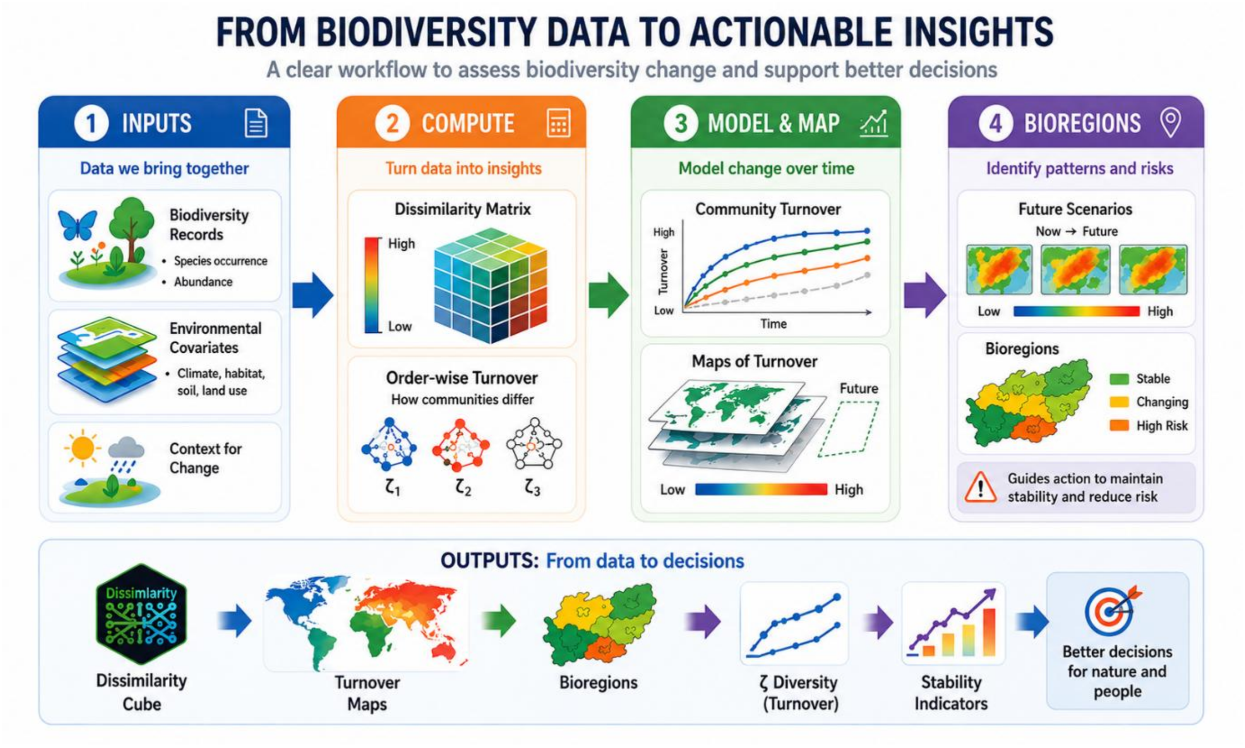
`dissmapr`

A Workflow for Compositional Dissimilarity and Biodiversity Turnover Analysis

MacFadyen, S., Yahaya, M., Kumschick, S., Hui, C.

EU Horizon Europe | Grant Agreement No. 101059592



What is `dissmapr`?

dissmapr is an open-source R package that provides a complete, reproducible workflow for analysing how biological communities change across space and time. Rather than focusing on the predicted occurrence of individual species, it quantifies variation in assemblage composition, providing a community-level perspective on biodiversity change. The package is designed to work with standardised species-occurrence records from biodiversity data infrastructures such as GBIF, and can be applied to any taxon, region, or time period where sufficient occurrence data are available.

The framework combines species-occurrence records with environmental covariates to compute multisite compositional turnover using order-wise metrics such as zeta diversity. These dissimilarity measures are then linked to environmental and geographic drivers using predictive models, allowing users to generate turnover surfaces, identify bioregions, and assess how community structure may reorganise under future conditions. By operating at the community level rather than the single-species level, dissmapr captures emergent patterns of ecological organisation that would otherwise remain hidden in species-by-species analyses.

Implemented as a modular R framework, dissmapr standardises data acquisition and linking, computes dissimilarity (including zeta-based and conventional indices), models drivers using Multi-Site Generalised Dissimilarity Modelling (MS-GDM with i-splines), and maps outputs as continuous turnover surfaces and bioregions. The same workflow propagates to scenarios (e.g. future climates) to forecast shifting community boundaries. Each step is fully scripted and parameterised, ensuring transparency and reproducibility from raw data to final mapped outputs.

By turning biodiversity records into mappable signals of community change, the Dissimilarity Cube helps identify stable regions, shifting assemblages, and areas at risk of ecological reorganisation. It therefore provides a scalable and policy-relevant approach for monitoring biodiversity structure, detecting emerging novel communities, and supporting spatial conservation planning. The outputs are designed to complement other B-Cubed products (Suitability Cubes, Invasibility Cubes) within an integrated framework for evidence-based biodiversity management.

Why Community-Level Analysis?

Ecological dissimilarity arises through species replacement and nestedness across environmental, spatial, and temporal gradients, reflecting processes such as environmental filtering, dispersal limitation, biotic interactions, and historical contingency. Understanding how entire communities, rather than individual species, respond to these gradients is essential for capturing the full dimensionality of biodiversity change. Single-species models, while powerful for targeted assessments, cannot reveal emergent properties of assemblages such as the coherence of species pools, the integrity of functional groups, or the degree to which novel community combinations are forming.

Traditional measures of beta diversity quantify differences between two assemblages but cannot capture the full structure of turnover among multiple sites. dissmapr extends this principle by formalising multisite dissimilarities within a multidimensional data framework, enabling consistent quantification, aggregation, and modelling of compositional change across spatial and temporal scales. This multisite perspective is critical because ecological processes operate simultaneously across many locations; pairwise comparisons provide only a partial view of the landscape-level reorganisation of communities.

For conservation planning and policy, community-level analysis offers several distinct advantages. It allows practitioners to identify areas where assemblage integrity is high and likely to persist, as well as transition zones where communities are actively reorganising. These insights are directly relevant for delineating conservation priorities, designing ecological networks, and monitoring the effectiveness of management interventions. Policy frameworks such as the Kunming-Montreal Global Biodiversity Framework increasingly require indicators that reflect the state and trajectory of ecosystems as a whole, not merely the status of individual species. Community-level turnover metrics, as generated by dissmapr, provide precisely this kind of integrated, scalable evidence base.

Furthermore, by linking compositional turnover to environmental drivers through predictive modelling, dissmapr enables scenario-based assessments that inform proactive rather than reactive management. Decision-makers can evaluate how proposed land-use changes, climate trajectories, or conservation strategies are likely to affect the spatial structure of biodiversity, supporting more robust and forward-looking policy decisions.

Zeta Diversity: Beyond Pairwise Comparisons

Zeta diversity generalises beta diversity by quantifying the number of species shared across i sites, extending analysis beyond pairwise comparisons. The framework defines zeta_i as the expected number of species common to exactly i sites drawn from a regional pool. Lower zeta-orders (e.g. zeta_2, zeta_3) represent turnover among rare or localised species, whereas higher orders capture turnover in widespread, common species. This order-wise decomposition provides a much richer characterisation of community structure than any single dissimilarity index, because it separates the contributions of rare and common species to overall compositional change.

The Venn diagram below illustrates this concept. Each circle represents a site containing a set of species, and the overlapping regions show species shared among increasing numbers of sites. At zeta-order 1 (a single site), zeta equals the local species richness. At zeta-order 2, it counts species shared between any pair of sites. As the order increases, progressively fewer species are held in common, and the resulting decline curve characterises the rate and structure of compositional turnover across the landscape.

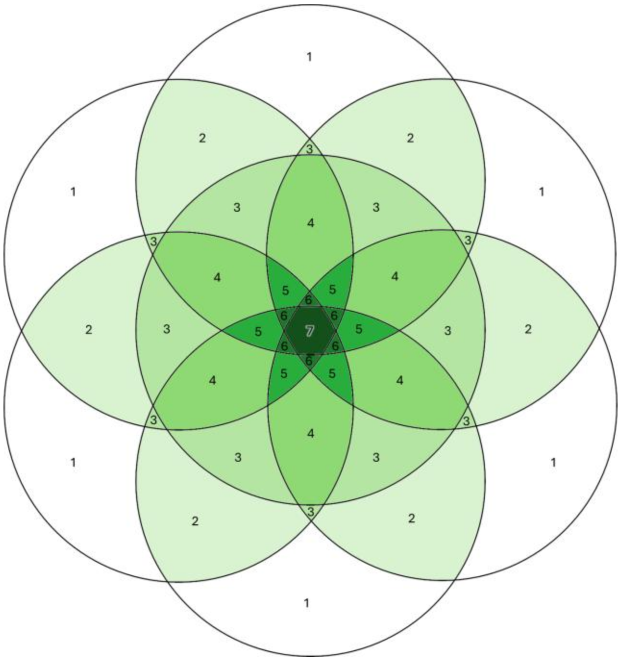


Figure 1: Conceptual Venn diagram of zeta diversity, showing species shared across increasing numbers of sites (zeta-orders).

Order-wise dissimilarity defines the scale of inference. At low zeta-orders, turnover is often governed by geographic distance and local environmental heterogeneity, while higher orders increasingly reflect broader climatic or sampling effects. The rate of zeta-decline with increasing order quantifies the overall rate of compositional turnover across the landscape, linking community similarity to underlying processes of assembly, dispersal, and environmental filtering. Steep declines indicate high turnover (few species in common among sites), while shallow declines suggest more homogeneous assemblages.

From a practical standpoint, the zeta framework enables analysts to ask targeted questions about different components of biodiversity. For example, conservation strategies aimed at protecting rare, range-restricted species would focus on patterns revealed at low zeta-orders, while assessments of broad-scale community coherence would draw on higher-order patterns. By computing dissimilarity at multiple orders and comparing the resulting surfaces, dissmapr provides a layered view of biodiversity structure that supports nuanced, context-specific decision-making.

The `dissmapr` Workflow

The Dissimilarity Cube is constructed by harmonising species occurrences, environmental predictors, and (optionally) traits into standard matrices, then transforming these inputs into order-wise dissimilarity and modelled turnover surfaces. The workflow is modular and fully scripted, so each stage is transparent and repeatable. Users can adapt individual components to their study system while maintaining methodological consistency across analyses.

Phase 1: Inputs and Setup

The first phase establishes the analytical framework by loading required libraries, defining the study region and spatial resolution, acquiring species-occurrence data (typically from GBIF via the rgbif interface), and assembling environmental predictor layers (e.g. climate variables, topography, land cover). Data are harmonised into a consistent spatial grid and formatted as site-by-species and site-by-environment matrices. Quality filters are applied to remove records with insufficient spatial precision or taxonomic reliability, and environmental layers are checked for collinearity and coverage.

Phase 2: From Data to Dissimilarity

In the second phase, the harmonised matrices are used to compute order-wise dissimilarity. Site-by-species matrices are screened for minimum species richness and site occupancy thresholds, then zeta diversity is calculated across a range of orders using the zetadiv package. The resulting dissimilarity values are linked to environmental and geographic predictors through Multi-Site Generalised Dissimilarity Modelling (MS-GDM), which uses monotonic i-spline regressions to capture nonlinear relationships between predictors and compositional turnover. This modelling step identifies which environmental gradients drive the greatest community change and at what thresholds transitions occur.

Phase 3: Prediction, Mapping, and Scenarios

The final phase applies the fitted MS-GDM models to predict continuous dissimilarity surfaces across the study landscape. These surfaces represent the expected degree of compositional turnover at each location relative to the broader community context. Bioregionalisation algorithms (hierarchical clustering of the predicted turnover surfaces) partition the landscape into spatially coherent bioregions that reflect underlying community structure. The same models can be propagated to alternative environmental scenarios (e.g. projected future climates) to forecast how bioregional boundaries and community composition may shift, enabling sensitivity analysis and change detection.

Throughout all three phases, the workflow integrates biodiversity records, environmental rasters, and optional trait or effort data into harmonised site-by-environment matrices. It computes compositional turnover metrics across multiple zeta-orders, fits MS-GDM models using i-spline regressions, and predicts continuous dissimilarity surfaces across landscapes. Each step logs its parameters and outputs, supporting full traceability from raw input data to final mapped products.

Zeta-MSGDM with `dissmapr`

The zeta-based Multi-Site Generalised Dissimilarity Model (MS-GDM) is the core analytical engine of dissmapr. It extends classical Generalised Dissimilarity Modelling by replacing pairwise beta diversity with order-wise zeta dissimilarity, enabling simultaneous analysis of compositional turnover across multiple sites. The model fits monotonic i-spline basis functions to each environmental predictor, producing partial-dependence curves that quantify how compositional turnover responds to environmental gradients. These curves reveal which variables drive community change and at what points along each gradient the rate of turnover is greatest. The resulting fitted model provides the foundation for spatial prediction and scenario analysis in subsequent workflow steps. By combining zeta diversity with i-spline regression, dissmapr captures nonlinear and threshold responses of community composition to environmental change. This approach accommodates the complex, often non-monotonic relationships between biodiversity turnover and climate, topography, or land-use gradients, producing ecologically interpretable outputs that inform both research and conservation planning.

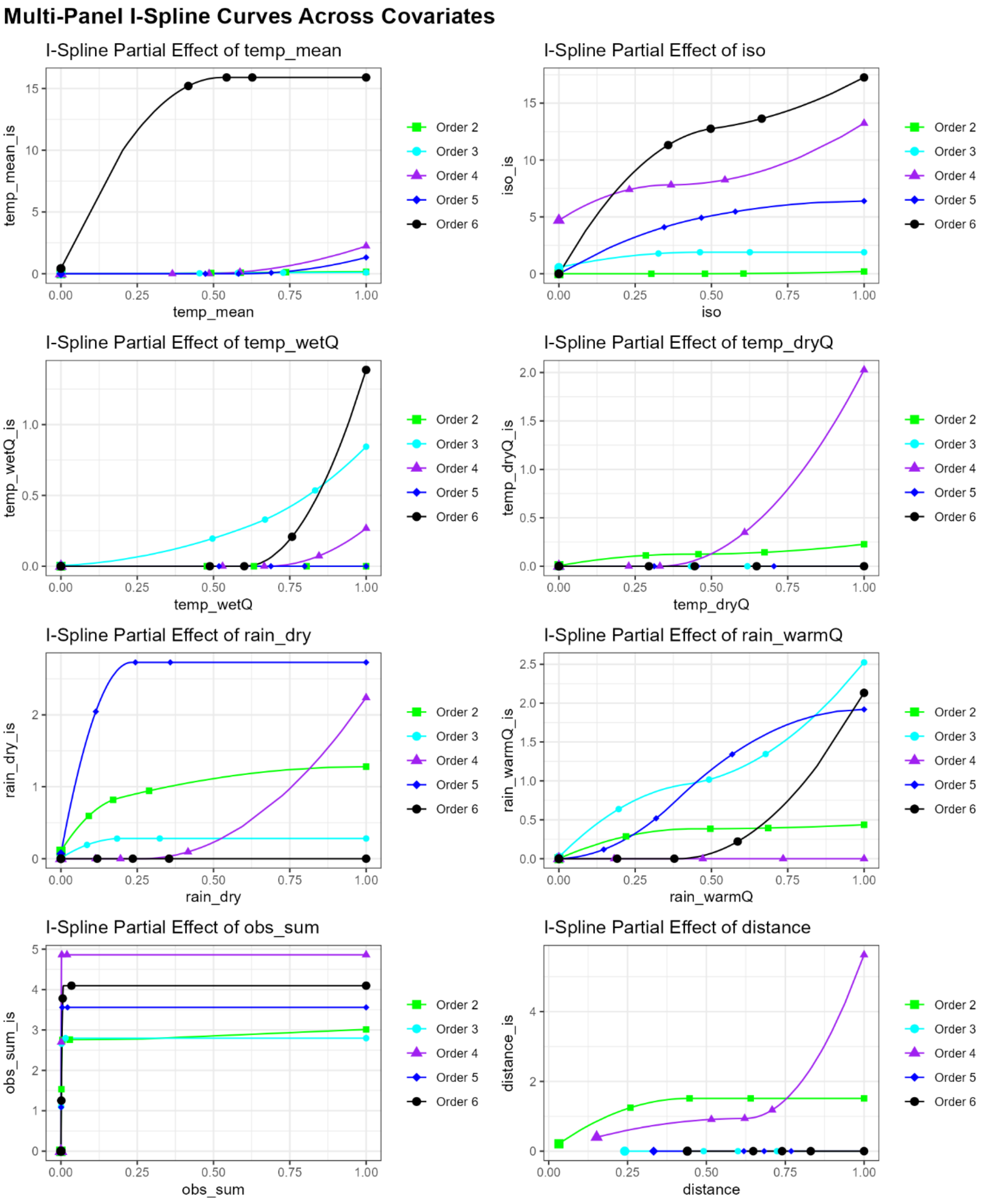


Figure 2: Multi-Site Generalised Dissimilarity Model (MS-GDM) fitting process, showing i-spline partial-dependence curves that quantify how compositional turnover responds to individual environmental gradients.

The figure above shows an example of the i-spline partial-dependence output produced by the run_ispline_models() function. Each panel corresponds to a different environmental predictor, and the curves depict the fitted i-spline response of compositional turnover to that variable. Steeper slopes indicate environmental ranges where small changes drive large shifts in community composition, while flat regions indicate relative compositional stability. Together, these panels provide a diagnostic summary of how environmental gradients partition into zones of rapid versus slow turnover.

Predict Community Turnover

Once the MS-GDM model is fitted, dissmapr generates continuous spatial predictions of compositional turnover across the study landscape. The fitted i-spline transformations are applied to gridded environmental rasters, producing turnover surfaces that represent predicted compositional dissimilarity relative to a reference configuration. These surfaces can be visualised as gradient maps, where colour variation reflects the magnitude and direction of community change across space.

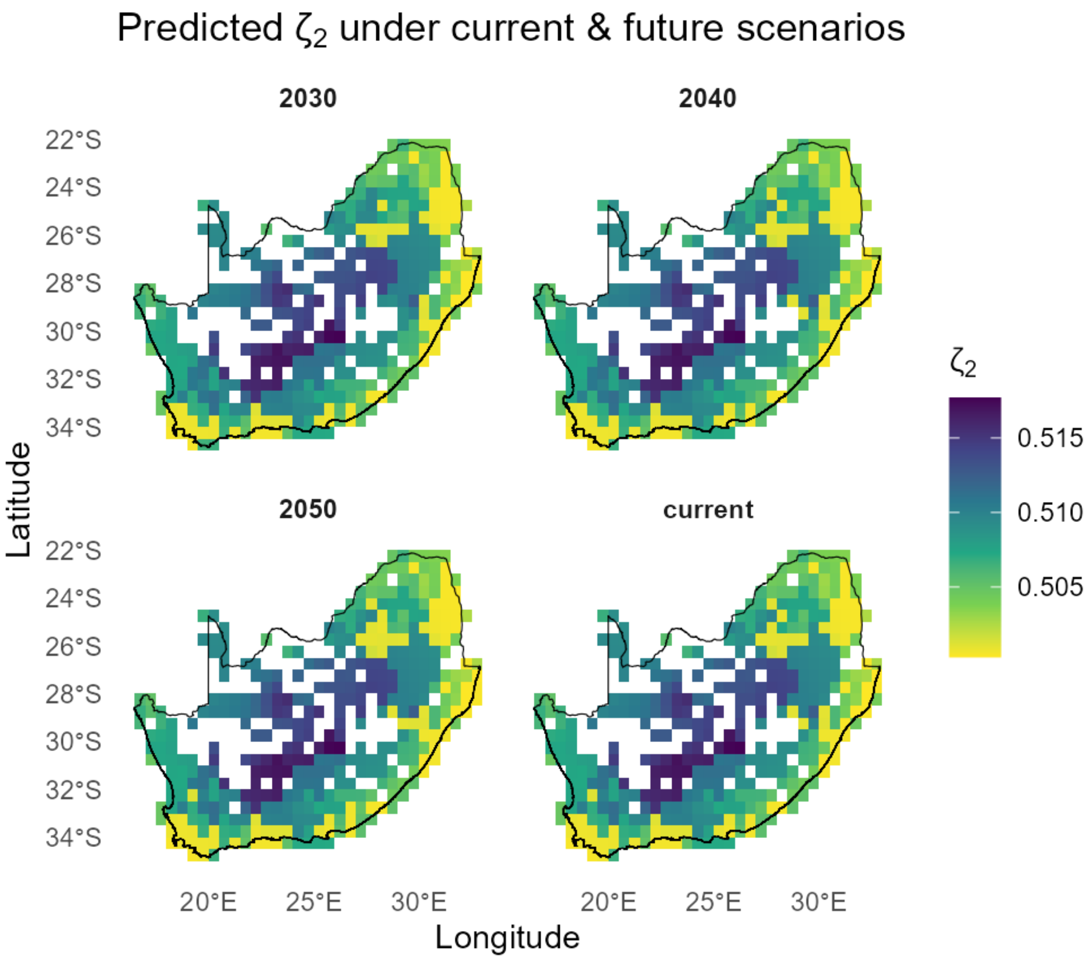


Figure 3: Predicted compositional turnover surfaces showing spatial patterns of community change across the study landscape, derived from MS-GDM model outputs applied to environmental rasters.

Bioregionalisation and Scenario Analysis

Bioregionalisation clusters continuous turnover surfaces into spatially coherent bioregions that summarise the structure of ecological communities. These emergent regions are not defined by administrative or political boundaries, but instead reflect the underlying patterns of species co-occurrence and environmental similarity. They provide an objective, data-driven basis for conservation planning and monitoring, enabling practitioners to identify ecologically meaningful units for management and to track changes in community structure over time. The bioregionalisation step in dissmapr applies hierarchical or k-medoids clustering to the predicted turnover surfaces, with the number of regions determined by ecological interpretability and statistical criteria. The resulting bioregions capture gradients in community composition and can be compared across taxa, time periods, or spatial scales. By overlaying bioregional maps with existing protected-area networks or land-use plans, conservation planners can evaluate representativeness and identify gaps in current protection.

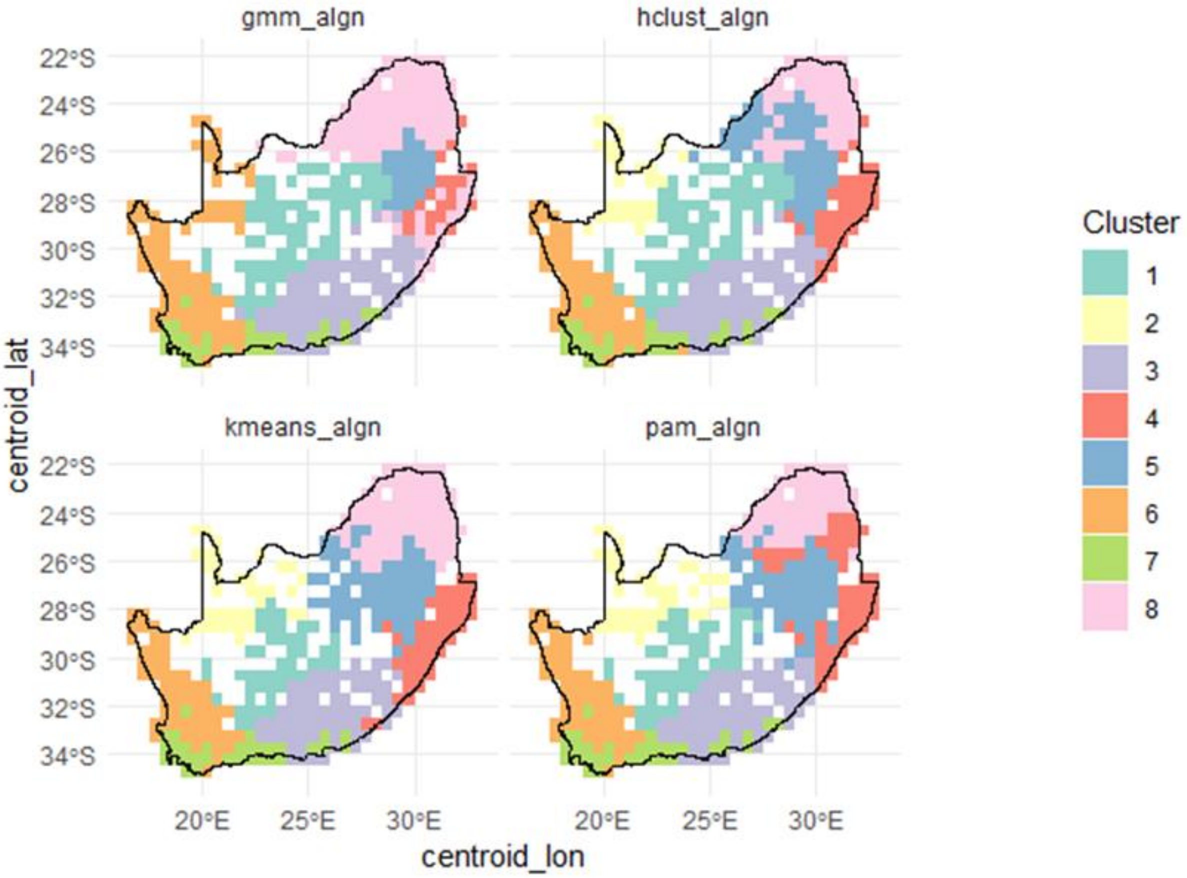


Figure 4: Bioregional partitions of South Africa derived from butterfly assemblage turnover analysis, illustrating data-driven delineation of ecologically coherent regions

Scenario propagation applies the same dissimilarity models to alternative climate or land-use projections, enabling the mapping of potential shifts in community composition and bioregional boundaries. This facilitates prediction of biodiversity responses under future environmental scenarios while maintaining methodological consistency with the baseline analysis. Users can compare current and projected bioregions to identify areas where community reorganisation is most likely, supporting proactive allocation of conservation resources. The map_bioreg() function partitions the predicted turnover surface into discrete bioregions using hierarchical or k-medoids clustering. The figure below shows four alternative bioregional classifications of South Africa derived from butterfly assemblage turnover at zeta-order 2, each generated with a different clustering algorithm. Comparing across methods highlights areas of consistent classification (core bioregional identities) versus areas of ambiguity (transition zones sensitive to methodological choice), providing a robust basis for delineating conservation units.

The map_bioregDiff() function quantifies and maps changes between baseline and scenario-projected bioregional boundaries. By comparing the current bioregional classification against one projected under a future climate scenario, the function identifies grid cells where bioregional identity is predicted to shift. The resulting change map highlights areas of expected community reorganisation, distinguishing stable regions from those where ecological transitions are anticipated. These outputs directly support forward-looking conservation planning by flagging priority areas for monitoring and adaptive management.

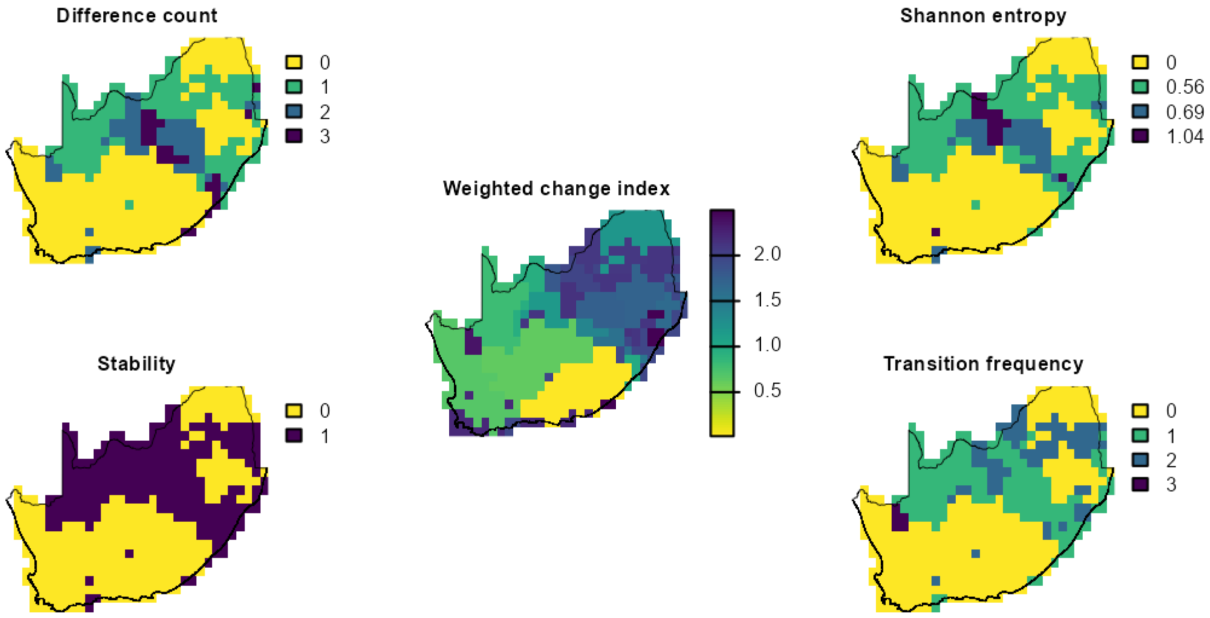


Figure 5: Areas of expected community reorganisation under alternative environmental futures.

Key Outputs

The dissmapr workflow produces a suite of interrelated outputs that collectively characterise the compositional structure of biodiversity across a study region. These outputs are designed to be interpretable both as standalone analytical products and as inputs to broader policy and planning frameworks.

Order-wise dissimilarity matrices: Quantitative matrices recording compositional dissimilarity among sites at each zeta-order, providing the raw material for all subsequent modelling and mapping steps.

Continuous turnover surfaces: Spatially explicit raster layers depicting predicted compositional turnover across the landscape, generated by applying fitted MS-GDM models to gridded environmental data.

Bioregional partitions: Classified maps delineating spatially coherent bioregions based on clustering of turnover surfaces, providing ecologically meaningful spatial units for planning and monitoring.

Change maps: Spatial layers comparing baseline and scenario-projected turnover or bioregional boundaries, highlighting areas of expected community reorganisation under alternative environmental futures.

Partial-dependence plots: Diagnostic visualisations showing the fitted i-spline response of compositional turnover to each environmental predictor, revealing the magnitude and shape of each driver's contribution.

The Dissimilarity Cube is structured along three orthogonal dimensions: Sites (the spatial units over which assemblages are defined), Predictors (environmental and geographic variables driving turnover), and Dissimilarity (zeta-order matrices or continuous surfaces). This cube structure enables flexible querying, aggregation, and comparison across taxa, regions, and time periods, and aligns with the broader B-Cubed data architecture for interoperable biodiversity information products.

Applications

For researchers, dissmapr provides a standardised, reproducible platform for investigating the drivers and spatial patterns of community turnover across taxa and regions. The modular workflow supports comparative analyses (e.g. contrasting turnover patterns among insects, plants, and vertebrates within the same landscape), temporal studies (tracking how assemblage composition changes over decades), and methodological experiments (evaluating the sensitivity of results to spatial resolution, zeta-order, or predictor selection). By automating data handling and model fitting, the package reduces analytical overhead and enables researchers to focus on ecological interpretation.

For conservation planners, the outputs of dissmapr offer a data-driven foundation for spatial prioritisation and network design. Bioregional maps identify ecologically coherent management units, turnover surfaces highlight transition zones where conservation investment may be most effective, and change maps pinpoint areas where community reorganisation is anticipated under future conditions. These products can inform systematic conservation planning exercises, gap analyses of existing protected-area networks, and the design of ecological corridors connecting areas of high compositional distinctiveness.

For policy, dissmapr contributes to the evidence base required by international biodiversity frameworks, including the Kunming-Montreal Global Biodiversity Framework and EU Biodiversity Strategy for 2030. Indicators derived from compositional turnover analysis (such as the rate of community change, the stability of bioregional boundaries, or the emergence of novel assemblages) can supplement species-level indicators with community-level information on ecosystem integrity and resilience. The integration of dissmapr within the B-Cubed framework, alongside Suitability Cubes and Invasibility Cubes, provides a comprehensive toolkit for assessing biodiversity status, trends, and threats at scales relevant to national and supranational reporting.

The modular workflow enables reproducible analyses of biodiversity structure and change across taxa, regions, and time periods. Whether applied to a single national dataset or scaled to continental analyses, dissmapr provides consistent, transparent outputs that support evidence-based decision-making for biodiversity conservation and management.

Package Functions

The dissmapr package consists of 10 core functions addressing key steps in biodiversity data analysis, from sourcing raw occurrence and environmental data to mapping bioregions and predicting compositional turnover. It also includes three specialised functions that automate the zeta-MSGDM workflow, and a suite of available metrics for order-wise computation. The package leverages existing tools such as the zetadiv package (specifically Zeta.msgdm) for MS-GDM, alongside widely used R libraries including terra, sf, and ggplot2.

Core Functions

`get_occurrence_data()` Import and harmonise biodiversity-occurrence data

`generate_grid()` Generate spatial grid and gridded summaries

`assign_mapsheet()` Add nearest mapsheet code and centre coordinates

`get_enviro_data()` Retrieve, crop, resample, and link environmental rasters to sampling sites

`format_df()` Format biodiversity records to long/wide

`compute_orderwise()` Compute order-wise metrics

`rm_correlated()` Remove highly correlated predictors

`predict_dissim()` Predict pairwise compositional turnover (zeta-dissimilarity) with richness

`map_bioreg()` Raster-based clustering and interpolation of bioregional data

`map_bioregDiff()` Map bioregional change metrics between categorical raster layers

Zeta-MSGDM Workflow

Functions to automate a zeta-MSGDM workflow that fits, extracts, and visualises i-spline models for any set of zeta orders in just three function calls:

`run_ispline_models()` Run multiple Zeta.msgdm i-spline models and return both models and combined i-spline table

`plot_ispline_lines()` Plot i-spline partial effects with quantile and start-point markers

`plot_ispline_boxplots()` Plot facetted boxplots for all i-spline basis columns

Available Metrics

Use helper_indices to choose a specific function for the func parameter in compute_orderwise, enabling the calculation of one of the following metrics:

`richness()` Calculate species richness

`turnover()` Calculate species turnover or beta diversity

`abund()` Calculate species abundance

`phi_coef()` Calculate Phi coefficient

`cor_spear()` Calculate Spearman's rank correlation

`cor_pears()` Calculate Pearson's correlation

`diss_bcurt()` Calculate Bray-Curtis dissimilarity

`orderwise_diss_gower()` Compute Gower dissimilarity between two site vectors

`mutual_info()` Calculate mutual information

`geodist_helper()` Calculate geographic distance via Haversine formula

Resources

Package and Documentation

- R package documentation: https://b-cubed-eu.github.io/dissmapr/
- GitHub repository: https://github.com/b-cubed-eu/dissmapr
- B-Cubed project documentation: https://b-cubed.eu/

Citation

MacFadyen, S., Yahaya, M.M., Kumschick, S., Hui, C. (2025). dissmapr: Compositional Dissimilarity and Biodiversity Turnover Analysis. R package version 0.1.0.

MacFadyen, S., Cortes Lobos, R.B., Di Musciano, M., Hui, C., Rocchini, D. (2025). Documentation on modelled data cubes. B-Cubed project deliverable D4.1. European Commission. EU Horizon Europe Grant Agreement No. 101059592.

Key References

Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters,* 9, 683-693.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalised dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions,* 13, 252-264.

Hui, C. & McGeoch, M.A. (2014). Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. *The American Naturalist,* 184, 684-694.

Latombe, G., Hui, C., McGeoch, M.A. (2015). Multi-site generalised dissimilarity modelling: using zeta diversity to differentiate drivers of turnover in rare and widespread species. *Methods in Ecology and Evolution,* 8, 431-442.

MacFadyen, S., Yahaya, M.M., Kumschick, S., Hui, C. (2025). *M14: Code Test Compositional dissimilarity and turnover*. Project milestone 14. European Commission (EU) Horizon Europe Grant Agreement No. 101059592.

MacFadyen, S., Yahaya, M.M., Kumschick, S., Hui, C. (2025). *M13: Code development Compositional dissimilarity and turnover*. Project milestone 13. European Commission (EU) Horizon Europe Grant Agreement No. 101059592.

McGeoch, M.A., Latombe, G., Andrew, N.R., Nakagawa, S., Nipperess, D.A., Roighi, M., Hui, C. (2019). Measuring continuous compositional change using decline and decay in zeta diversity. *Ecology,* 100, e02832.