test

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 usefull for several discinct 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 anthropogenic 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 relationsships (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 priorisation 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 flaws. First, there is contention which method should be used to determine biovalidity. Different studies use different methods, which reduces comparability among them. It remians 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, two practices have taken hold. Biovalidity is judged against unrealistic heurisitcs without ecological foundations or against statistical significane. Effect size heuristics exist for some commonly used metrics but not all. They are not specifically tailored to evaluate biovalidity. They are based on practically irrelevant considerations. 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.

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 statistcally significant. A move toward *a priori* determined relevant effect sizes seems frutiful (Popovic et al. 2024). Other ways to adress these issues have been proposed but will be considered here (Maier and Lakens 2022; Wulff and Taylor 2023; Crouch 2025)

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 biological data. These simulations 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.

1.1 Methods

1.1.1 The data sets

We have compiled occurrence data sets of diatoms, aquatic macroinvertebrates, fish, and aquatic macropyhtes (Tables S1-S4).  
These data sets range from few sites () to many, from moments in time() to long-term time series, and from local scales () to multiple countries ().

We will train HMSC models with each data set to predict communities for similar but artificial environments. Environments which are simulated to follow typology systems with known qualities.  
Evaluating the congruence between simulated biotic communities and artificial typology systems over many data sets will provide us with ecologically plausible benchmarks for the selected evaluation metrics.

Each data set will be subsampled. Within each data set, we will stratify the subsampling by year. Within each year, we will restrict the data to the three consecutive months with the most samples. If the number of samples in these months () < 100, this year will not be evaluated for this data set. If samples, our subsampling methodology will incorporate a descending-frequency approach based on one hundred-sample increments. The process will begin by dividing by 100 (rounded down) to determine the number of increments . For each increment , we will generate random subsamples, each containing samples. When the remainder from dividing by 100 exceeds 0.5, we will include one additional sample at the maximum size. This approach ensures thorough sampling coverage while maintaining computational feasibility. To maintain computational feasibility, the maximum number of total subsets per data set is six. For each subsample, we will determine the spatial scale (minimum, median, and maximum spatial distance between sampling locations) with the sf R package (ref) and the taxonomic resolution as fraction of taxa identified to species, genus, family, or higher level.  
Lastly, we will remove taxa that occur in less than 5% of sampling sites for this subset.

1.1.2 HMSC Model

To each subset we will fit a Hierarchical model of species communities (HMSC, Ovaskainen and Abrego 2020) to establish the relationship between taxon occurrence 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. The model will be fit with the Hmsc R package [3.0-13] (Tikhonov et al. 2022) and a probit residual distirbution. For each subset an initial model will be fit with two chains, 2000 samples, a thinning of two, and a transient period of 2000 samples.  
Model fit will be evaluated with the potential scale reduction factor [PSRF; Gelman and Rubin (1992)] as implemented in the coda package [0.19-4.1] (Plummer et al. 2006). The PSRF compares the variance between multiple chains to the variance within each chain to assess chain convergence. If 10% of PSRFs are > 1.1 the model is refit. We will iteratively increase the chain lengths by increments of 1000 until <10% of PSRFs are > 1.1. If this is not achieved with a sample size of 10000, the subset will be discarded from further analyses. Thinning is always the number of chains divided by 1000, and the length of the transient equals the number of samples.

1.1.3 Environmental data

We will compile and extensive data base of environmental variables relevant for freshwater organisms (Table 1). Each variable will be summarized at the catchment scale. Catchments will be derived from the EU Hydro DEM catchment data base (EEA 2019).

Table 1.1: 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) |
| Valley Bottom Flatness Index | EcoDataCube | Witjes et al. (2023) |
| 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 will be captured Moran’s Eigenvector Maps [MEM; Dray et al. (2006)]. MEM transform Euclidean spatial distance matrices into eigenfunctions. MEM gives rise to a large number of spatial eigenfunctions (typically number of samples - 1). Different methods to select from them have been proposed an lead to significantly different outcomes (Bini et al. 2009). We will use all spatial eigenfunctions that correlate statistically significant (at ) with the residuals of logistic 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. MEM will be implemented with the adespatial R package [0.3-2.4] (Dray et al. 2024).

1.1.4 Modellings species communities

We will interpret all the variance explained by environmental variables as abiotic filtering, the variance explained by spatial eigenvectors as spatial filtering, and the variance of the latent variables as biotic interactions. The last point is contentions (Dormann et al. 2018; e.g., Blanchet et al. 2020; Vallé et al. 2023), as the latent variables can also capture unmeasured environmental drivers. In our case this will likely include local scale variables such as sediment composition. We expect therefore to slightly overestimate the relevance of biotic interactions and will keep this in mind in any interpretations of the results. Tjur’s (Tjur 2009) will be used to evaluate the overall explained variation in the biotic data.  
The unexplained variance , will be interpreted as stochasticity in community composition. Each of these values will be averaged across species. For each fitted model, we thus obtain a vector of four numbers, which sum to 1, estimating the assembly mechanisms of this meta-community. Over all the models we will run, we will be able to derive a multivariate probability distribution of assembly mechanisms. As each mechanism itself can range between 0 and 1, their distribution can likely be captured by a beta distribution. The Dirichlet distribution is a multivariate distribution with beta marginals, and thus able to represent this.

1.1.5 Simulating typologies

The fitted models will be used to simulate new biotic data at the same sampling locations. While spatial coordinates will remain the same, we will slightly adjust the environmental variables. Before, we alter the variables we will cluster the sites into environmental types, i.e., we will create artificial typologies.  
These typologies will not be based on all the available environmental variables but rather on a random subset. For each subsample we will draw nine sets of environmental variables, randomizing both the number of variables (between three and all) and their identity. These selected environmental variables will then be submitted to a k-means clustering algorithm. We will test between two and ten clusters and pick the solution with the highest average silhouette width [ASW; Kauffman and Rousseeuw (1990)] which has at least ten samples in the smallest cluster. If all classifications have a group with 10 samples, we will pick the solution with the most classes for which only one class has 10 samples and balance the classes by moving the samples from other classes that are closest to the small classes centroid to that class until it has 10 members. The quality of the final k-means classification is assessed with the ASW. Additionally, we compute a fuzzy C means classification [FCM; Dunn (1973)] of the same data. We will use the same number of centers as in the k-means classification and the implementation available in the vegclust [] R package (De Cáceres et al. 2013). The quality of this fuzzy classification will be quantified with the normalized partitioning entropy (Bezdek 2013).

After creating this artificial typology system, we will alter the environmental variables. These alterations will alter the quality of the typology system by dilating (i) the distance between centroids and (ii) the distances between samples and their centroid. An good typology has high compactness (sites close to centroid) and separation (centroids far apart). By randomly varying these two properties thirty times for each artificial typology, we create a total of 270 landscapes for each subset. To modify separation, we first compute the location of overall centroid in environmental space. We then randomly draw a scaling factor between 0 and 2 from a uniform distribution. The new centroids are then computed as . To adjust the compactness, we draw a second scaling factor between -1 and 1. We multiply the distance between original site location and original centroid with and add the result to the new centroid locations to obtain the new site location.

1.1.6 Evaluating typologies - the theory

We will evaluate each typology with a selection of recommended or practically used methods. We provide an extensive and harmonized explanation of all metrics in the supplementary materials (SX).

1.1.7 Evaluating typologies - the theory

We will compute Jaccard distance matrices of predicted biotic communities using the parallelDist package (Eckert 2022). Classification strength, ANOSIM, PERMANOVA will be computed with functions from the vegan package (Oksanen et al. 2022). will be computed with purpose written code and using the zetadiv package (Latombe et al. 2018). Indicator values will be computed with indicspecies (De Cáceres and Legendre 2019) and ISAMIC with labdsv (Roberts 2007). All code is available ADD GITHUB REPO

1.1.8 Deriving benchmarks

To derive benchmarks from these evaluations we will use Random Forest regression analyses using the tidymodels framework (reference). The analysis will be performed separately for each evaluation metric and taxon. In total, we will fit 80 random forest models (four taxa and 20 evaluation metrics). The Random Forest model will be implemented with Ranger (reference) 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.   
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.

1.1.9 Software

To ensure replicability, we will use the groundhog [3.2.2] (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 (Wang et al. 2022), Checks of posterior distributions were conducted with coda … (), Data wrangling will be done with data.table [1.16.2] (Barrett et al. 2024), dplyr [] [@], lubridate [] [@],

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 p-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.

# 2 Acknowledgements

# References

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>

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>

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>

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

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

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>

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

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>

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>

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>

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

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>

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>

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

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>

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>