactions*, implicitly incorporated in the model via local limits to diversity, play a central role.

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 dynamics have been demonstrated (Blowes et al., 2019). Here we build upon this work by showing that such intrinsic constraints propagate up to the metacommunity scale and ultimately determine macroecological structure represented by the OFD.

We tested for stationarity in the first thee moments of the catchment ODFs, however the results were inconclusive. Weak evidence of a greater tendency toward stationarity in the variance of the OFD was found, however, at present, the reason for this apparent difference in the time scale of fluctuations remains unclear. Stationarity has previously been used to evidence intrinsic regulation of species richness in communities (Gotelli et al., 2017). There is good reason to believe that metacommunity dynamics occur at a rate slower than local community dynamics (O'Sullivan et al., 2019) and as such that substantially longer time series would be required to demonstrate stationarity in metacommunity properties. In fact, the average compositional similarity as a function of time lag shown in Fig. 3A suggests that the $30$ year period for which data are available corresponds in many cases to around one 'half-life' of catchment-scale turnover (BC similarity drops to $0.5$, assuming turnover can be modelled as exponential decay to zero). This confirms the intuition that even at multi-decadal scales, time series are likely to be too short to make strong claims regarding the stationarity of fluctuations. However, we argue that the fact that the OFD is a *distribution* rather than a scalar property like species richness, and the fact that this distribution is consistently strongly biased, can be seen as *prima facie* evidence of some intrinsic regulation akin to an entropic

constraint.

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.

The OFD and the species abundance distribution (SAD), while likely related, are not functionally equivalent phenomena in ecology. 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). The niche width and dispersal length heterogeneities described above which, while not sufficient to explain all properties of the OFD, do play a role that is absent from the SAD. Furthermore, 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, despite the links between these two key macroecological patterns, evidence of ecological drivers of the OFD should be seen as distinct from what was already known regarding the mechanisms responsible for the SAD..

The model used to explain the shape of the OFD is essentially the simplest possible dynamical model of a multi-species metacommunity subject to local biodiversity regulation. The LSPOM builds upon previous work on a related model, the Lotka-Volterra Metacommunity Model. 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, 2021b). 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., 2021b) in a way that is closely analogous to the stochastic colonisations and extinctions that occur in PO type models. Thus, the LSPOM

should be seen as an empirically falsifiable abstraction of the more complex LVMCM.

# 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.

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.

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.

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.

Barbara Drossel, Paul G Higgs, and Alan J McKane. The influence of predator–prey population dynamics on the long-term evolution of food web structure. *Journal of Theoretical Biology*, 208(1):91–107, 2001.

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.

Luke J Harmon and Susan Harrison. Species diversity is dynamic and unbounded at local and continental scales. *The American Naturalist*, 185(5):584–593, 2015.

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.

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. Community composition exceeds area as a predictor of long-term conservation value. *In preparation*, 2021a.

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, 2021b.

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.

Samraat Pawar. Community assembly, stability and signatures of dynamical constraints on food web structure. *Journal of theoretical biology*, 259(3):601–612, 2009.

Eric R Pianka. Latitudinal gradients in species diversity: a review of concepts. *The American Naturalist*, 100(910): 33–46, 1966.

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.

Daniel L Rabosky and Allen H Hurlbert. Species richness at continental scales is dominated by ecological limits. *The American Naturalist*, 185(5):572–583, 2015.

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.

John W Tukey. Box-and-whisker plots. *Exploratory data analysis*, pages 39–43, 1977.

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.

Katsuhiko Yoshida. Evolutionary dynamics of species diversity in an interaction web system. *Ecological Modelling*, 163(1-2):131–143, 2003.

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 A meeting of models: a brief comparison of the LVMCM and the LSPOM

The LVMCM is a spatially explicit, non-neutral, population-dynamical metacommunity model. It describes the time evolution of all species' population biomasses as a function of local environmental conditions, ecological interactions and dispersal. In contrast, the LSPOM is a spatially and mechanistically implicit, ecologically neutral, probabilistic patch occupancy model. Despite these seemingly irreconcilable structural differences, the LSPOM captures both the dominant processes and the emergent steady-state biodiversity distribution of the LVMCM. Crucially, it's parametric simplicity means the LSPOM can be fitted directly to observation. Elsewhere (O'Sullivan et al., 2021a) we developed a method for extracting spatial and environmental distributions and local species richness records from empirical observations and inputting these directly into LVMCM simulations. Parameters can then be tuned in order to reproduce the best fitting LSPOM mixing rate of the observation, reproducing the

empirical macroecology. In Fig. S4 we show a example of this process. While the simplifying assumptions of the LVMCM–in particular fact that abiotic niche widths are typically kept fixed for all species–mean the simulated OFD deviates slightly from the observation (Benito et al., 2018), it is clear that the neutral LSPOM matches both to a reasonable approximation.

## S1.3 Fitting the LSPOM to data

The steady state of the simple locally saturated patch occupancy model can be solved analytically, and this analytic solution can be fitted to a given OFD allowing estimation of the mixing rate $m$. The steady state OFD of the LSPOM is such that, in a large metacommunity with a total of $Nn$ local populations, the mean number of species occupying $x$ sites is $Nns_x$, where for $m = 0$

$$s_x = \frac{1 - Q(x-1, \lambda - 1)}{x\lambda} \tag{S1}$$

and for $m > 0$

$$s_x = \frac{1}{x\lambda}\left[1 - \left(\frac{m}{\lambda+m}\right)^x - Q(x, \lambda-1) + \exp\left(\frac{(\lambda-1)\lambda}{m}\right)\left(\frac{m}{\lambda+m}\right)^{x-1} Q\left(x, \frac{(\lambda-1)(\lambda+m)}{m}\right)\right], \tag{S2}$$

with $Q(x, y) = \Gamma(x, z)/\Gamma(z)$ denoting the regularised incomplete gamma function. Noting that the total number of species in this metacommunity is $Nn\sum_x s_x$, the definition of $\lambda$ translates to

$$\lambda - 1 = \frac{Nn}{Nn\sum_x s_x} = \frac{1}{\sum_x s_x}. \tag{S3}$$

This is solved numerically by adjusting $\lambda$. R code demonstrating the fitting procedure is available in the accompanying Github repository.

## S1.4 Fitting Pigolotti result to data

... R code demonstrating the how to estimate $\alpha$ and $m$ from metacommunity time series is available in the accompanying Github repository.

# S2 Supplementary figures

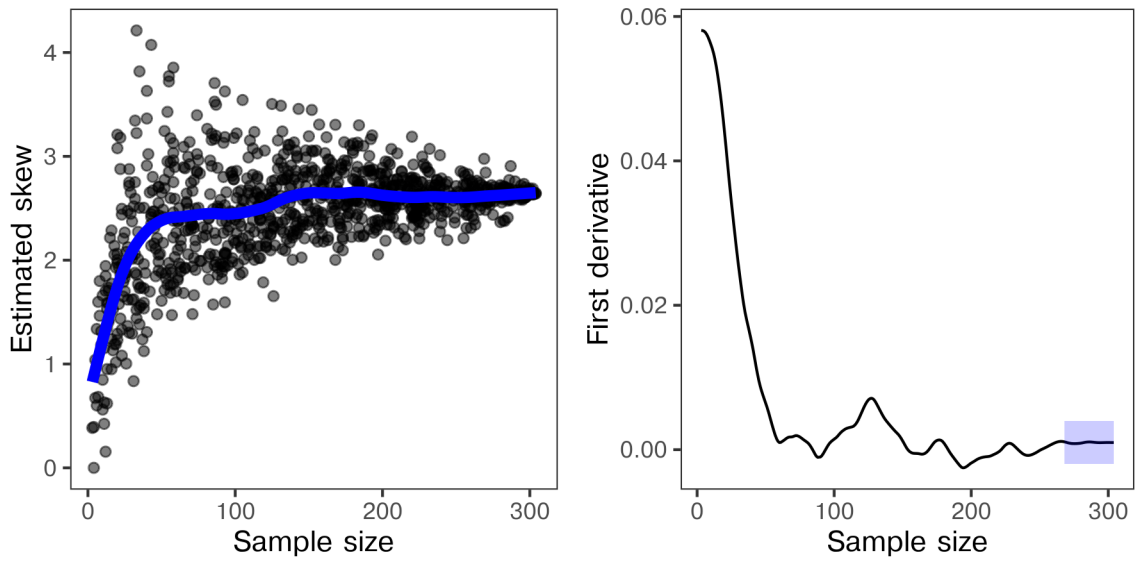


Figure S1: **Testing the accuracy of the skewness estimate for a given catchment-year.** Skewness of the OFD computed after randomly subsampling species to various 'depths' (number of species). The average of the first derivative evaluated over the last 10% of the sub-sample depth (blue box) is taken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

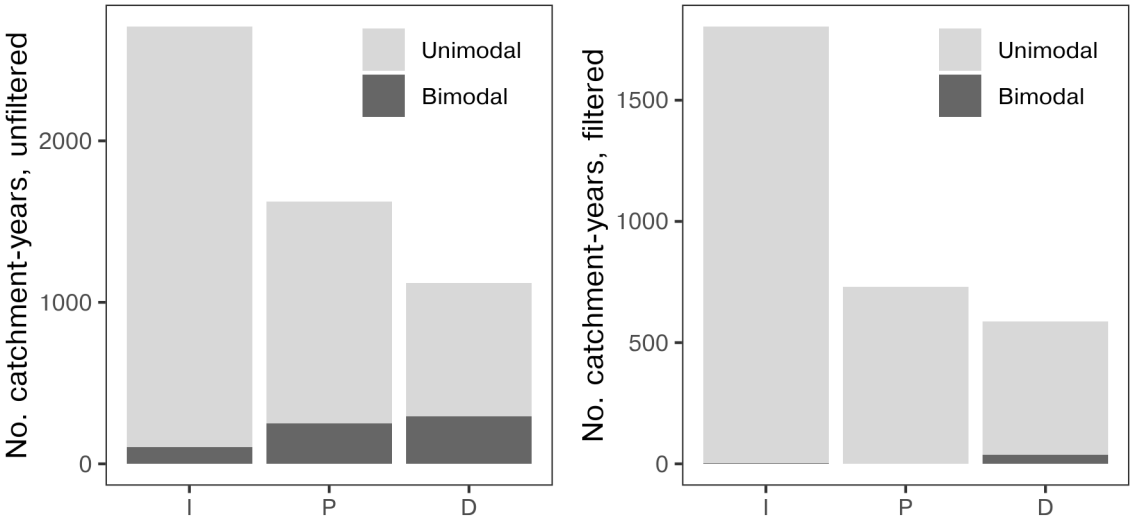


Figure S2: **Catchment-years for which unimodal and bimodal OFDs were detected.** The number of catchment-years for which significant bimodality in the OFD was detected (Tokeshi method, $p < 0.05$) with and without filtering the data.

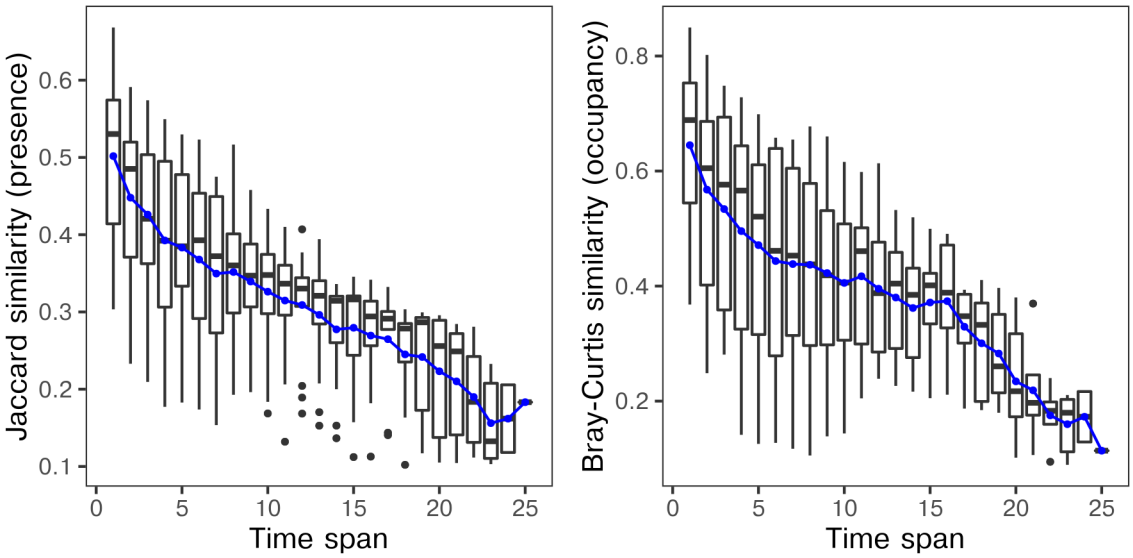


Figure S3: **The decay in compositional similarity with temporal distance.** The decay in macroinvertebrate similarity with regards to species presence-absence and site occupancy at the catchment scale as a function of temporal distance (time span) for a representative catchment (Ouse Upper and Bedford).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

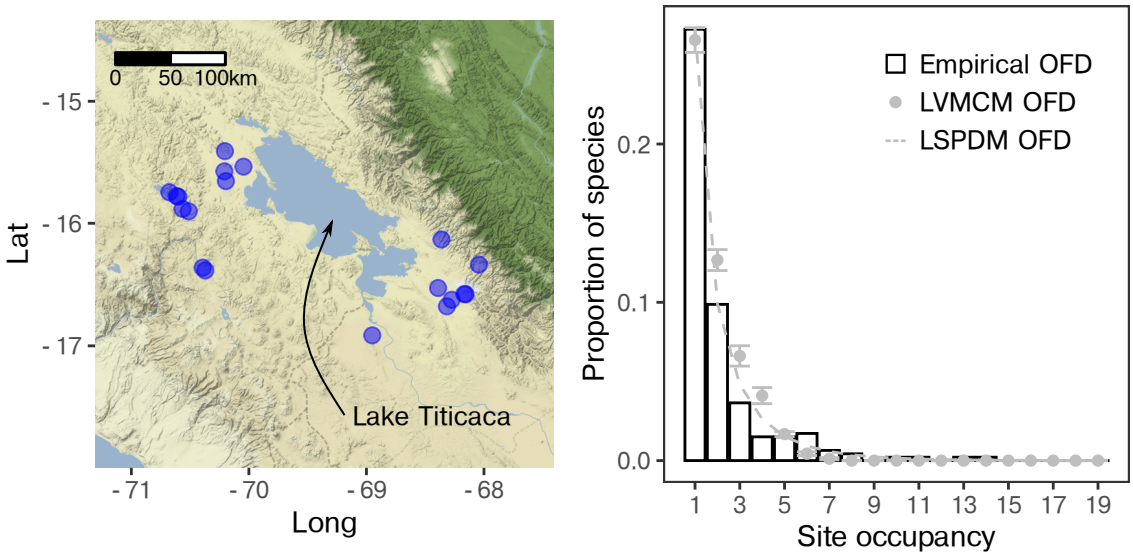


Figure S4: **Comparing the LVMCM and the LSPOM** Taking spatial, environmental and ecological data from a survey of Andean lake diatoms ((Benito et al., 2018), survey sites indicated by blue points), we show how the LSPOM (dashed line) reproduces the OFD of both the observation (bars) and the 'full' LVMCM (points, error bars indicate standard deviation over 10 replicate simulations). As such, the simpler LSPOM can form a useful bridge between the natural ecosystem and the parameter rich simulation (O'Sullivan et al., 2021a).

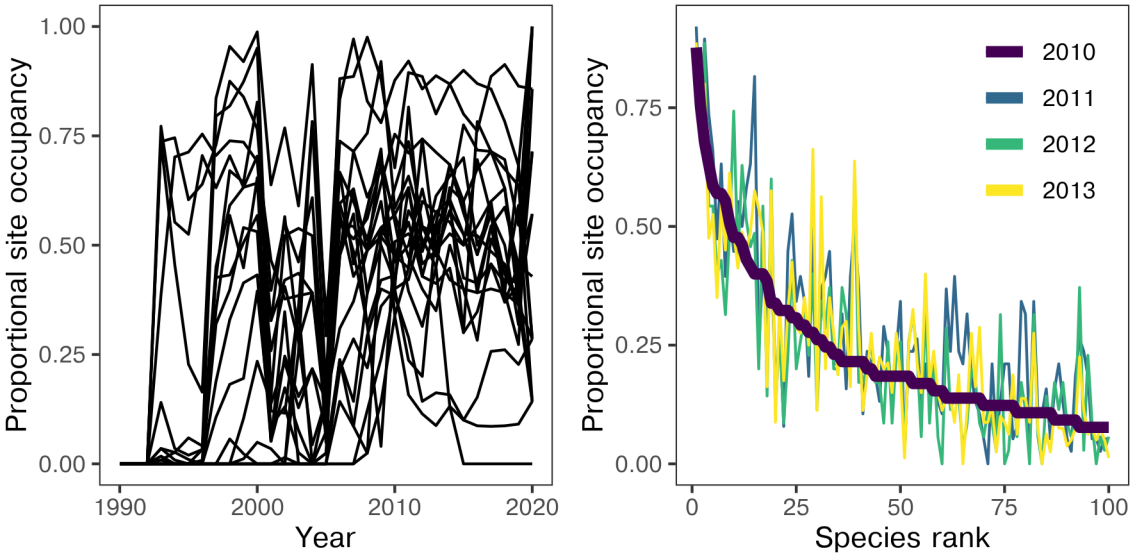


Figure S5: **Species level trends in site occupancy.** A: Proportional site occupancy of the 20 macroinvertebrate species with highest site occupancy in 2010 plotted over the entire time series (Ouse Upper and Bedford). B: Rank occupancy curve of the 100 species with highest site occupancy in 2010 (thick 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.

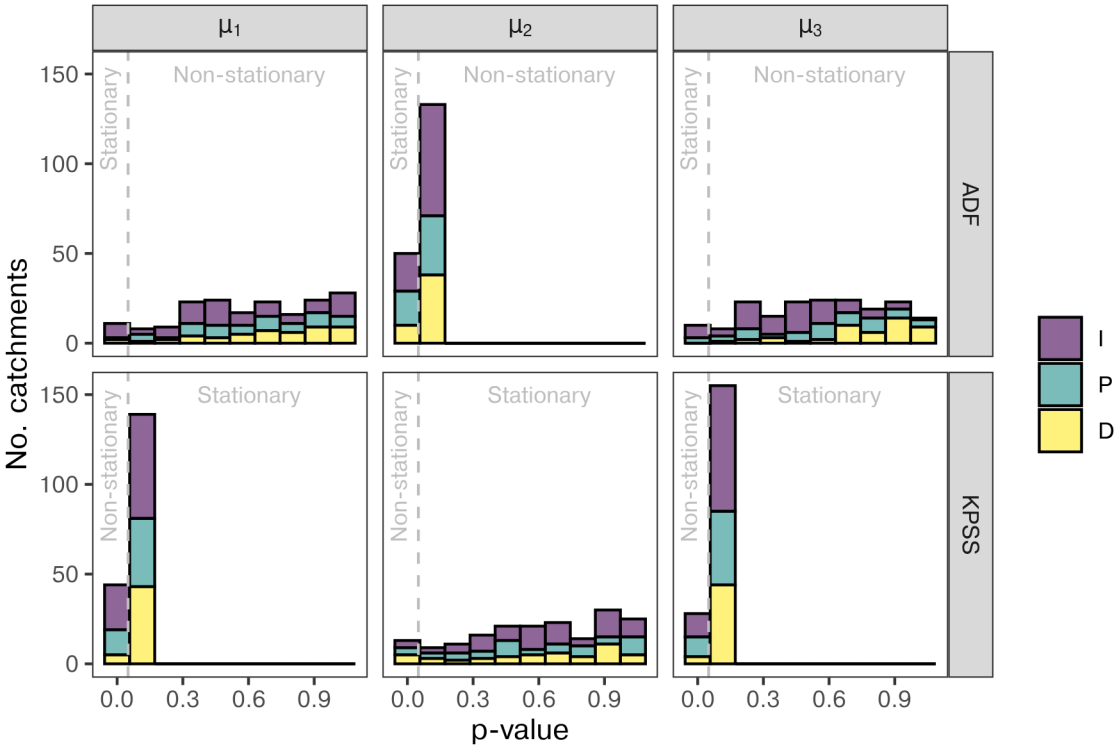


Figure S6: **The significance estimates of the ADF and KPSS tests.** The ADF and KPSS tests applied to the time series of the mean ($\mu_1$), variance ($\mu_2$) and skewness ($\mu_3$) of the catchment OFDs. Shown here are the $p$-values of the tests with $0.05$ indicated by the dashed vertical line. 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 ($\mu_2$) of the OFD.

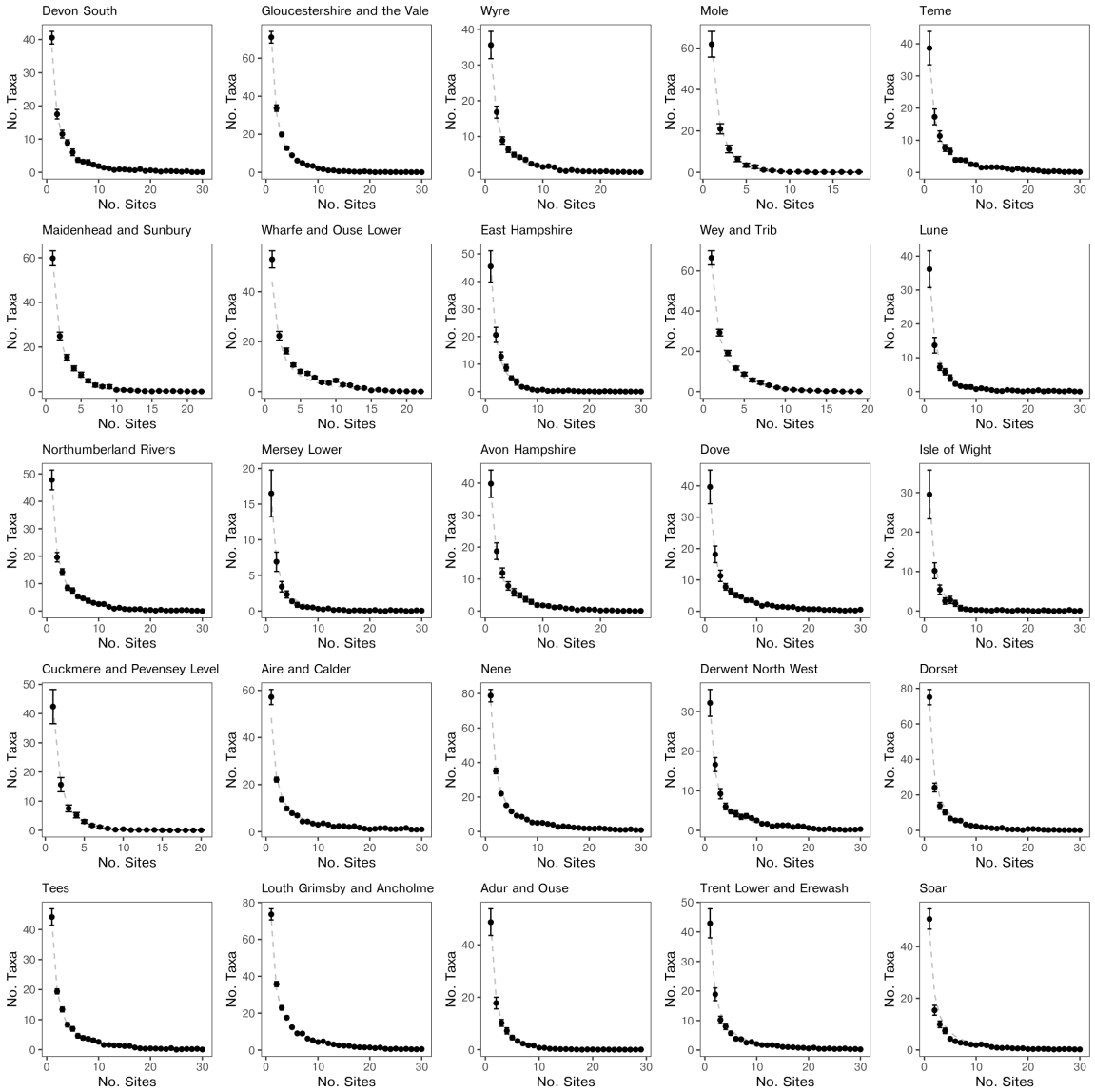


Figure S7: **Time averaged macroinvertebrate site occupancy distributions for 25 random catchments.** The dashed lines represent the fitted LSPOM.

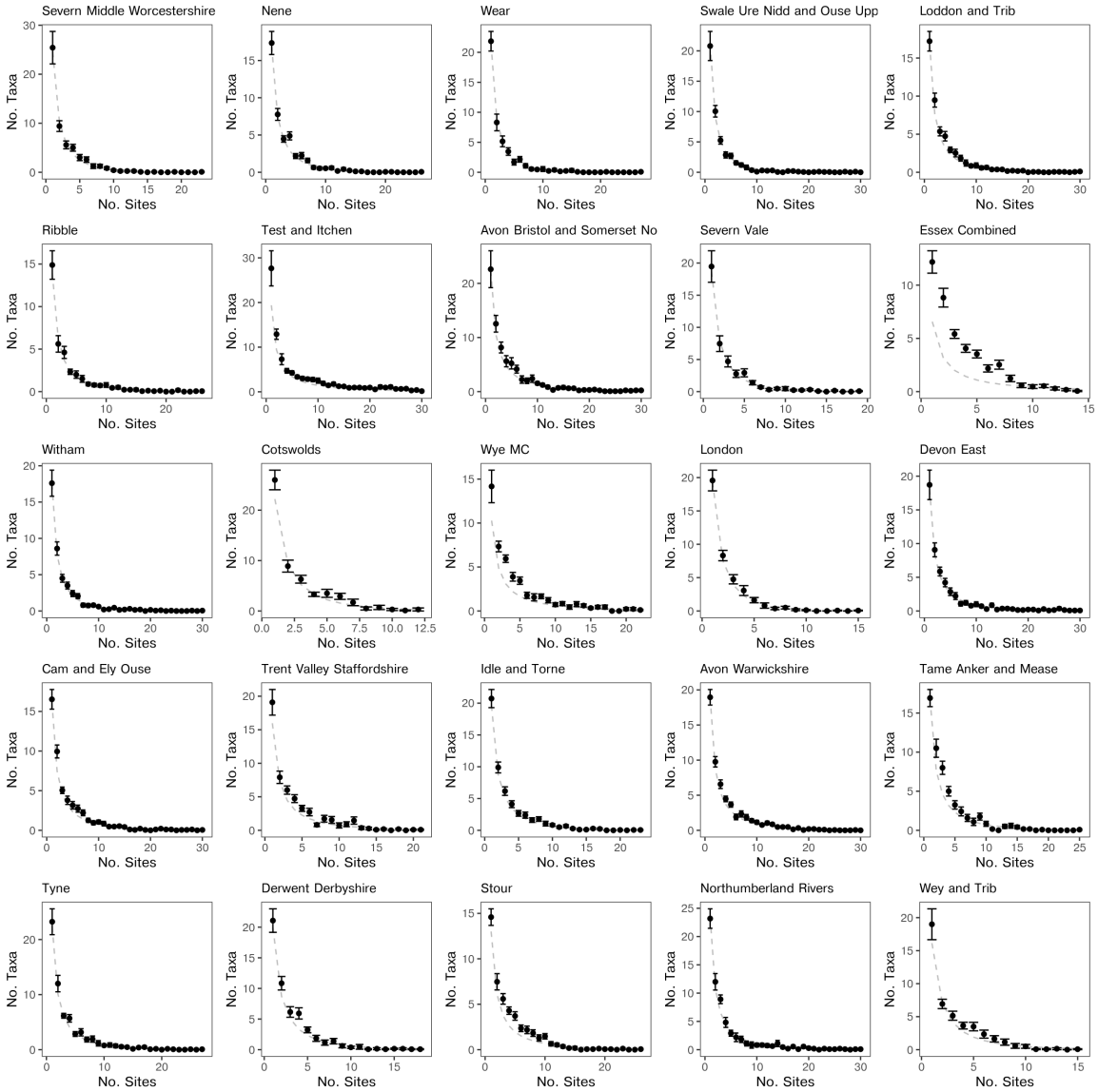


Figure S8: **Time averaged macrophyte site occupancy distributions for** 25 **random catchments.** The dashed lines represent the fitted LSPOM.

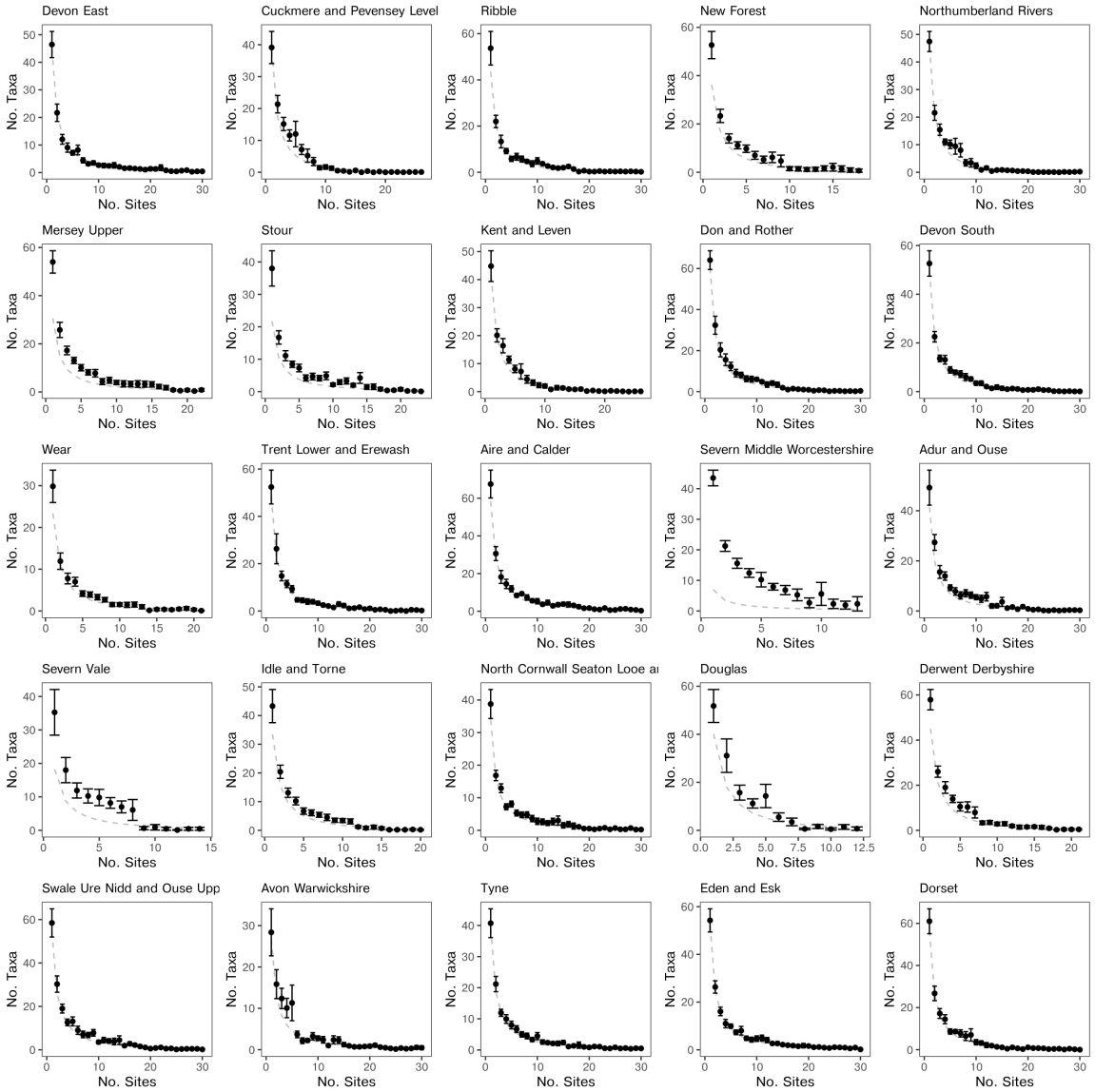


Figure S9: **Time averaged diatom site occupancy distributions for** 25 **random catchments.** The dashed lines represent the fitted LSPOM.

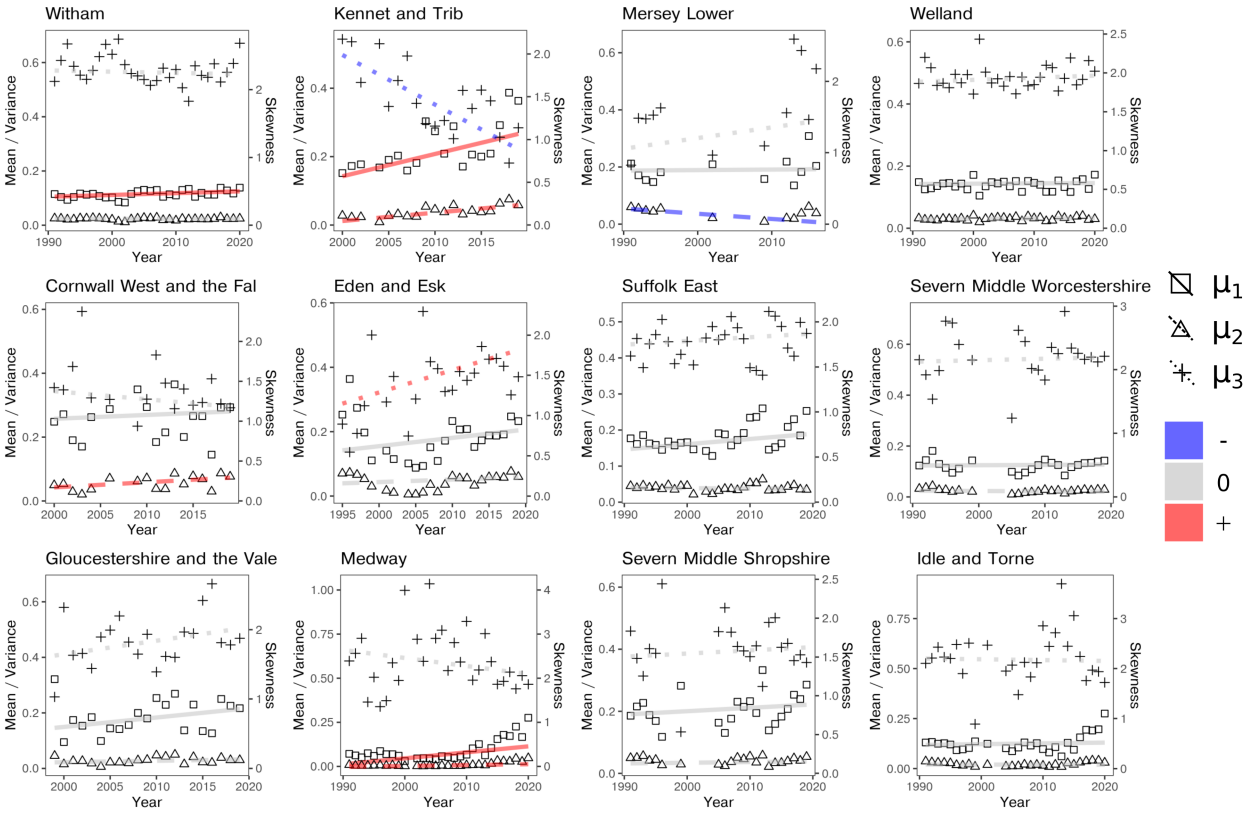


Figure S10: **Macroinvertebrate OFD Moment expansion time series for** 12 **random catchments.** Mean ($\mu_1$), variance ($\mu_2$) and skewness ($\mu_2$) of the OFD plotted against time. The slopes of the regressions are the Sen's slope estimates while the intercepts are defined as the difference between the median skewness and the median year scaled by the estimated slope (Conover, 1999). Colour indicates negative (blue), positive (red) or non-signficant slope assessed using the Mann-Kendall test of temporal trend.

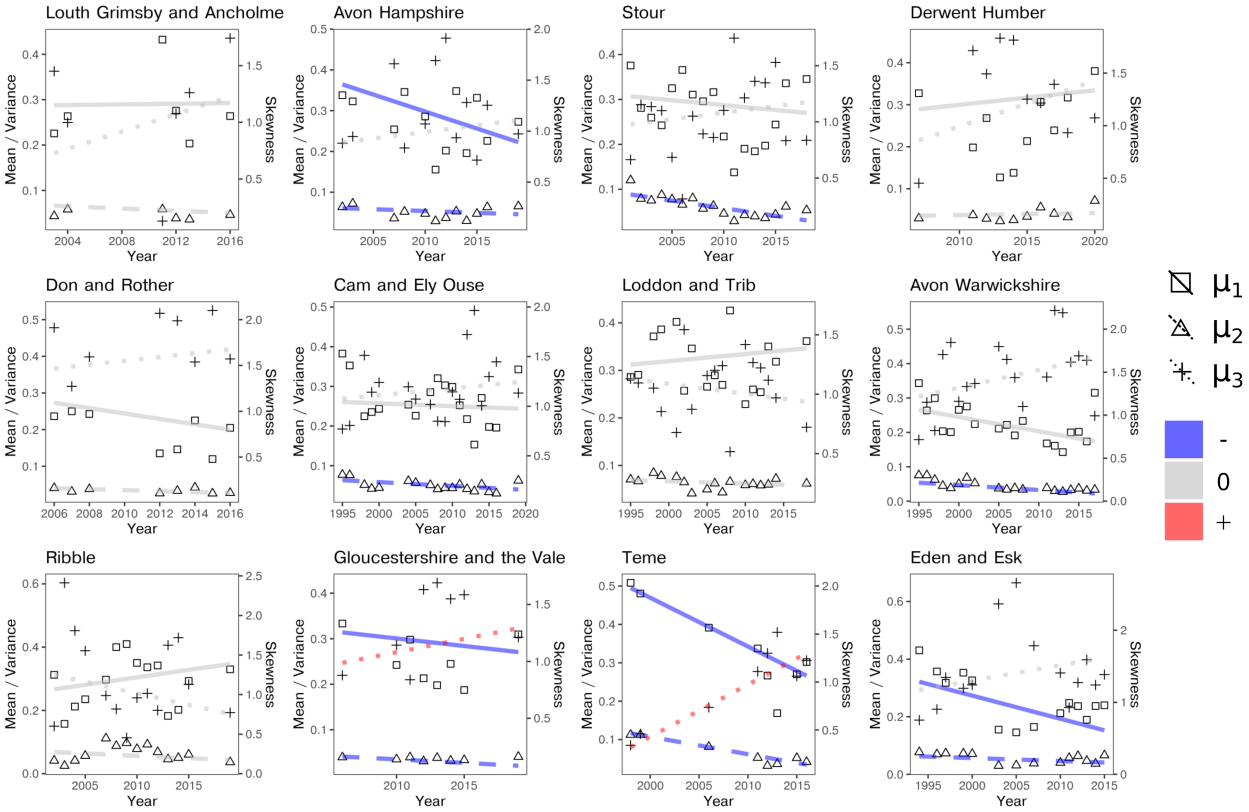


Figure S11: **Macrophyte OFD Moment expansion time series for** 12 **random catchments.** Plotted as in Fig.S10

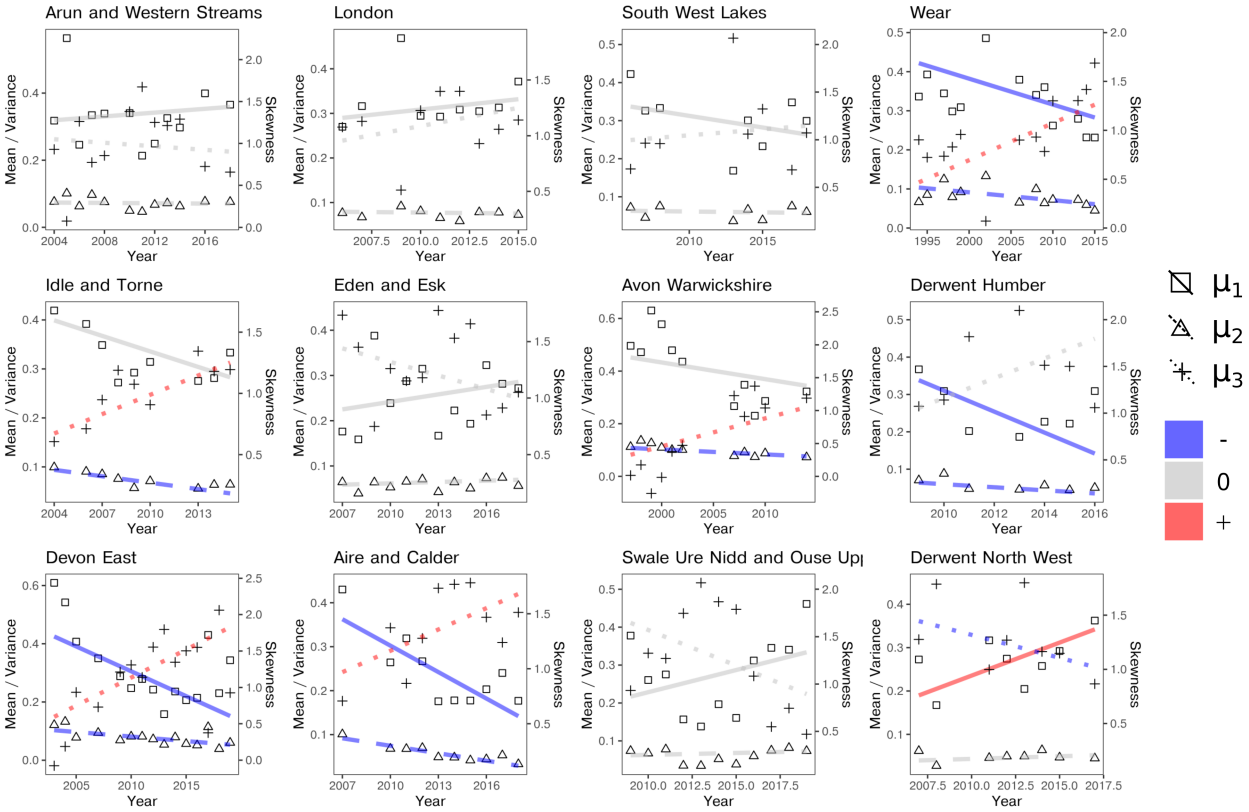


Figure S12: **Diatom OFD Moment expansion time series for** 12 **random catchments.** Plotted as in Fig.S10