Context-Dependent Benchmarks for Ecological Typology Biovalidity Assessment

Jonathan F. Jupke

# 1 Introduction

A tool in the management and research of ecosystems is to divide them into different types: river types, ecoregions, habitat types. These examples serve to highlight the breadth of systems, spatial scales, and methods employed in this exercise. These classifications can be useful for several distinct reasons. They are a easy to use representation of the environmental context and might be considered as a latent, discrete gradient capturing the relevant environmental variables like their continuous counterpart, the ordination axis. This environmental context can then be used to calibrate indicator metrics in biomonitoring (Szoszkiewicz et al. 2019). This is necessary as responses of biota to natural gradients (Alahuhta et al. 2017; Bogotá-Gregory et al. 2023) and anthropocentric stressors (Denison et al. 2021; Pharaoh et al. 2024; Bevan et al. 2024) can differ between ecosystem types. Such context or region-dependence has also been observed for trait-environment relationships (Yang et al. 2018; Burner et al. 2021; Kopp et al. 2023). They are also used as a controlling variable when planning experiments. When we try to minimize the impact certain environmental features, we can focus our sampling to a small selection of types (ref). We often hold different protection goals for different regions of habitats - a typology can then facilitating the selection, combination, or prioritization of management actions. Lastly, they can also serve as a proxy for species ranges, where detailed information is lacking (Pinkert et al. 2023).

Typologies are models. Rather than describing nature, they superimpose their simplified view. A view that can be usefull if they meet the assumption that ecological communities within types are more similar to oneanother than to communities from other types. Generally this property is known as external cluster validity (insert short description), and in this context more specifically as biological validity (ref). With the advent of typology systems in different management schemes (Clean Water act, Water Framework Directive, more examples - also check CWA again.) interest in biovalidation grew and noumerous systems have to date been evaluated against noumerous (new word) taxonomic groups.

Provide some examples.

The available studies on the biovalidity of typology systems suffer from two limitations. First, there is contention which method should be used to determine biovalidity. Different studies use different methods, which reduces comparability among them. It remains difficult to compare the biovalidity of typology systems across taxa and places, because we lack an empirically grounded notion of how biovalid a biovalid typology systems is. In lieu, three practices have taken hold. Biovalidity is judged against unrealistic heuristics without ecological foundations, against statistical significance, or ordination diagrams (e.g., Khanal et al. 2025). Dichotomous decisions based on *p*-values suffer from a slew of issues. Here, we wish to highlight X issues relevent whithin this context. P-value depend on effect size and sample size. Large scale ecological studies typically have a large number of sampling sites, thus rendering even miniscule effect sizes statistically significant. Other ways to dress these issues have been proposed but will not be considered here (Maier and Lakens 2022; Wulff and Taylor 2023; Crouch 2025) Visual assessments of ordination diagrams are not standardized and ad-hoc in nature. More crucially, ordination diagrams only show the variation captured by two or three latent variables and rarely incorporate uncertainty estimates. Model-based ordinations allow the visualization of uncertainty of 95% Credible Intervals (Hoegh and Roberts 2020) or prediction errors (Niku et al. 2019) often severely altering apparent patterns. Illustrative examples are provided here().

Effect size heuristics exist for some commonly used metrics but not all. They are not specifically tailored to ecological systems but describe the mathematical properties of the metrics. The fact that these heuristics are disconnected from ecological realities is evident as they are context-, scale-, and taxon-independent. All three factors should influence the result.

A move toward *a priori* determined relevant effect sizes seems fruitful (Popovic et al. 2024).

We aim to provide a comprehensive set of benchmarks for different biovalidity metrics. These benchmarks will cover different taxonomic groups and different spatial scales. We will train a large number regression models on independent empirical data sets and use these models to simulate new data. The simulated communities will be based on artificial typology systems along a gradient of low to high expected biovalidity, thus enabling us to map plausible ranges for different metrics, spatial scales, and taxa.

# 2 Methods

Add short intro.

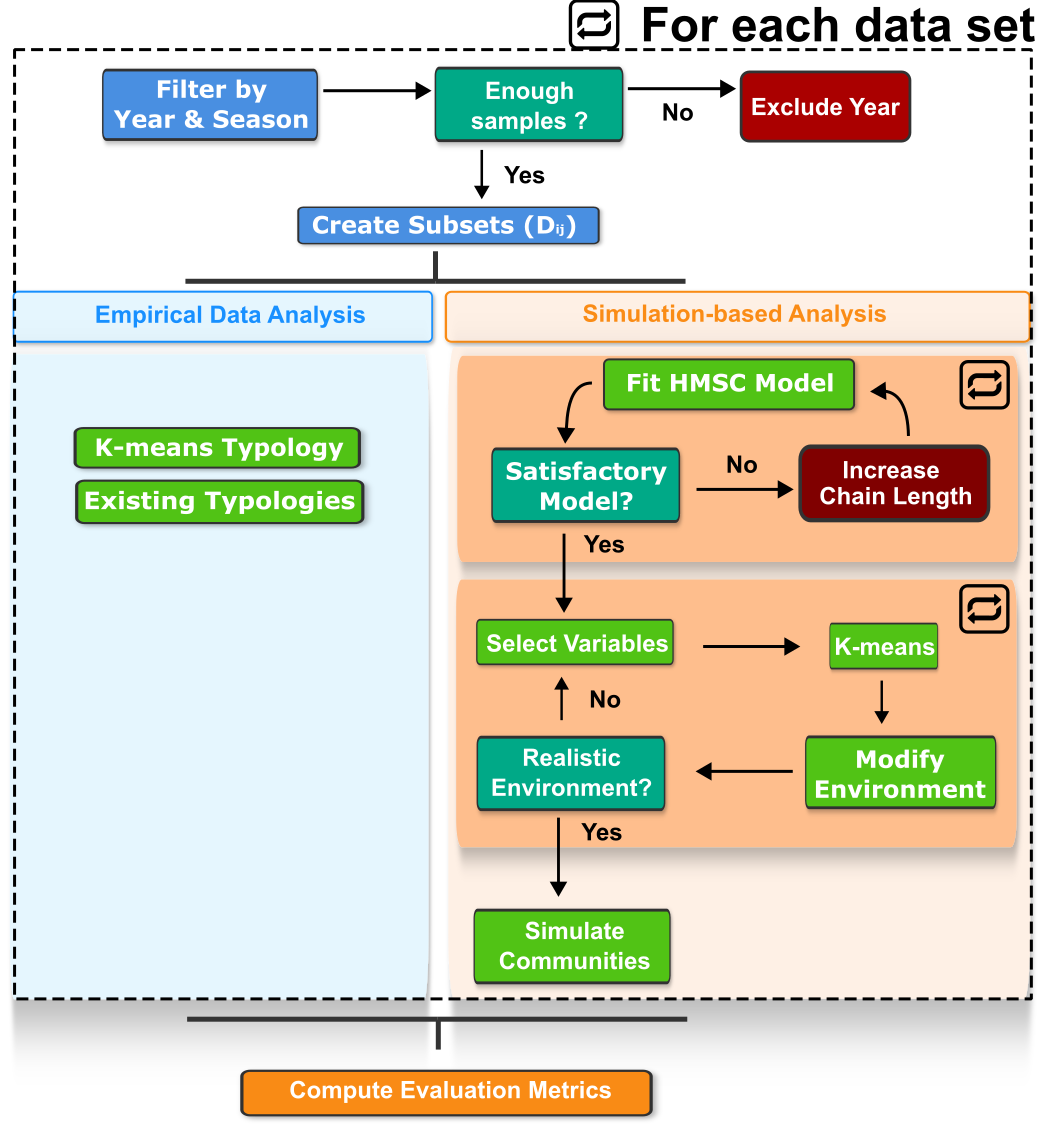


Figure 2.1: placeholder

We have compiled occurrence and abundance data sets of diatoms, aquatic macroinvertebrates, fish, and aquatic macropyhtes (Tables S1-S4, Figure 2.2, Box A Figure 2.1). These data sets range from few samples (19 samples) to many (59044 samples), from moments in time (16 days) to long-term time series (more than 48 years). Most originate form national or federal monitoring campaigns for the Water Framework Directive, though not all. Details on the sampling procedures for every data set are provided in the aforementioned tables. Despite the similar sampling approaches across most data sets, we analyzed each data set separately to minimize potential bias through variations in sampling. We followed two approaches to establish benchmarks for a selection of evaluation metrics. First, we computed a set of evaluation metrics for each of these data sets using a k-means classification of their environments and a selection of pan-European or global typology systems. Second, we greatly expand the number of analyzed typology systems by using simulated occurrence and abundance data sets. These communities were simulated from joint species distribution models that were previously fit on the empirical data sets. We will slightly alter the environmental conditions of these empirical data sets to either improve or diminish the expected performance of typology systems. Evaluating the congruence between simulated data sets and simulated typology systems over many data sets provides us with ecologically plausible benchmarks under a wide variety of circumstances.

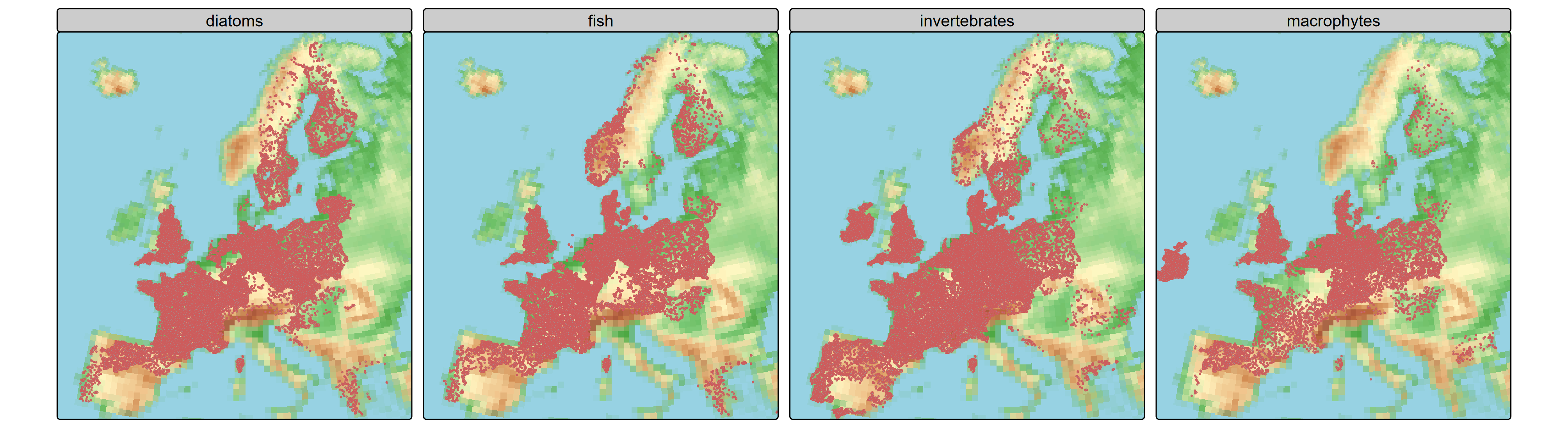


Figure 2.2: Spatial distribution sampling sites across Europe.

2.1 Data preparation

2.2 2.1 Data preparation

For both approaches, within each data set , we performed separate analyses for each year. For each year, we further restricted the data to the three consecutive months with the most samples , lessening the impact of long-term and seasonal trends. The samples from each will be distributed between clusters later in the analysis and we wanted to prevent very small clusters (< 10 samples) from producing extreme results. Since we compare different numbers of clusters for each , we require a minimum sample size of to ensure adequate statistical power. Years with are excluded from the evaluation of data set .

For data sets meeting the minimum threshold, we created subsets using a logarithmic sampling strategy. We generated subsets using a geometric sequence with initial value and multiplier :

Each computed value was rounded to the nearest integer, and duplicates after rounding were removed. We retained only subset sizes , and always included the complete data set as the final subset if not already present in the sequence.

For example, with , this strategy yields seven subsets: with 50 samples, with 75, with 113, with 169, with 254, with 381, and with 564 samples. If abundance data were available, we conducted all analyses on both abundance and occurrence data. To maintain computational feasibility, if the number of subsets exceeded 15, we retained the 15 largest subsets, ensuring the complete data set was always included.

For each , we determined the spatial scale (minimum, median, and maximum spatial distance between sampling locations), position (median, minimum and maximum latitude and longitude), and the taxonomic resolution as fraction of taxa identified to species, genus, family, or higher level.

2.2.1 Modelling

For the first approach we did not perform any modelling. The analyses for this approach continue at section @ref{sec:Evaluating-typologies}. For the second approach, we fit a Hierarchical model of species communities (HMSC, Ovaskainen and Abrego 2020), to each , to establish the relationship between biota and environmental variables, space, biotic interactions, and stochasticity. HMSC is a hierarchical Bayesian latent variable regression model, which is fit with block-conditional Gibbs MCMC sampler. We used a probit residual distribution for occurrence data and a Lognormal Poisson for abundance data. Each model was initially fit with two chains, 2000 MCMC samples, a thinning of two, and a transient period of 4000 samples.  
The model fit was evaluated with the potential scale reduction factor [PSRF; Gelman and Rubin (1992)]. The PSRF compares the variance between multiple chains to the variance within each chain to assess chain convergence. If 10% estimated parameters have a PSRFs of > 1.1 the model will be refit.  
We will iteratively increase the chain length by increments of 1000 until < 10% of PSRFs are > 1.1. Thinning is always the number of MCMC samples divided by 1000, and the length of the transient equals the number of MCMC samples.  
If > 10% of parameters have a PSRF of > 1.1 at a chain length of 10000 MCMC samples, was discarded from further analyses. We also discarded models with sub par predictive performance. This was evaluated in through the Area under the receiver operating characteristic curve (AUROC, Hanley and McNeil 1982) and a series of posterior predictive checks (Gelman et al. 1996; Conn et al. 2018) These posterior predictive checks compared the observed community patterns to patterns generated from 150 posterior draws per fitted model. We used the last 50 draws from each chain. We assessed four diagnostic patterns: (1) beta diversity structure — the distribution of pairwise dissimilarities (Jaccard for occurrence, Bray-Curtis for abundance), (2) species co-occurrence patterns quantified with the C-score metric (Stone and Roberts 1990), (3) species-environment relationships measured as weighted-average optima (ter Braak and Prentice 1988) for the ten most abundant taxa, and (4) species prevalence distributions capturing the commonness-rarity spectrum across all taxa. For each diagnostic, we calculated the proportion of observed values falling outside the 95% envelope of simulated values. Models were flagged for poor fit if three or more diagnostics indicated substantial deviation (|z| > 2, where z is the standardized effect size). Flagged models with AUC < 0.75 or >10% of parameters with PSRF > 1.1 at 10,000 MCMC samples were excluded from benchmark derivation (n = X models excluded). Detailed methods for each posterior predictive check are provided in Supplementary Methods SX.

If the AUROC < 0.75 or the

Likewise, was not used if its predictive quality was bad (AUC < 0.75).

2.2.2 Preparing environmental and spatial predictors

We have compiled and extensive data base of environmental variables relevant for freshwater organisms (Table 1). Each variable was be summarized at the catchment scale. As catchments we used the EU Hydro DEM catchment data base (EEA 2019).

Table 2.1: Environmental variables used to fit and simulate from Hmsc models

| **Variable** | **Data product** | **Reference** |
| --- | --- | --- |
| Rainfall erosivity | GloREDa | Panagos et al. (2023) |
| Soil Organic Carbon | European Soil Database v2.0 | Panagos et al. (2022) |
| Soil pH in Water | EcoDataCube | Witjes et al. (2023) |
| Slope | Hydrography90m | Amatulli et al. (2022) |
| Roughness | Geomorpho90m | Amatulli et al. (2020) |
| Floodplain Area | Potential Flood prone Area | EEA (2020) |
| Mean annual discharge | Copernicus C3S | Berg et al. (2021) |
| Minimum annual discharge | Copernicus C3S | Berg et al. (2021) |
| Saturated Soil Water Content | Soil Hydraulic Properties | Tóth et al. (2015) |
| Upstream catchment area | HydroATLAS | Linke et al (2019) |
| Maximum annual discharge | Copernicus C3S | Berg et al. (2021) |
| Snow Cover | HydroATLAS | Linke et al (2019) |
| Segment Sinuosity | EU Hydro | EEA (2019) |
| Mean Valley Bottom Flatness Index | EcoDataCube | Witjes et al. (2023) |
| Maximum Valley Bottom Flatness Index | EcoDataCube | Witjes et al. (2023) |
| Mean Stream Power Index | Hydrography90m | Amatulli et al. (2022) |
| Maximum Stream Power Index | Hydrography90m | Amatulli et al. (2022) |
| Lake Index | HydroLAKES | Messager et al. (2016) |
| Elevation | Hydrography90m | Amatulli et al. (2022) |
| Groundwater Table Depth | HydroATLAS | Linke et al. (2019) |
| Glaciated Area in Catchment | Randolph Glacier Inventory | RGI Consortium (2017) |
| Precipitation of Wettest Month | CHELSA-BIOCLIM+ | Brun et al. (2022) |
| Precipitation of Driest Month | CHELSA-BIOCLIM+ | Brun et al. (2022) |
| Minimum Temperature Coldest Month | CHELSA-BIOCLIM+ | Brun et al. (2022) |
| Maximum Temperature Warmest Month | CHELSA-BIOCLIM+ | Brun et al. (2022) |
| Bedrock Geology | IHME | Cornu et al. (2013) |
| Catchment soil type | European Soil Database v2.0 | Panagos et al. (2022) |

The effects of spatial proximity were captured with Moran’s Eigenvector Maps (MEM, Dray et al. 2006). MEM transform Euclidean spatial distance matrices into eigenfunctions, giving rise to a large number of spatial eigenfunctions (typically number of samples - 1). Within each model, we used all spatial eigenfunctions that correlate statistically significant (at ) with the residuals of logistic or Poisson regression fit to the single taxa. Before the selection, all *p*-values are adjusted for multiple testing using Holm’s step-down procedure (Westfall and Young 1993). Bini et al. (2009) has shown that this way of selecting spatial eigenfunctions has minimal impact on the estimated regression coefficients for environmental variables.

2.2.3 Modelling species communities

With each successful model, we used variation partitioning to establish the fraction of variation explained by (i) each explanatory variable, (ii) environmental variables in total, (iii) spatial variables in total, and (iv) latent variables.  
The explained variation across environmental variables can be interpreted as the strength of abiotic filtering. Similarly, the explained variation across spatial variables (MEMs) can be interpreted as the strength of spatial processes (dispersal limitation, mass effect). We interpreted the variation explained by latent variables as a signal of biotic interaction, though this is contentious (Dormann et al. 2018; e.g., Blanchet et al. 2020; Vallé et al. 2023), as the latent variables can also capture unmeasured environmental drivers. Tjur’s (Tjur 2009) was used to evaluate the overall explained variation in occurrence models and xxx in abundance models.  
The unexplained variance or , was interpreted as stochasticity in community composition. Each of these values was be averaged across species for each .

2.2.4 Simulating environments

Each fitted model was used to simulate new biotic data sets . For these simulations, we slight adjusted the environmental variables, creating simulated environments, . Let denote the complete set of environmental variables in the original data set , i.e., the explanatory variables in the HMSC models, excluding MEMs. For each , we randomly selected subsets of environmental variables , where was randomly chosen between 3 and . For brevity, we omit the subscripts where context makes them clear (e.g., writing instead of , and instead of ). This partitioned the original environmental variables into two disjoint sets: , the variables used to define simulated types and , the neglected variables. This randomization was performed across multiple iterations , generating different variable selections for each iteration.

Let denote the restriction of to the variables . Each was submitted to a k-means clustering. We tested between two and ten clusters and picked the solution with the highest average silhouette width (ASW, Kauffman and Rousseeuw 1990), which had at least ten samples in the smallest cluster. If all classifications had a group with < 10 samples, we picked the solution with the most classes for which only one class had < 10 samples and balanced the classes by moving the samples from other classes that are closest to the small classes centroid to that class until it had 10 members. Additionally, we compute a fuzzy C means classification (FCM, Dunn 1973) of the same data and the number of centers determined in the k-means classification. Based on this k-means classification, we altered the values of by adjusting the position of cluster centroids relative to each other and by shifting individuals sites closer to or farther way from the centroid of their cluster. The combination of this modified and , represented a simulated environment .

A good typology is separated and compact, i.e., the centroids are far from each other and each site is close to its centroid. By randomly varying these two properties for each , we create a total of 200 for each .

To modify separation, we computed the location of the overall centroid in environmental () space. We then randomly drew a dilation factor where is a uniform distribution. The new centroids were computed as:

To adjust the compactness, we drew a second dilation factor and multiplied the distance between each site location and respective centroid by .

The we added this distance to the new centroids .

The ranges for dilation factors were determined through an exploratory analyses. We tested progressively wider ranges for both and across a subset of data sets and evaluated what proportion of simulated environments passed the Isolation Forest and Mahalanobis distance filters (described below). The ranges and represent the widest intervals that still allowed sufficient simulations (~30% pass rate) to generate the target number of acceptable environments (200 per ) within a computationally feasible number of attempts. See supplementary materials for a short exposition of these analyses.

We were mindful not to create unrealistic non-analog environments. Each was subjected to two sanity checks. We tested whether the simulated combination was similar to existing environmental conditions observed in the same Environmental Zones (Metzger et al. 2005). We chose the Environmental Zones as regionalization because (i) they span Europe, (ii) they are based on variation in environmental rather than biotic conditions , and (iii) they received relatively favorable results in previous studies (Jupke et al. 2023). Most of our data sets spanned multiple Environmental Zones. In these cases, we merged all zones covered by the respective data set. We used the environmental variables form all the EU HydroDEM catchments in a given combination of Environmental Zones () as training data for an Isolation Forest (Liu et al. 2008) to identify implausible environmental combinations. An Isolation Forest is an unsupervised anomaly detection algorithm that identifies outliers by randomly partitioning data through recursive splitting. The method exploits the principle that anomalies, being few and different, require fewer random partitions to be isolated compared to normal points, i.e., catchments. Each catchment receives an anomaly score () based on its average path length across multiple isolation trees, where shorter paths indicate likely anomalies. With the trained forest, we computed the anomaly scores for , i.e., . Then we determined the z-scores for each catchment in , as

A combination of dilation scores () passes the Isolation Forest test if at least 75% of samples have a z-score 1. For the second test we again used all the catchments in the respective environmental zones as a base line. We computed the means and covariance matrix of and then tested the Mahalanobis distance between and the original data. Using a chi-squared test, we evaluate how likely we would have been to observe , given the means and covariances of . A combination of dilation scores passed the Mahalanobis test if at least 90% of samples are within the central 99% of the distribution. If a passed both tests, it was used to simulate biotic communities. We iterated over this procedure until we had 200 for each . To obtain these, created 300 for each . Each is used to dilate the variables , each resulting was evaluated through isolation forest and Mahalanobis distance tests. For each , a maximum number of 30 was used. For each , a maximum was 200 different was evaluated. For s where less than 200 passed both tests, we used all that passed.

The underlying k means and FCM classifications of where again evaluated for each using ASW and the normalized partitioning entropy (Bezdek 2013), respectively. These classifications represent the simulated typologies, which were evaluated in the next step.

2.2.5 Simulating communities

For each validated simulated environment we simulated an abundance or occurrence data set . Simulations were performed with the predict() function from Hmsc on the pooled chains of the accepted model (converged and posterior predictive checks). By default this method produces one predicted data set for each sample in the pooled MCMC-chain.  
We opted to use the last fifty samples from each unpooled chain, i.e., a total of 150 simulated communities., the same number we used for the posterior predicitve checks. We averaged over these 150 communities to obtain one predicted community. For abundance data we took the arithmetic average of each taxon at each site. For occurrence data, we also took the arithmetic mean, which produces a number between 0 and 1. We used this number as the probability when drawing from a Bernoulli distribution to obtain the final simulated, averaged occurrence data set, .

2.2.6 Evaluating typologies

Both original and simulated data sets were evaluated with a section of methods. An extensive and harmonized explanation of all the metrics can be found in the supplementary materials (SX). We evaluated biovalidity with Analysis of Similarities (ANOSIM, Clarke 1993), Classification Strength (CS, Van Sickle 1997), Permutational Multivariate Analysis of Variance (PERMANOVA, Anderson 2001), Indicator Species Analysis Minimizing Intermediate Constancies (ISAMIC, Roberts 2007), and Area under the zeta diversity decline curve (AUC, Jupke et al. 2023). All distance based methods used Jaccard distance matrices for occurrence data and Bray-Curtis for abundance data of . The ANOSIMs were run in a pairwise fashion for each possible combination of types, as suggested in Clarke et al. (2014). As test statistics, we evaluated the minimum, mean, and maximum pairwise ANOSIM R values for each . For PERMANOVA, we evaluated the and the statistic. For the computation of the AUC, we sightly deviated from the original proposal in Jupke et al. (2023). While we still evaluated -diversity for ranks 1 to 10 and still scaled with the diversity (i.e., ), we also divided each by a baseline which is the of inter-type comparisons. For CS and ISAMIC, we evaluated the eponymous test statistics.

These evaluation approaches do not work properly for the fuzzy classification. For these approaches, we computed two additional metrics. Our first approach involves fitting binomial multiresponse GLMs (Wang et al. 2012) using the presence-absence data as response variables and the type-membership as explanatory variables. This method shows how much of the variation in occurrences can be explained by the type-memberships. We will quantify the model fit with the weighted average if explained deviance across all taxa. The deviances are weighted by the total number of taxon occurrences to prevent an undue influence of rare species. The best possible score is a 1, indicating the type-membership is able to perfectly predict the occurrence probability of every species. The worst score is 0, indicating the type-membership is independent of occurrence probabilities of all taxa.

Our second approach consist of comparing the similarity of two communities in biological space (i.e., their Jaccard distance based on community composition) with the distance in type-membership space, i.e., the Jensen-Shannon distance between their type-membership vectors. This comparison will be achieved with a Mantel test (Mantel 1967) which computes the Pearson correlation between values in the distance matrices.

To provide a further baseline, we also evaluated each original data set using the same k-means approach as described above, Illies Freshwater Ecoregions (Illies 1978), the Environmental Zones, GloRiC (Ouellet Dallaire et al. 2019), Freshwater Ecoregions of the World (Abell et al. 2008), EEA Biogeographic Regions (Evans 2005), Hydroecological regions of Europe (Wasson et al. 2007), and the Broad River Types (Lyche Solheim et al. 2019) with the same metrics.

Classification strength, ANOSIM, PERMANOVA will be computed with functions from the vegan package [2.7-1] (Oksanen et al. 2022). will be computed with purpose written code and using the zetadiv package [1.3.0] (Latombe et al. 2018). Indicator values will be computed with indicspecies [1.8.0] (De Cáceres and Legendre 2019) and ISAMIC with labdsv [2.1-2] (Roberts 2007). Some of the computations will be parallelized with the R packages foreach [1.5.2] (Microsoft and Weston 2022) and doParallel [1.0.17] (Corporation and Weston 2022). All code is available in the [associated github repository](https://github.com/JonJup/AquaticTypologyBenchmark).

2.2.7 Deriving benchmarks

To derive context-specific benchmarks from these evaluations, we will use Random Forest regression analyses using the tidymodels R package [1.4.0] (Kuhn and Wickham 2020). We will fit separate random forests for each evaluation metric and taxon. In total, we will fit 80 random forest models (four taxa and 20 evaluation metrics). The key innovation of this approach is that benchmarks are conditional predictions: for a given combination of system properties (taxonomic group, spatial scale, sample size, ecological structure), the model predicts the expected range of biovalidity metric values. The random forest model will be implemented with Ranger [] with three hyperparameters subject to optimization: the number of trees, the number of variables randomly sampled at each split, and the minimum node size. We will employ a grid search strategy for hyperparameter tuning, using a grid for each parameter. The data will be split into training (75%) and testing (25%) sets. Ten-fold cross-validation will be applied to the training set to evaluate model performance across hyperparameter combinations.   
The random forest models include three categories of predictors:  
1. Typology characteristics: number of variables considered in the artificial typology, cumulative importance of these variables (from variation partitioning), average silhouette width of environmental variables, dilation of points relative to centroids, dilation of centroids relative to each other  
2. Sampling design: minimum, maximum, mean, and median spatial distance among samples (capturing spatial scale), number of samples (statistical power), taxonomic resolution of the data set (fraction identified to species vs. higher levels)  
3. Ecological structure: fraction of variation explained by environmental variables, spatial processes, biotic interactions, and stochasticity. These capture fundamental differences in community assembly - systems dominated by environmental filtering should show different biovalidity patterns than those structured primarily by dispersal limitation or biotic interactions.

As predictors the model will use the number of variables considered in the artificial typology, the cumulative importance of these variables (according to the variation partitioning), the average silhouette width of the environmental variables in the typology, the dilation of points relative to their centroid, the dilation of centroids relative to each other, the minimum, maximum, mean, and median distance among samples, the number of samples, the fraction of variation explained by environmental variables, space, biotic interactions, and stochasiticity, and the taxonomic resolution of the data set. Following hyperparameter tuning, we select the model configuration that minimized root mean square error (RMSE) across folds. The final model will be fitted to the complete training set and evaluated on the held-out test set. Permutation-based variable importance will be specified to quantify predictor contributions. To interpret the fitted models, we generated two types of visualizations: (1) variable importance plots using permutation-based importance scores to identify the most influential predictors, and (2) partial dependence plots for the top predictors to visualize the marginal effect of each variable on the predicted response while holding other variables constant. This fitted model will be able to provide benchmarks.

2.2.8 Software

computed with the parallelDist R package [0.2.6] (Eckert 2022). All analyses were run with (todo: ist with das richtige Wort?) Singularity containers (Kurtzer et al. 2017). We used R version … and Python version …. (todo add version). The complete code and the container are available under … (todo add link). The Hmsc models were fit with Hmsc-HPC (Rahman et al. 2024) [version] with the sf R package [1.0-21] (Pebesma 2018) (Simonsohn and Gruson 2025) to use the package versions that were available on the 1st of December 2024. These are also the versions that are referenced throughout the text. Multiresponse GLMs were implemented with mvabund [4.2.1] (Wang et al. 2022), Checks of posterior distributions were conducted with coda [0.19-4.1] (Plummer et al. 2006). Data wrangling will be done with data.table [1.16.2] (Barrett et al. 2024). Jenson-Shannon distance will be computed with philentropy [0.9.0] (Drost 2018). The isolation forest was implemented with isotree [0.6.1-4] (Cortes 2025). MEM will be implemented with the adespatial R package [0.3-2.4] (Dray et al. 2024). as implemented in the coda package [0.19-4.1] (Plummer et al. 2006). The FCM was implemented with the vegclust [2.0.3] R package (De Cáceres et al. 2013).

Consider a set of sites classified into distinct types , where represents the complete typology system. Let be a site and the function that returns its type assignment. For any site , let represent the mean dissimilarity between and all other sites within its assigned type . For any other type , let denote the mean dissimilarity between and all sites in type . The neighbor type of , denoted as , is defined as the type (excluding ) that minimizes the mean dissimilarity:

The silhouette width quantifies how well fits within its assigned type compared to its neighbor type. For site , it is computed as:

The Average Silhouette Width (ASW) is calculated as the mean of individual silhouette widths across all sites:

The ASW ranges from -1 to 1, where -1 indicates the poorest possible type assignments and 1 represents optimal type assignments.

The Classification Strength (CS, Van Sickle 1997) provides another measure to evaluate typology systems by comparing similarities within and among types. Given a similarity matrix with elements where and index over sites, we define the within type similarity for type T as:

where if sites and both belong to type (i.e., ), and otherwise. represents the total number of pairwise comparisons within type , ensuring represents the mean similarity among sites of type . The among-type similarity for type is defined analogously:

where is the number of comparisons between sites in type and sites in all other types The type-specific classification strength is calculated as:

This measure quantifies how much more similar sites are to others within their assigned type compared to sites in different types. The difference to SW is, that CS considers all other types and not only the neighbor. The Classification Strength for the complete typology system is computed as a weighted average of type-specific Classification Strengths:

where is the number of sites in type and is the total number of sites. Similar to the ASW, CS values range from -1 to 1, with higher values indicating more distinct and well-defined types.

The Analysis of Similarities (ANOSIM, Clarke 1993) evaluates typology systems using ranked similarities rather than raw similarity values. Let be the matrix of ranked similarities derived from , where lower ranks are assigned to higher similarities. For any pair of sites , the element represents the rank of the corresponding similarity . We define as the mean rank of within-type similarities and as the mean rank of between-type similarities. The ANOSIM R statistic is computed as:

where Like ASW and CS, the ANOSIM R statistic ranges from -1 to 1. A permutational *p*-value for R can be computed.

Permutational multivariate analysis of variances (PERMANOVA, Anderson 2001) evaluates typology systems by analyzing the variance in dissimilarities among sites. For a dissimilarity matrix with elements , we compute two sums of squares: The total sum of squares:

The within-type sum of squares:

The pseudo F-ratio is then computed as:

Larger F-ratios indicate greater separation between type centroids in multivariate space. The statistical significance of the F-ratio can be assessed through a permutation test.

Non-geometric measures evaluate typology systems directly from abundance or occurrence data, without requiring similarity metrics. These include the indicator species analysis (IndVal, Dufrêne and Legendre 1997), Indicator Species Analysis Minimizing Intermediate Constancies (ISAMIC, Roberts 2023), and Area Under the ζ-diversity Decline Curve (AUCζ, Jupke et al. 2023).  
 The indicator species analysis combines two components for each taxon in type : concentration () and relative frequency (). A perfect indicator (IndVal = 1.0) occurs exclusively in one type (perfect concentration) and is present in all samples of that type (maximum frequency). For the group-equalized IndVal variant, which accounts for uneven sample sizes across types (De Cáceres et al. 2010), we compute the concentration component as:

where is the group-equalized sum of abundances in type , and is the total group-equalized abundance. The frequency component as:

where is the number of occurrences of taxon in type , and is the total number of sites in type . The group-equalized IndVal for taxon in type is then:

For each species, only the highest IndVal is considered.

Statistical significance of each IndVal is assessed through a permutation test. Since multiple tests are performed (one for each taxon), the resulting p-values are adjusted using Holm’s step-down procedure to control for multiple comparisons (Westfall and Young 1993). Two metrics evaluate the overall quality of the typology system: 1. The proportion of significant indicator taxa (at ), denoted as :

where is the indicator function, the number of species and is the adjusted -value for taxon . 2. The mean adjusted p-value across all taxon-type combinations:

A superior typology system is characterized by a higher and a lower .

The Indicator Species Analysis Minimizing Intermediate Constancies (ISAMIC) evaluates how consistently species occur within types. For each taxon and type , we calculate:

The metric quantifies the predictability of species occurrence patterns. Values close to 1 indicate that species tend to be either consistently present or consistently absent within types Values close to 0 indicate that species occurrences are unpredictable, with frequencies close to 0.5 Higher ISAMIC values suggest a more effective typology system.

The Area Under the Zeta diversity Curve (AUCζ, Jupke et al. 2023) quantifies the rate at which the number of shared taxa decreases along an increasing number of sampling sites. The zeta (ζ) diversity of degree represents the mean number of species shared across sites (Hui and McGeoch 2014).

where represents the set of species in site . For any typology system, we compute the ζ-diversity within-types using sites from the same type and the ζ-diversity between-types using sites from different types. As ζ-diversity declines monotonically with increasing , we can compute the area under the decline curves to quantify the speed and magnitude of turnover. The AUCζ for is calculated as:

We divide by to obtain results that are independent of species richness. The effectiveness of a typology system can be quantified by the ratio:

where values > 1 indicate lower turnover within types than between types.

Lastly, we will compute one evaluation metric for a fuzzy classification. Using gradual instead of binary type-membership, such classifications offer advantages for capturing ecological gradients but are less commonly used. The evaluation criteria we have discussed thus far assume crisp classifications and cannot be applied to fuzzy classifications. Therefore, we will use two approaches specific to fuzzy classifications. To our knowledge, these approaches have not been previously applied to bio validation of typology systems.

Our first approach involves fitting binomial multiresponse GLMs (Wang et al. 2012) using the presence-absence data as response variables and the type-membership as explanatory variables. This method shows how much of the variation in biota can be explained by the type-memberships. We will quantify the model fit with the weighted average if explained deviance across all taxa. The deviances are weighted by the total number of taxon occurrences to prevent an undue influence of rare species. The best possible score is a 1, indicating the type-membership is able to perfectly predict the occurrence probability of every species. The worst score is 0, indicating the type-membership is independent of occurrence probabilities of all taxa.

Our second approach consist of comparing the similarity of two communities in biological space (i.e., their Jaccard distance based on community composition) with the distance in type-membership space, i.e., the Jensen-Shannon distance between their type-membership vectors. This comparison will be achieved with a Mantel test (Mantel 1967) which computes the Pearson correlation between values in the distance matrices.

# 3 Supplementary Methods SX: Posterior Predictive Validation of HMSC Models –>

3.1 Overview

Posterior predictive checks assess whether models can reproduce key ecological patterns observed in empirical data (Gelman et al. 1996). Rather than relying solely on aggregate fit statistics (e.g., AUC), we evaluated whether fitted HMSC models reproduced four fundamental community patterns: beta diversity structure, species co-occurrence, species-environment relationships, and prevalence distributions. For each fitted model, we generated 150 simulated data sets by sampling from the posterior distribution of model parameters. Specifically, we selected the last 50 samples samples from each MCMC chain.

3.1.1 Beta Diversity Structure

We evaluated whether models reproduced the observed distribution of pairwise dissimilarities. For each data set (observed and simulated), we computed all pairwise community dissimilarities using Jaccard distance for occurrence data or Bray-Curtis distance for abundance data. We then compared the 5th, 25th, 50th, 75th, and 95th percentiles of their distributions.

For each quantile, we calculated the 95% envelope across the simulated data sets (i.e., the 2.5th and 97.5th percentiles of simulated quantile values). Models were flagged if any observed quantile fell outside its corresponding envelope. Additionally, we computed a standardized effect size for the median dissimilarity

. This metric indicates whether the model systematically over- or under-estimates compositional turnover.

3.1.2 Species Co-occurrence Patterns

Co-occurrence patterns reflect biotic interactions, shared environmental preferences, and assembly history (Gotelli 2000). The C-score quantifies the prevalence of “checkerboard” patterns—species pairs that rarely co-occur at sites (Stone and Roberts 1990).

For each data set, we calculated the C-score across all species pairs with occurrences. For species and , the checkerboard score is

, where and are the number of sites each species occupies, and is the number of sites where both occur. The C-score is the mean of across all species pairs. Higher C-scores indicate more checkerboard patterns (potential competitive exclusion); lower scores suggest aggregation (positive associations).

We calculated the C-score for observed data and each of the 150 simulated data sets, then computed the standardized effect size:

Models were flagged if |z| > 2.

3.1.3 Species-Environment Relationships

Even if community-level predictions are adequate, individual species may show incorrect environmental responses. We evaluated whether the ten most abundant species exhibited consistent environmental optima between observed and simulated data.

For each species and environmental variable , we calculated the abundance-weighted mean: c\_he = Σ(a\_ih × e\_i) / Σ(a\_ih)

where is the abundance of species at site , and is the value of environmental variable at site . For presence-absence data, . This yields the “center of gravity” of each species along each environmental gradient.

For each species-environment combination in the observed data, we calculated the corresponding optimum in each of the 150 simulated data sets. We then computed the standardized effect size

Individual relationships were flagged if . We summarized model performance as the proportion of species-environment combinations flagged. Models with >20% of relationships flagged were considered poorly calibrated.

3.1.4 Species Prevalence Distribution

The prevalence distribution — how many sites each species occupies — reflects fundamental macroecological patterns (Gaston and Blackburn 2000). Most communities show right-skewed distributions with many rare species and few ubiquitous species. Models that fail to reproduce this pattern may inadequately capture ecological ground truth.

For each data set, we calculated prevalence for each species as the proportion of sites occupied. We then summarized the prevalence distribution with five statistics: (1) number of rare species (prevalence < 0.05), (2) number of uncommon species (0.05 ≤ prevalence < 0.25), (3) number of common species (0.25 ≤ prevalence < 0.75), (4) number of ubiquitous species (prevalence ≥ 0.75), and (5) distribution skewness.

For each statistic, we calculated standardized effect sizes comparing observed to simulated values. Models were flagged if the number of rare species or ubiquitous species deviated substantially () from simulated distributions.

3.1.5 Decision Rules

For each model, we tallied the number of flagged diagnostics (0-4). Models with any flagged diagnostics were considered inadequate for generating communities. These models were excluded from benchmark derivation.

3.1.6 Software Implementation

All posterior predictive checks were implemented in R version [X.X.X]. Beta diversity calculations used the vegan package [2.7-1] (Oksanen et al. 2022). C-scores were computed with custom functions adapted from EcoSimR (Gotelli et al. 2015). Posterior sampling from HMSC models used the predict() function from Hmsc [3.3-7] (Tikhonov et al. 2022). Code for all diagnostics is available in the associated GitHub repository (<https://github.com/JonJup/AquaticTypologyBenchmark>).

3.2 Notation Guide

|  |  |  |
| --- | --- | --- |
| Symbol | Description | Example/Notes |
| **Data Structure** |  |  |
|  | a Data set | Individual monitoring dataset |
|  | Data set , year | Data from one dataset in a specific year |
|  | Data set , year , subset | Specific sample size subset (e.g., 100, 200, 300 samples) |
|  | Sample size | Number of samples in |
|  | Simulated biotic data | Community occurrence/abundance matrices |
| **Environmental Data** |  |  |
|  | Complete set of environmental variables |  |
|  | Number of environmental variables | Total available predictors |
|  | Original environmental data | Empirical environmental conditions |
|  | Simulated environmental data | Modified environmental conditions |
|  | Selected environmental variables | Variables used to define typology |
|  | Neglected environmental variables | Complement set; variables not used in typology |
|  | Environmental Zone reference data | All catchments within relevant Environmental Zones |
|  | Subset of variable indices | Randomly selected for each iteration |
|  | Size of subset | Number of variables in the typology (3 to ) |
| **Dilation and Clustering** |  |  |
|  | Cluster centroid | Mean position of sites in a cluster |
|  | Overall centroid | Mean position across all clusters |
|  | Centroid of cluster | Position of specific cluster center |
|  | Modified centroid | After separation dilation |
|  | Separation dilation factor | ; moves centroids apart/together |
|  | Compactness dilation factor | ; adjusts within-cluster spread |
|  | Dilation score combination |  |
|  | Modified distance | Distance from site to modified centroid |
| **Iteration and Indexing** |  |  |
|  | Site index | Individual sampling location () |
|  | Cluster/type index | Individual cluster or type () |
|  | Subset index | Index for sample size subsets |
|  | Iteration index | Simulation iteration () |
|  | Number of iterations | Total simulation iterations (target: 200) |
|  | Taxon index | Individual species () |
|  | Number of taxa | Total species richness |
|  | Total number of sites | Across all types |

**Note on subscript hierarchy:** Single subscripts denote specific instances (e.g., = dataset ), double subscripts add temporal detail (e.g., = dataset in year ), and triple subscripts add further subdivision (e.g., = dataset , year , subset ). Superscripts denote modifications (e.g., ) or complements (e.g., ).

# 4 Acknowledgements

# References

Abell R, Thieme ML, Revenga C, et al (2008) Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. BioScience 58:403–414. <https://doi.org/10.1641/B580507>

Alahuhta J, Ecke F, Johnson LB, et al (2017) A comparative analysis reveals little evidence for niche conservatism in aquatic macrophytes among four areas on two continents. Oikos 126:136–148. <https://doi.org/10.1111/oik.03154>

Amatulli G, Garcia Marquez J, Sethi T, et al (2022) Hydrography90m: A new high-resolution global hydrographic dataset. Earth System Science Data 14:4525–4550. <https://doi.org/10.5194/essd-14-4525-2022>

Amatulli G, McInerney D, Sethi T, et al (2020) Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. Scientific Data 7:162. <https://doi.org/10.1038/s41597-020-0479-6>

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecology 26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>

Barrett T, Dowle M, Srinivasan A, et al (2024) Data.table: Extension of ‘data.frame‘

Berg P, Photiadou C, Bartosova A, et al (2021) [Hydrology related climate impact indicators from 1970 to 2100 derived from bias adjusted European climate projections](https://doi.org/10.24381/cds.73237ad6)

Bevan PA, Ferreira GB, Ingram DJ, et al (2024) Regional Biomes outperform broader spatial units in capturing biodiversity responses to land-use change. Ecography e07318. <https://doi.org/10.1111/ecog.07318>

Bezdek JC (2013) Pattern Recognition with Fuzzy Objective Function Algorithms. Springer Science & Business Media

Bini LM, Diniz-Filho JAF, Albaladejo RG, et al (2009) Coefficient shifts in geographical ecology: An empirical evaluation of spatial and non-spatial regression. Ecography 32:193–204. <https://doi.org/10.1111/j.1600-0587.2009.05717.x>

Blanchet FG, Cazelles K, Gravel D (2020) Co-occurrence is not evidence of ecological interactions. Ecology Letters 23:1050–1063. <https://doi.org/10.1111/ele.13525>

Bogotá-Gregory JD, Jenkins DG, Lima FCT, et al (2023) Geomorphological habitat type drives variation in temporal species turnover but not temporal nestedness in Amazonian fish assemblages. Oikos e09967. <https://doi.org/10.1111/oik.09967>

Brun P, Zimmermann NE, Hari C, et al (2022) CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometre-resolution

Burner RC, Stephan JG, Drag L, et al (2021) Traits mediate niches and co-occurrences of forest beetles in ways that differ among bioclimatic regions. Journal of Biogeography 48:3145–3157. <https://doi.org/10.1111/jbi.14272>

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Austral Ecology 18:117–143. <https://doi.org/10.1111/j.1442-9993.1993.tb00438.x>

Clarke KR, Gorley RN, Somerfield PJ, Warwick RM (2014) Change in Marine Communites: An Approach to Statistical Analysis and Interpretation, 3rd edn. PRIMER-E, Plymouth

Conn PB, Johnson DS, Williams PJ, et al (2018) A guide to Bayesian model checking for ecologists. Ecological Monographs 88:526–542. <https://doi.org/10.1002/ecm.1314>

Consortium R (2017) Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0: Technical Report, Global Land Ice Measurements from Space, Colorado, USA. Digital Media.

Cornu J-F, Eme D, Malard F (2013) The distribution of groundwater habitats in Europe. Hydrogeology Journal 21:949–960. <https://doi.org/10.1007/s10040-013-0984-1>

Corporation M, Weston S (2022) doParallel: Foreach parallel adaptor for the ’parallel’ package

Cortes D (2025) [Isotree: Isolation-based outlier detection](https://doi.org/10.32614/CRAN.package.isotree)

Crouch DJM (2025) The false evidence rate: An approach to frequentist error rate control conditioning on the observed *P* value. Proceedings of the National Academy of Sciences 122:e2415706122. <https://doi.org/10.1073/pnas.2415706122>

De Cáceres M, Legendre P (2019) Indicspecies

De Cáceres M, Legendre P, He F (2013) Dissimilarity measurements and the size structure of ecological communities. Methods in Ecology and Evolution 4:1167–1177. <https://doi.org/10.1111/2041-210X.12116>

De Cáceres M, Legendre P, Moretti M (2010) Improving indicator species analysis by combining groups of sites. Oikos 119:1674–1684. <https://doi.org/10.1111/j.1600-0706.2010.18334.x>

Denison CD, Scott MC, Kubach KM, Peoples BK (2021) Integrating Regional Frameworks and Local Variability for Riverine Bioassessment. Environmental Management 68:126–145. <https://doi.org/10.1007/s00267-021-01479-6>

Dormann CF, Bobrowski M, Dehling DM, et al (2018) Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. Global Ecology and Biogeography 27:1004–1016. <https://doi.org/10.1111/geb.12759>

Dray S, Bauman D, Blanchet G, et al (2024) adespatial: Multivariate multiscale spatial analysis

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling 196:483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>

Drost H-G (2018) Philentropy: Information Theory and Distance Quantification with R. Journal of Open Source Software 3:765. <https://doi.org/10.21105/joss.00765>

Dufrêne M, Legendre P (1997) Species Assemblages And Indicator Species:the Need For A Flexible Asymmetrical Approach. Ecological Monographs 67:345–366. [https://doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2](https://doi.org/10.1890/0012-9615(1997)067%5b0345:SAAIST%5d2.0.CO;2)

Dunn JC (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3:32–57. <https://doi.org/10.1080/01969727308546046>

Eckert A (2022) parallelDist: Parallel distance matrix computation using multiple threads

EEA (2019) [EU-Hydro – River Network Database](https://doi.org/10.2909/393359a7-7ebd-4a52-80ac-1a18d5f3db9c)

EEA (2020) Potential flood-prone area extent

Evans D (2005) Some Uses of the Map of the Natural Vegetation of Europe for Natura 2000. Application and Analysis of the Map of the Natural Vegetation of Europe 195–204

Gaston KJ, Blackburn TM (2000) Pattern and process in macroecology. Blackwell Science, Oxford, England; Malden, Massachusets, USA

Gelman A, Meng X-L, Stern H (1996) [Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies](https://www.jstor.org/stable/24306036). Statistica Sinica 6:733–760

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical science 7:457–472

Gotelli NJ (2000) NULL MODEL ANALYSIS OF SPECIES CO-OCCURRENCE PATTERNS. Ecology 81:2606–2621. [https://doi.org/10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081%5b2606:NMAOSC%5d2.0.CO;2)

Gotelli NJ, Hart EM, Ellison AM (2015) EcoSimR: Null model analysis for ecological data. R package version 01 0 10:

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36. <https://doi.org/10.1148/radiology.143.1.7063747>

Hoegh A, Roberts DW (2020) Evaluating and presenting uncertainty in model-based unconstrained ordination. Ecology and Evolution 10:59–69. <https://doi.org/10.1002/ece3.5752>

Hui C, McGeoch MA (2014) Zeta Diversity as a Concept and Metric That Unifies Incidence-Based Biodiversity Patterns. The American Naturalist 184:684–694. <https://doi.org/10.1086/678125>

Illies J (1978) Limnofauna europaea. Fischer Stuttgart

Jupke JF, Birk S, Apostolou A, et al (2023) European river typologies fail to capture diatom, fish, and macrophyte community composition. Science of The Total Environment 896:165081. <https://doi.org/10.1016/j.scitotenv.2023.165081>

Kauffman L, Rousseeuw P (1990) Finding Groups in data: An Introduction to Cluster Analysis. John Wiley & Sons Inc, Hoboken, New Jerse

Khanal GP, Hartmann A, Leitner P, et al (2025) Benthic macroinvertebrate assemblages foster the ecoregional approach in near-natural streams of the Bhutan Himalaya. Hydrobiologia. <https://doi.org/10.1007/s10750-025-05995-y>

Kopp D, Stoddard JL, Hill R, et al (2023) Joint species distribution models reveal taxon-specific sensitivities to potential anthropogenic alteration. Freshwater Science 726283. <https://doi.org/10.1086/726283>

Kuhn M, Wickham H (2020) Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles.

Kurtzer GM, Sochat V, Bauer MW (2017) Singularity: Scientific containers for mobility of compute. PLOS ONE 12:e0177459. <https://doi.org/10.1371/journal.pone.0177459>

Latombe G, McGeoch MA, Nipperess DA, Hui C (2018) Zetadiv : An R package for computing compositional change across multiple sites, assemblages or cases. Ecology

Linke S, Lehner B, Ouellet Dallaire C, et al (2019) Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data 6:283. <https://doi.org/10.1038/s41597-019-0300-6>

Liu FT, Ting KM, Zhou Z-H (2008) Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining. IEEE, pp 413–422

Lyche Solheim A, Globevnik L, Austnes K, et al (2019) A new broad typology for rivers and lakes in Europe: Development and application for large-scale environmental assessments. Science of The Total Environment 697:134043. <https://doi.org/10.1016/j.scitotenv.2019.134043>

Maier M, Lakens D (2022) Justify Your Alpha: A Primer on Two Practical Approaches. Advances in Methods and Practices in Psychological Science 5:1–14

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer research 27:209–220

Messager ML, Lehner B, Grill G, et al (2016) Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nature Communications 7:13603. <https://doi.org/10.1038/ncomms13603>

Metzger MJ, Bunce RGH, Jongman RHG, et al (2005) A climatic stratification of the environment of Europe: A climatic stratification of the European environment. Global Ecology and Biogeography 14:549–563. <https://doi.org/10.1111/j.1466-822X.2005.00190.x>

Microsoft, Weston S (2022) Foreach: Provides foreach looping construct

Niku J, Hui FKC, Taskinen S, Warton DI (2019) Gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in r. Methods in Ecology and Evolution 10:2173–2182. <https://doi.org/10.1111/2041-210X.13303>

Oksanen J, Simpson GL, Blanchet FG, et al (2022) Vegan: Community ecology package

Ouellet Dallaire C, Lehner B, Sayre R, Thieme M (2019) A multidisciplinary framework to derive global river reach classifications at high spatial resolution. Environmental Research Letters 14:024003. <https://doi.org/10.1088/1748-9326/aad8e9>

Ovaskainen O, Abrego N (2020) Joint Species Distribution Modelling. Cambridge University Press, Cambridge, United Kingdom

Panagos P, Hengl T, Wheeler I, et al (2023) Global rainfall erosivity database (GloREDa) and monthly R-factor data at 1 km spatial resolution. Data in Brief 50:109482. <https://doi.org/10.1016/j.dib.2023.109482>

Panagos P, Van Liedekerke M, Borrelli P, et al (2022) European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. European Journal of Soil Science 73:e13315. <https://doi.org/10.1111/ejss.13315>

Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10:439–446. <https://doi.org/10.32614/RJ-2018-009>

Pharaoh E, Diamond M, Jarvie HP, et al (2024) Potential drivers of changing ecological conditions in English and Welsh rivers since 1990. Science of The Total Environment 946:174369. <https://doi.org/10.1016/j.scitotenv.2024.174369>

Pinkert S, Sica YV, Winner K, Jetz W (2023) The potential of ecoregional range maps for boosting taxonomic coverage in ecology and conservation. Ecography e06794. <https://doi.org/10.1111/ecog.06794>

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence diagnosis and output analysis for MCMC. R News 6:7–11

Popovic G, Mason TJ, Drobniak SM, et al (2024) Four principles for improved statistical ecology. Methods in Ecology and Evolution 2041–210X.14270. <https://doi.org/10.1111/2041-210X.14270>

Rahman AU, Tikhonov G, Oksanen J, et al (2024) Accelerating joint species distribution modelling with Hmsc-HPC by GPU porting. PLOS Computational Biology 20:e1011914. <https://doi.org/10.1371/journal.pcbi.1011914>

Roberts DW (2023) Labdsv: Ordination and multivariate analysis for ecology

Roberts DW (2007) R package labdsv

Simonsohn U, Gruson H (2025) Groundhog: Version-control for CRAN, GitHub, and GitLab packages

Stone L, Roberts A (1990) The checkerboard score and species distributions. Oecologia 85:74–79. <https://doi.org/10.1007/BF00317345>

Szoszkiewicz K, Jusik S, Pietruczuk K, Gebler D (2019) The Macrophyte Index for Rivers (MIR) as an Advantageous Approach to Running Water Assessment in Local Geographical Conditions. Water 12:108. <https://doi.org/10.3390/w12010108>

Tikhonov G, Ovaskainen O, Oksanen J, et al (2022) Hmsc: Hierarchical model of species communities

Tjur T (2009) Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination. The American Statistician 63:366–372. <https://doi.org/10.1198/tast.2009.08210>

Tóth B, Weynants M, Nemes A, et al (2015) New generation of hydraulic pedotransfer functions for Europe. European Journal of Soil Science 66:226–238. <https://doi.org/10.1111/ejss.12192>

Vallé C, Poggiato G, Thuiller W, et al (2023) Species associations in joint species distribution models: From missing variables to conditional predictions. Journal of Biogeography jbi.14752. <https://doi.org/10.1111/jbi.14752>

Van Sickle J (1997) Using Mean Similarity Dendrograms to Evaluate Classifications. Journal of Agricultural, Biological, and Environmental Statistics 2:370–388. <https://doi.org/10.2307/1400509>

Wang Y, Naumann U, Eddelbuettel D, et al (2022) [Mvabund: Statistical methods for analysing multivariate abundance data](https://doi.org/10.32614/CRAN.package.mvabund)

Wang Y, Naumann U, Wright ST, Warton DI (2012) Mvabund - an R package for model-based analysis of multivariate abundance data: *The* mvabund R package. Methods in Ecology and Evolution 3:471–474. <https://doi.org/10.1111/j.2041-210X.2012.00190.x>

Wasson J-G, Chandesris A, Garcia-Bautista A, et al (2007) European Hydro-Ecoregions

Westfall P, Young S (1993) Resampling-based multiple testing: Examples and methods for p-Value adjustment. Wiley

Wulff JN, Taylor L (2023) How and why alpha should depend on sample size: A Bayesian-frequentist compromise for significance testing. PsyArXiv

Yang J, Cao M, Swenson NG (2018) Why Functional Traits Do Not Predict Tree Demographic Rates. Trends in Ecology & Evolution 33:326–336. <https://doi.org/10.1016/j.tree.2018.03.003>