mrangr: An R package for mechanistic simulation of metacommunities

Katarzyna Markowska [*1], Michał Wawrzynowicz[1], Lechosław Kuczyński[1]

[1] Population Ecology Lab, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland

[*] Corresponding author: katarzyna.markowska@amu.edu.pl

ORCID's

Katarzyna Markowska 0000-0001-5646-4611

Michał Wawrzynowicz 0009-0001-3729-3439

Lechosław Kuczyński 0000-0003-3498-5445

Abstract

Validating complex correlative algorithms requires synthetic spatial data in which the mechanisms underlying community assembly are known. However, generating realistic "ground truth" datasets remains computationally challenging, primarily because many spatial simulators implicitly conflate the fundamental and realised niches in their environmental inputs. Here, we introduce `mrangr`, a high-performance `R` package designed for the mechanistic, spatially explicit simulation of multispecies communities. Built upon the `terra` spatial ecosystem, the software implements a population-level generalised Lotka-Volterra (GLV) framework across lattice grids. By strictly separating abiotic carrying capacities from an asymmetric interaction matrix, `mrangr` enables the realised niche to emerge dynamically through local demography, biotic interactions and dispersal. A defining feature of the package is the integrated "Virtual ecologist" module: dedicated observation layer that applies spatial masking and probabilistic detection errors to the simulation outputs, mimicking the constraints of

empirical biodiversity surveys. We demonstrate the computational flexibility of the package through three case studies: (i) generating synthetic spatiotemporal arrays to quantify the scale-dependent effects of dispersal on α, β, and γ diversity; (ii) constructing an *in-silico* sandbox to isolate the competition-colonisation trade-off; and (iii) benchmarking the ability of correlative models to successfully reconstruct fundamental niches under strong biotic constraints. By providing a complete, end-to-end generative data pipeline, `mrangr` enables computational ecologists to rigorously benchmark statistical models, optimize sampling geometries, and rigorously test hypotheses at the interface of theoretical ecology and spatial data science.

Key-words: Metacommunity simulation; Synthetic data generation; Process-based modelling; Virtual ecologist; Statistical benchmarking; Fundamental and realised niche

Running Head: Mechanistic metacommunity simulation in R

1. Introduction

Predictive macroecology increasingly relies on sophisticated statistical and machine learning algorithms to infer complex ecological relationships from spatiotemporal observational data. While methods such as Species Distribution Models (SDMs) and Joint Species Distribution Models (JSDMs) offer powerful descriptive capabilities and are widely used to untangle community assembly processes and reveal niche limits (Seoane et al., 2023), rigorously validating their predictive accuracy requires a "ground truth": synthetic data in which the mechanisms underlying population dynamics and biotic interactions are perfectly known. Metacommunity theory (M. A. Leibold et al., 2004; Thompson et al., 2020) provides the ideal mathematical framework for generating such data, explicitly linking density-independent abiotic responses, density-dependent biotic interactions, and dispersal across spatial scales.

However, a major bottleneck in generating realistic synthetic data is the conflation of the fundamental and realised niches within simulation engines. Many spatially explicit simulators rely on input suitability rasters that implicitly incorporate biotic constraints. This restricts the ability to generate datasets in which environmental filtering and community processes vary independently . In order to effectively benchmark correlative algorithms, a generative framework must treat the fundamental niche strictly as an abiotic boundary condition. This allows the realised niche to emerge dynamically through spatial simulation as a product of biotic interactions and dispersal.

Although the `R` programming environment provides a robust ecosystem for ecological data science, there is still a need for off-the-shelf, high-performance generative models that maintain this strict niche separation. Existing process-based metacommunity simulators often necessitate computational or theoretical trade-offs. For example, tools like `gen3sis` (Hagen et al., 2021) and the Python-based `MetaIBM` (Lin et al., 2024) frequently rely on individual-based modelling (IBM) or indirect fundamental niche overlap. While IBMs are highly detailed, they become computationally prohibitive across large, high-resolution geospatial grids and often lack explicit, asymmetric interaction matrices.

Conversely, highly flexible environments such as `metaRange` (Fallert et al., 2025) require users to code these functionalities themselves. Consequently, there is a pressing need for an efficient computational pipeline capable of simulating complex, asymmetric community interactions over large spatial scales.

To address this gap, we introduce `mrangr`, an `R` package designed as a scalable engine for spatially explicit, mechanistic metacommunity simulation in which dispersal, demography, and biotic interactions can be explicitly parametrised. Based on the computational solutions developed in the `rangr` package (Markowska et al., 2025), `mrangr` extends these capabilities to simulate multi-species systems by embedding a population-level generalized Lotka-Volterra (GLV) framework within the `terra` spatial ecosystem, allowing for highly efficient matrix operations across spatially explicit environmental rasters. By defining species relationships through an asymmetric interaction matrix, it explicitly models diverse biotic interactions, including competition, facilitation, and predation. Crucially, the package couples this generative engine with a "Virtual ecologist" module, enabling researchers to filter the simulated ground truth using realistic observation models. This dual architecture - generating mechanistically pure data and applying targeted sampling biases - provides a comprehensive toolkit for generating synthetic spatial datasets to benchmark ecological algorithms and assess the robustness of macroecological inferences.

2. Software architecture and computational workflow

The `mrangr` package is designed as a modular, high-performance simulation pipeline in `R`. To handle large-scale spatial data efficiently, the package's spatial infrastructure is built entirely upon the `terra` ecosystem (Hijmans, 2026), which leverages C++ for the rapid processing of multidimensional raster arrays. This architecture allows `mrangr` to seamlessly ingest empirical environmental datasets as abiotic boundary conditions, ensuring high interoperability with standard Geographic Information System (GIS) workflows.

Computationally, the core architecture of `mrangr` is designed to mechanistically decouple the fundamental niche from the realised niche (Figure 1). Users initialise the simulation by providing intrinsic growth rates (r) and spatially explicit carrying capacity grids (K), which define the fundamental niche of each species independently. The generative engine then iterates across discrete time steps, computing the emergent realised metacommunity through the integration of:

1. Demography and interaction engine: Within each raster cell, abundances of all species forming the local community are updated using a generalised Lotka-Volterra (GLV) model. Biotic interactions are governed by a user-supplied, asymmetric matrix (a), which allows interspecific dynamics to be resolved based on current cell densities.
2. Dispersal module: Following local demographic updates, spatial connectivity is achieved using species-specific dispersal kernels, executed across the simulation grid.
3. Observation layer: To bridge the gap between these theoretical mechanisms and empirical reality, the package includes a "Virtual ecologist" module. This layer applies probabilistic error functions and spatial masking to simulate the specific sampling geometries and observation biases of real-world biological surveys.

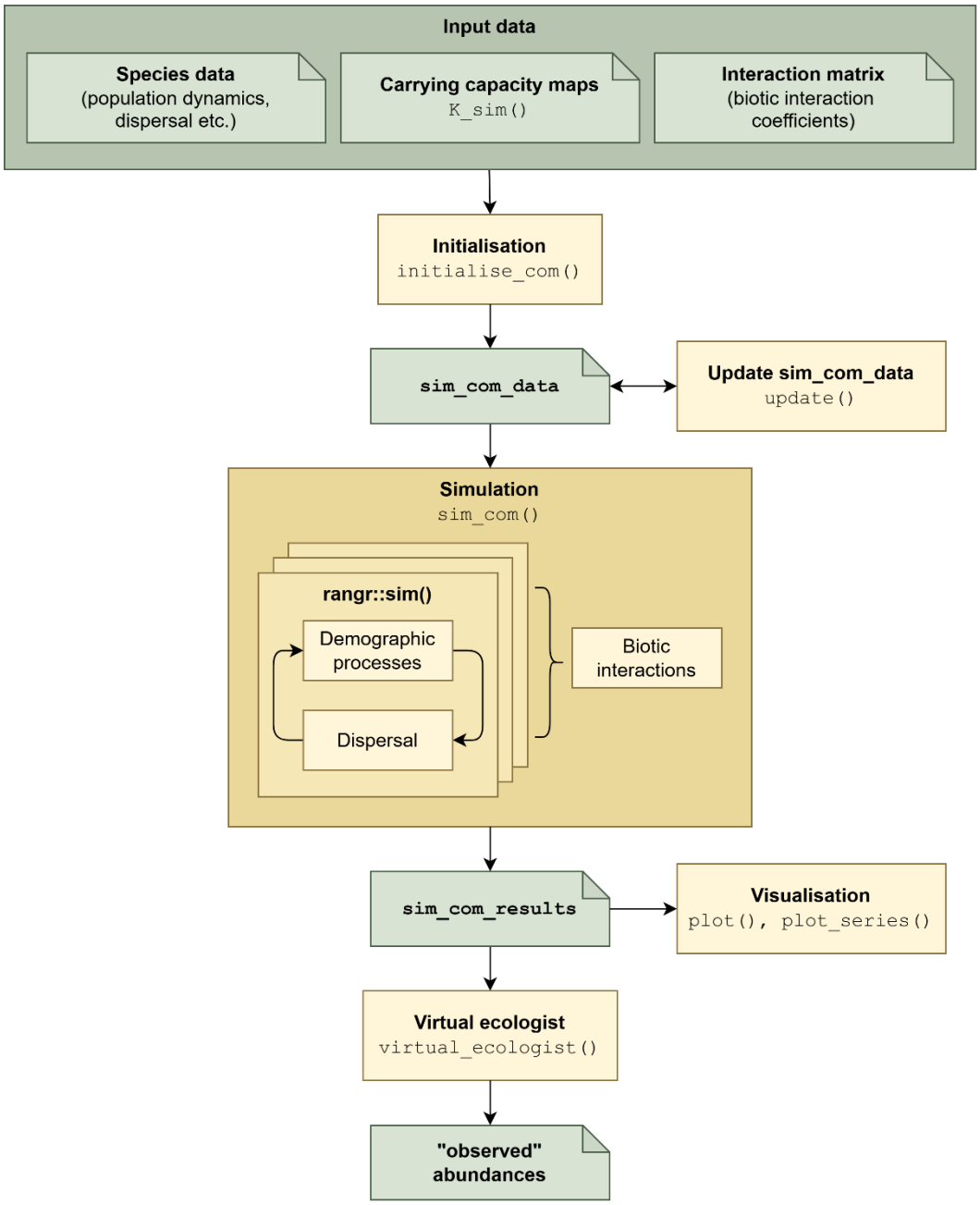

Figure 1. Conceptual framework and operational workflow of the `mrangr` package. The schematic illustrates the mechanistic decoupling of drivers: the fundamental niche is strictly defined by input carrying capacity maps, while the realized niche emerges dynamically from the integration of biotic interactions, demographic rates, and dispersal. The workflow progresses from initialisation to the "Virtual ecologist" module, which simulates observational errors. Green rectangles represent data objects (inputs and state variables), while yellow rectangles represent the package's core functions governing metacommunity dynamics and sampling.

By generating paired datasets - the true, underlying spatial arrays and the sparse, "observed" spatiotemporal records - `mrangr` provides a complete synthetic data pipeline. This structural fidelity allows researchers to input simulated datasets directly into standard analytical pipelines (e.g., SDMs or occupancy models), providing a rigorous computational platform for benchmarking statistical algorithms against a known ground truth. A comprehensive overview of the supported computational modules and observational processes is provided in **Error! Reference source not found.**.

Table 1. Architectural overview of the `mrangr` computational framework, distinguishing between the ecological state model (the numerical engine simulating true spatial abundance) and the observation model (the data-degradation pipeline generating synthetic survey datasets).

Computational module	Ecological function	Software implementation & data structures
ECOLOGICAL STATE MODEL		
Abiotic boundary conditions (Fundamental niche)	Defines the potential spatial range and local carrying capacity strictly via environmental parameters, independent of community interactions.	Users input spatially explicit grid arrays (K), supplied either as static `terra` rasters or generated dynamically from multidimensional environmental layers via the `K_sim()` function.
Biotic filtering (Realised niche)	Modifies the fundamental niche by reducing abundance (competition, predation) or expanding it (facilitation), creating the realised distribution.	Driven by a user-supplied asymmetric interaction matrix (a), the package uses a generalised Lotka-Volterra (GLV) algorithm to iteratively solve for local cell abundances, allowing the realised niche to emerge dynamically.
Spatial connectivity (Dispersal)	Simulates spatial fluxes across the lattice grid. Low rates cause dispersal limitation, preventing species from reaching suitable patches. High rates drive mass effects (source-sink dynamics) and rescue effects.	Users control the spread via the `kernel_fun` parameter in `initialise_com()`. This allows for modelling constrained diffusion (limitation) or fat-tailed distributions (long distance dispersal) to simulate different isolation scenarios.
Stochastic noise (Ecological drift)	Injects stochasticity to simulate random variance in local population density.	Demographic stochasticity is inherent to the simulation. Additional noise can be introduced into demographic rates or environmental layers using `initialise_com()` or `update()` functions.

OBSERVATION MODEL		
Synthetic sampling (Virtual ecologist)	Distorts the 'true' simulated abundance arrays through spatial masking and probabilistic errors to generate benchmarking datasets that mimic empirical survey constraints.	Executed via the `virtual_ecologist()` function. Users define spatial sampling geometries (e.g., random, systematic) and apply probability distributions via the `obs_error` parameter to simulate imperfect detection rates.

3. Core features and ecological capabilities

While `mrangr` inherits the spatially explicit architecture of the `rangr` package (Markowska et al., 2025), it expands its scope to handle complex, multispecies networks. The following features distinguish the package as both a robust simulation engine for theoretical ecology and a practical tool for benchmarking statistical methods (e.g., SDMs and JSDMs).

3.1. Mechanistic simulation of the realised niche

Biotic interactions are modelled via a square numeric matrix where each element a_{ij} represents the *per-capita* interaction strength of species j on species i. Mechanistically, this coefficient defines the change in the carrying capacity of species i caused by a single individual of species j. Consequently, the realised niche is calculated dynamically: at each time step, the effective carrying capacity of a focal species is derived by modifying its fundamental niche (K_{fund}) by the net biotic influence of the community.

The core engine uses a generalised Lotka-Volterra (GLV) framework to update the effective carrying capacity for species i at time t in a given spatial cell:

$$K_{i,t} = max\left(K_{i,fund} + \sum_{j=1}^{S}\left(a_{i,j} \cdot N_{j,t-1}\right), 0\right)$$

where S is the total number of species, $K_{i,fund}$ is the fundamental niche of the species i, $N_{j,t-1}$ is the abundance of interacting species, and the $max(\ldots, 0)$ function ensures that carrying capacity remains non-negative. Because the software applies interactions strictly to the

carrying capacity rather than intrinsic growth rates, it structurally enforces the theoretical distinction between environmental filtering (the fundamental niche) and biotic constraints (the realised niche).

3.2. An accessible workflow for ecologists

The package has been designed to minimise the coding burden on researchers. The workflow relies on just two primary functions: `initialise_com()` prepares the simulation by linking spatial maps with demographic parameters, and `sim_com()` executes the spatial iterations over time. This intuitive setup allows ecologists to easily loop over parameter sets to conduct sensitivity analyses or generate large datasets.

3.3. Simulating invasion dynamics

The package offers specialised functionality to simulate species invasions. Users can designate specific species as invaders and schedule their introduction at defined time steps, rather than initializing them at the start of the simulation. This temporal flexibility makes it straightforward to simulate biological invasions, priority effects, and community assembly dynamics over time.

3.4. The "Virtual ecologist" module

To connect theoretical simulations with empirical data, the package includes a `virtual_ecologist()` module (Zurell et al., 2010). This function mimics the constraints of real-world biodiversity surveys by applying spatial sampling designs (e.g., surveying only a small percentage of the landscape) and imperfect detection probabilities (e.g., using Binomial or log-Normal error distributions). As data quality issues, such as spatial sampling gaps and detection biases, remain critical vulnerabilities in predictive modelling (Barbosa & Alves-Souza, 2025), it is essential to build on recent efforts to make virtual species more realistic (Malinowska et al., 2023) by generating "observed" datasets alongside known, mechanistically derived ground truths. This dual-output framework allows researchers to rigorously

benchmark statistical methods and quantify how sampling limitations affect ecological inference (Meynard et al., 2019).

3.5. Synthetic landscape generation

For theoretical experiments, `mrangr` provides a built-in environmental generator. The `K_sim()` function uses Gaussian Random Fields to create spatially autocorrelated landscapes. Ecologists can easily generate complex, cross-correlated environmental gradients to test how spatial heterogeneity influences metacommunity structure.

3.6. GIS integration

Unlike standalone simulators that rely on abstract spatial grids, `mrangr` is built entirely on the popular `terra` ecosystem (Hijmans, 2026). Ecologists can directly import empirical GIS data, such as high-resolution climate rasters or habitat maps, to serve as the abiotic boundary conditions for their simulations.

3.7. Computational efficiency

Spatially explicit simulations can be computationally expensive, particularly when scaled up to cover large landscapes or a high number of species. `mrangr` addresses this issue by delegating intensive spatial operations to the `terra` package, which is optimised using the `C++` programming language. This enables the package to retain the flexibility and readability of pure `R` code, while delivering the performance required for handling large landscape grids and extensive replication. Furthermore, the package is designed to support parallel execution. As demonstrated in the provided case studies, users can easily distribute replicates across processor cores using standard `R` parallelisation tools (e.g., `parallel`, `pbapply`), which makes it feasible to conduct extensive sensitivity analyses and robustly estimate parameter uncertainty.

4. The **mrangr** workflow

The `mrangr` package is designed around a linear, reproducible workflow that guides users from data preparation to final synthetic sampling. The pipeline consists of four distinct stages:

4.1. Step 1: Defining the environment and biological parameters

Before running a simulation, the user must define the abiotic and biotic rules of the metacommunity. The spatial environment is established by providing carrying capacity maps (the fundamental niche) for each species. These can be empirical GIS layers imported via the `terra` package or synthetic landscapes generated using the `K_sim()` function. The biotic rules are defined by supplying species-specific demographic rates (e.g., intrinsic growth rates) and an interaction matrix (a), which explicitly specifies the biotic interactions between all species in the network.

4.2. Step 2: Initialising the metacommunity

Once the input data are prepared, the `initialise_com()` function is used to construct the baseline simulation object. This function binds the spatial rasters, demographic parameters, and interaction matrix together. At this stage, the user also defines the dispersal kernel (how far and fast individuals move) and the initial starting distributions of the species. The output is a structured `R` object containing the spatial abundance of the metacommunity at time zero ($t = 0$).

4.3. Step 3: Running the simulation engine

With the metacommunity initialised, the `sim_com()` function executes the temporal simulation loop. For each time step, the function iteratively updates the carrying capacity landscape. Within every spatial cell, local population abundances are recalculated using the generalised Lotka-Volterra equations, resolving the density-dependent biotic interactions to form the realised niche. Following local demography, the dispersal module redistributes individuals across the grid. The simulation

returns a multidimensional spatial array containing the "true" abundance of every species across all time steps.

4.4. Step 4: Simulating the observation process

The final, optional step in the workflow is to translate the deterministic "true" abundances into realistic, noisy survey data. Using the `virtual_ecologist()` function, users can apply spatial sampling designs (e.g., random plots) and observation models (e.g., imperfect detection or counting errors). This transforms the fully realised spatial grids into standard ecological data formats, such as presence/absence matrices or count data frames, ready to be analysed by standard statistical tools like Species Distribution Models (SDMs) or Joint Species Distribution Models (JSDMs). By controlling the sampling intensity, users can generate datasets to test specific methodological bottlenecks, such as assessing algorithmic accuracy across varying sampling efforts (Fernandes et al., 2018) or validating model evaluation techniques for rare species with low sample sizes (Collart & Guisan, 2023).

5. Case studies

We present three case studies to validate the simulator against established ecological theory. The first two examples benchmark `mrangr` against known biological patterns: the influence of dispersal on biodiversity scaling and the dynamics of competition-colonisation trade-offs. The third example demonstrates the package's methodological utility, evaluating the limitations and potential of inferring fundamental niches from observation-based data.

The three case studies were run in the same exemplary simulation environment, defined on a 20×20 grid (400 cells) with a 1 km resolution, assuming a coordinate system EPSG:2180. Trends in simulated parameters were quantified and visualised using Generalized Additive Models for Location, Scale and Shape (GAMLSS) to capture non-linear responses and heteroscedasticity.

5.1. Case study 1: Simulating the scale-dependent effects of dispersal on biodiversity

Dispersal is the fundamental process connecting local communities and shaping biodiversity patterns at multiple scales. In metacommunity theory, varying dispersal rates creates a well-documented spatial pattern: dispersal promotes local coexistence through the rescue effect, yet potentially undermines regional diversity by homogenizing distinct communities (Mouquet & Loreau, 2003). Increased dispersal should theoretically elevate local richness (α-diversity) by overcoming dispersal limitation, while simultaneously eroding compositional turnover (β-diversity) through mass effects. At the regional scale (γ-diversity), these opposing forces may generate a unimodal response, where biodiversity peaks at intermediate dispersal rates that balance colonisation against competitive exclusion. Testing these predictions empirically is challenging due to the difficulty of manipulating dispersal traits. Here, we demonstrate how `mrangr` can be used to rigorously test these macroecological hypotheses by simulating metacommunities across a controlled gradient of dispersal ranges while keeping niche requirements and interaction strengths constant.

In this example, the metacommunity consisted of 20 species. For each simulation replicate, species-specific carrying capacity maps (K) were generated using spatially autocorrelated log-normal distributions. Biotic interactions were modelled via an asymmetric interaction matrix (a) with coefficients drawn from a normal distribution. The experimental gradient focused on dispersal ability. We varied the mean dispersal distance from 100 m to 3000 m across 30 discrete intervals. Dispersal was modelled using an exponential kernel, where the rate parameter is defined as $1/mean$ distance.

We performed 100 independent replicates for each dispersal scenario. Each simulation ran for 20 time steps, sufficient to allow the metacommunity to reorganise from its initial state under the imposed dispersal and interaction constraints. At the final time step, we calculated diversity metrics based on Hill numbers with $q = 1$ (exponential of Shannon entropy):

1. Alpha diversity (α): calculated as the mean local diversity across all 400 grid cells.

2. Gamma diversity (γ): calculated based on the total pooled abundance of each species across the entire landscape.
3. Beta diversity (β): derived using additive partitioning: $\beta = \gamma - \alpha$.

The simulation results confirm the opposing effects of dispersal on biodiversity across spatial scales, reproducing classic theoretical predictions (e.g., Mouquet & Loreau, 2003):

1. Local enrichment (α-diversity): As predicted, local species richness increased monotonically with dispersal ability (Figure 2a). At low dispersal rates, local communities become impoverished due to local extinctions and dispersal limitation. Increasing connectivity allows species to colonise and persist in suboptimal patches ('sink' habitats) via the rescue effect, thereby increasing local diversity.
2. Spatial homogenization (β-diversity): Conversely, compositional turnover declined sharply as dispersal increased (Figure 2b). Long dispersal distances effectively homogenise the metacommunity, eroding the spatial distinctions driven by environmental heterogeneity.
3. The regional trade-off (γ-diversity): The response of regional diversity highlights the tension between local enrichment and spatial homogenisation (Figure 2c). Gamma diversity increases rapidly at low dispersal distances as species overcome dispersal limitation, eventually saturating at a stable plateau. Unlike simple theoretical models that predict a decline in diversity at high dispersal rates due to global competitive exclusion, our results indicate that spatial heterogeneity in carrying capacity provides sufficient refuge for inferior competitors. In this high-dispersal regime, species sorting mechanisms allow species to efficiently track their environmental optima without being displaced from the landscape entirely, maintaining high regional diversity despite extensive mixing.

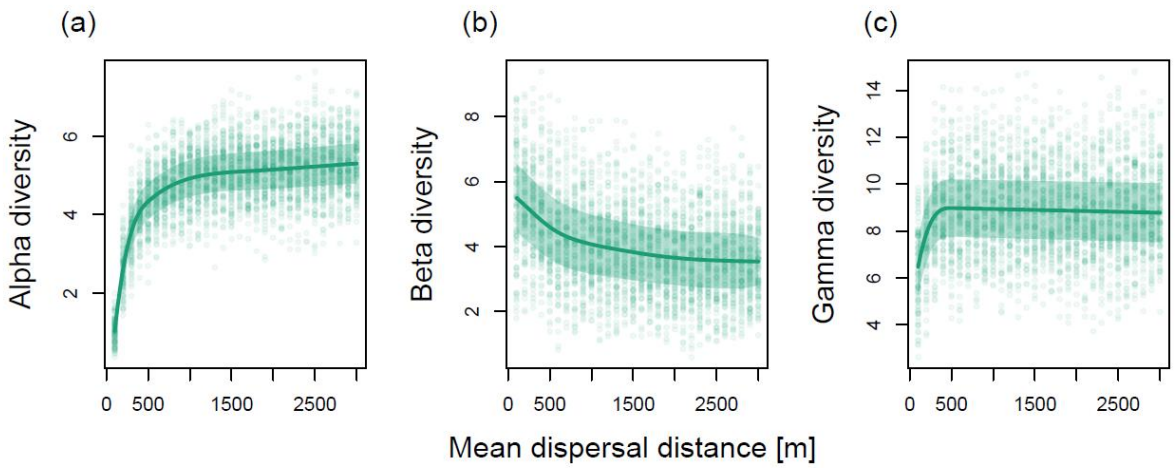

Figure 2. Response of metacommunity diversity components to mean dispersal distance. Scatterplots display (a) alpha, (b) beta, and (c) gamma diversity for metacommunities simulated with a regional pool of 20 species. Points represent individual simulation runs. Solid lines indicate the median and shaded regions represent the interquartile range, modelled using a Gaussian Location-Scale GAM (GAMLSS).

5.2. Case study 2: Simulating the competition-colonisation trade-off

A fundamental puzzle in community ecology is explaining how inferior competitors avoid exclusion in landscapes dominated by superior species. The competition–colonisation trade-off hypothesis provides a classic metacommunity solution, proposing that species coexist by partitioning the landscape based on dispersal ability rather than resource use (Tilman, 1994). In this framework, inferior competitors persist as "fugitive species" by investing in superior colonisation rates, allowing them to occupy vacant patches before slower-dispersing dominants arrive to displace them.

Identifying this exact trade-off in empirical systems is notoriously difficult because observed spatial patterns are often distorted by environmental heterogeneity and complex trait correlations. Here, we demonstrate how `mrangr` can be used as an *in-silico* sandbox to isolate and test this hypothesis computationally.

In this example, the metacommunity consisted of just two virtual species. To isolate the effect of dispersal on coexistence, we controlled for environmental preferences by enforcing complete fundamental niche overlap. Both species were assigned identical spatial habitat requirements, differing only in their competitive fitness within that niche:

1. The superior competitor (Species 1): Assigned a baseline carrying capacity generated via a log-normal distribution.
2. The inferior competitor (Species 2): Assigned a carrying capacity 20% lower than Species 1, across the entire landscape.
3. Biotic interactions: We modelled strong, symmetric competition between the species ($a = -1$). Under these conditions - identical fundamental niches and distinct fitness levels - the theory predicts the deterministic exclusion of Species 2 by Species 1 in every grid cell.

We introduced a trade-off where the inferior competitor (Species 2) compensated for its lower fitness with superior dispersal. We fixed the mean dispersal distance of Species 2 at 1000 m and systematically varied the mean dispersal of the superior competitor (Species 1) across three scenarios:

1. No trade-off (Control): Species 1 also disperses 1000 m (equal dispersal, unequal fitness).
2. Moderate trade-off: Species 1 disperses 100 m (10× disadvantage).
3. Strong trade-off: Species 1 disperses 10 m (100× disadvantage).

We performed 40 independent replicates of each scenario, with each replicate comprising 50 time steps. To evaluate whether spatial niche partitioning (via colonisation ability) could prevent exclusion despite a lack of niche differentiation, we tracked the relative abundance of the inferior competitor.

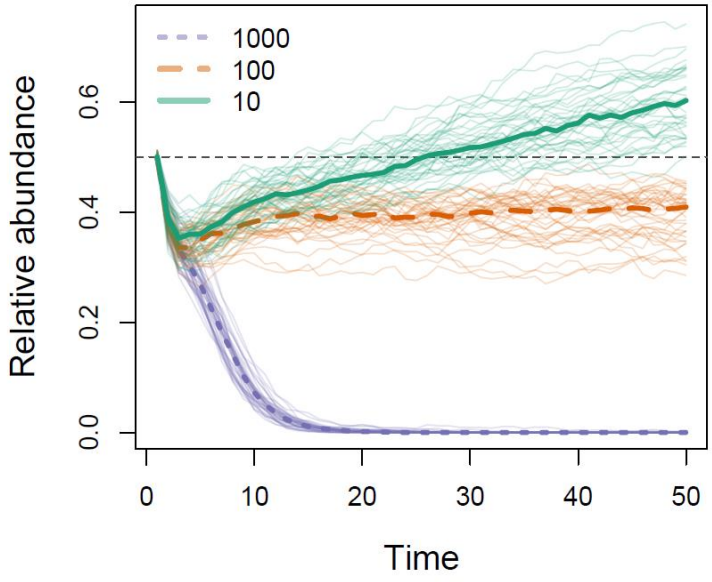

Figure 3. Testing the competition-colonization trade-off. Temporal dynamics of the inferior competitor's relative abundance over 50 simulation steps. The inferior competitor (Species 2) has a fixed high dispersal distance (1000 m) but lower competitive fitness ($K_2 = 0.8 \times K_1$). Thin lines represent individual simulation trajectories (n=40), while thick lines indicate the median. Scenarios differ by the mean dispersal distance of the superior competitor (Species 1): 1000 m (violet, dotted line), 100 m (orange, dashed line), and 10 m (green, solid line).

The simulations demonstrate that a dispersal advantage can effectively counteract competitive exclusion (Figure 3). In the absence of a trade-off, when both species shared equal dispersal capabilities, the inferior competitor was rapidly driven towards extinction (violet dotted line). However, as the strength of the trade-off increased, the persistence of the inferior competitor improved significantly. In the strongest trade-off scenario (green solid line), in which the superior competitor was severely dispersal-limited (10 m), the inferior competitor successfully exploited vacant space and achieved numerical dominance despite its lower fitness.

5.3. Case study 3: Benchmarking the statistical recovery of the fundamental niche

Estimating the fundamental niche from field data is complicated by two filters: biotic interactions, which constrain the realised distribution, and observational errors, which distort detection. Although correlative statistical models, such as SDMs, are typically used for this task, it is well documented that they are limited in their ability to account for complex biotic constraints (Jarnevich et al., 2015). Consequently, ecological field data rarely reflect pure environmental potential (Soberón, 2007). However, recovering this baseline is essential for forecasting species responses to novel environments. In this example, we use `mrangr` to simulate a known ground truth and systematically evaluate whether statistical models can reliably reconstruct the fundamental niche.

Spatially autocorrelated environmental variables were generated using Gaussian Random Fields via the `K_sim()` function. The metacommunity consisted of 5 virtual species. For each species, the fundamental niche (carrying capacity, K) was defined as a log-linear function of the environmental covariates, ensuring a known ground truth for species-environment relationships. To model the realised niche, we generated asymmetric interaction matrices (a) where off-diagonal elements were drawn from a normal distribution $N(0, \delta^2)$. We systematically varied the interaction strength parameter, δ, across a gradient to simulate metacommunities ranging from purely abiotic-driven ($\delta = 0$) to highly interactive systems ($\delta = 3$).

Simulations were initialised with abundances drawn from a Poisson distribution with expectations equal to the local carrying capacity ($\lambda = K$). The system evolved over 50 time steps. The first 10 steps served as a burn-in period, during which the community reached a quasi-equilibrium state. To replicate the spatiotemporal structure of empirical monitoring datasets, we employed the "Virtual ecologist" module across the subsequent 40 time steps. We sampled 10% of the available site-time combinations (`prop = 0.1`) and introduced observational error using a binomial distribution with detection probability $p = 0.5$, mimicking the imperfect detection typical of wildlife surveys.

We attempted to reconstruct the fundamental niche from the sampled realised abundances using Generalized Linear Mixed Models (GLMMs) fitted via the `glmmTMB` package. The models included the true environmental covariates as predictors. We evaluated the performance of these reconstructions against the true fundamental niche (K) using three metrics:

1. Bias of β: The absolute difference between the estimated environmental coefficient and the true coefficient used to generate K.
2. Sign agreement: The proportion of simulations where the model correctly identified the direction of the environmental response (positive/negative).
3. Correlation with K: The Spearman rank correlation between the spatially predicted abundance surface and the true carrying capacity map.

Our simulations demonstrate that interaction strength substantially impairs the statistical recovery of the fundamental niche. As the interaction strength increased, the spatial correlation between the reconstructed niche and the true carrying capacity declined non-linearly, effectively uncoupling realised abundance from environmental potential (Figure 4c). At the same time, the absolute bias in the estimated environmental coefficients (β) increased (Figure 4a), indicating that biotic constraints systematically distort the perceived magnitude of environmental preferences. Most critically, under strong biotic regulation, the sign agreement dropped toward 0.5 (Figure 4b), which is equivalent to random guessing. This implies that in highly interactive communities, standard correlative models frequently misidentify positive environmental associations as negative (and vice versa), resulting in spurious niche estimates driven by community dynamics rather than abiotic suitability.

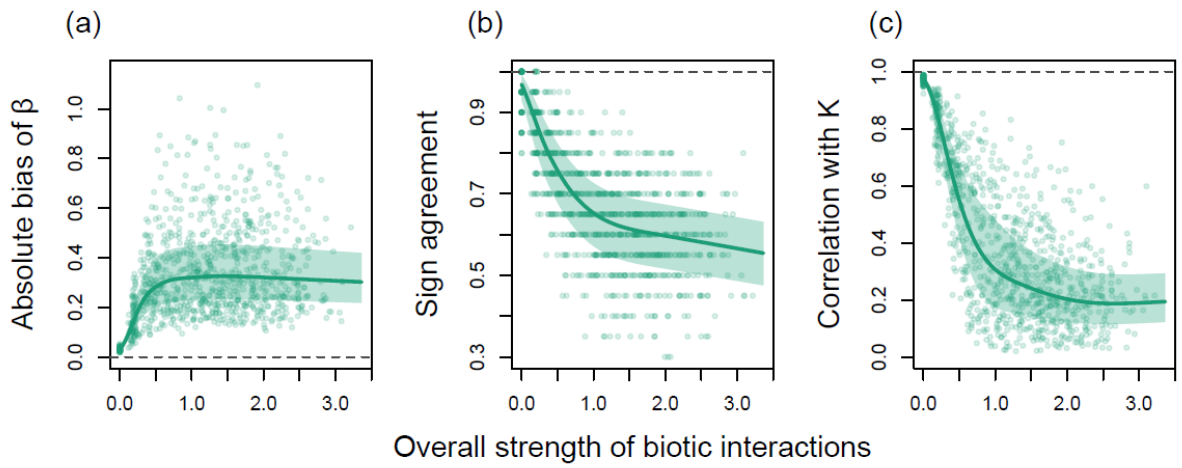

Figure 4. Influence of biotic interaction strength on the accuracy of fundamental niche estimation by the "Virtual ecologist" . Interaction strength is defined as the mean absolute value of off-diagonal elements in the interaction matrix. Estimation performance is evaluated via: (a) absolute bias of slope estimates (β); (b) sign agreement (proportion of estimated slopes matching the true sign); and (c) correlation between estimated abundances and true carrying capacity (K). Points represent individual metacommunities. Solid lines indicate the median and shaded regions represent the interquartile range, modelled using a Gaussian Location-Scale GAM (GAMLSS). Dashed horizontal lines indicate reference values for optimal performance (zero bias or perfect agreement/correlation).

6. Conclusions

While the metacommunity concept was traditionally framed within four distinct paradigms - species sorting, mass effects, patch dynamics, and neutral theory (M. A. Leibold et al., 2004) - recent work has shifted towards a unified, process-based framework (M. A. Leibold et al., 2022). However, a key challenge remains: operationalising this synthesis to accurately link observed spatial patterns back to their underlying generative processes (M. Leibold et al., 2025). To address this, `mrangr` simplifies complex metacommunity dynamics into a computationally tractable pipeline driven by three fundamental axes - space, time, and species - linked through explicit ecological processes.

Although this abstraction is necessarily simplified, `mrangr` successfully captures both the biotic interactions that drive local species coexistence, and the spatiotemporal population dynamics that determine community assembly. A defining architectural feature of the package is the explicit separation of the fundamental and realised niches. This ensures that simulated co-distribution patterns dynamically emerge from environmental potential and biotic networks that are parametrised as completely independent inputs. Furthermore, by incorporating stochasticity across local demography and spatial dispersal, `mrangr` enables the generation of robust parameter distributions from replicated simulations, which is essential for rigorous statistical inference.

Ultimately, the package provides an accessible and high-performance framework to explore limits of ecological inference, particularly the identifiability of species interactions from observational data (Barbier et al., 2025). By enabling fully spatially explicit virtual experiments with arbitrary biotic topologies and an integrated observation module, `mrangr` moves macroecology far beyond traditional, static null models. It serves as a comprehensive benchmarking platform for testing advanced analytical methods that map metacommunity dynamics onto species co-distribution patterns, such as Joint Species Distribution Models (Christopher D. Terry et al., 2023; Morueta-Holme et al., 2016; Ovaskainen et al., 2017, 2019). Through this mechanistic control, researchers can replicate established biodiversity patterns while computationally exploring complex frontiers, such as disentangling abiotic filtering from competitive exclusion, assessing the interplay between niche and fitness differences, or forecasting the spatiotemporal dynamics of biological invasions.

7. Authors' contributions

KM: Conceptualization, Project administration, Software, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing; **MW**: Validation, Writing - Original Draft, Writing - Review & Editing; **LK**: Conceptualization, Funding acquisition, Methodology, Software, Supervision, Visualization, Writing - Original Draft, Writing - Review & Editing;

8. Funding Statement

The study was supported by the National Science Centre (NCN) in Poland (grant no. 2018/29/B/NZ8/00066). The computational resources supporting this work were provided by the Poznań Supercomputing and Networking Centre (PCSS) (grant no. pl0090-01).

9. Conflict of interest statement

The authors declare no conflicts of interest.

10. Data availability statement

- The `mrangr` package is available on CRAN (https://cran.r-project.org/package=mrangr). It comes with built-in function documentation and a vignette demonstrating its main functionality and workflow logic.
- Contact details for issues with the package: https://github.com/popecol/mrangr/issues
- The package's source code can be accesses on GitHub (https://github.com/popecol/mrangr) and Zenodo (https://doi.org/10.5281/zenodo.18641951).
- Package's website is hosted at https://popecol.github.io/mrangr/.
- The code and data used in the case studies are available on Zenodo (https://doi.org/10.5281/zenodo.18643290).
- The preprint of this manuscript is available at: https://ecoevorxiv.org/repository/view/11808/.

11. References

Barbier, M., Bunin, G., & Leibold, M. A. (2025). Getting more by asking for less: Linking species interactions to species co-distributions in metacommunities. *Peer Community Journal*, *5*. https://doi.org/10.24072/pcjournal.483

Barbosa, W. L., & Alves-Souza, S. N. (2025). Data quality issues in data used in species distribution models: A systematic literature review. *Ecological Informatics*, *91*, 103378. https://doi.org/10.1016/j.ecoinf.2025.103378

Christopher D. Terry, J., Langdon, W., & Rossberg, A. G. (2023). Codistribution as an indicator of whole metacommunity response to environmental change. *Ecography*, *2023*(7), e06605. https://doi.org/10.1111/ecog.06605

Collart, F., & Guisan, A. (2023). Small to train, small to test: Dealing with low sample size in model evaluation. *Ecological Informatics*, *75*, 102106. https://doi.org/10.1016/j.ecoinf.2023.102106

Fallert, S., Li, L., & Cabral, J. S. (2025). metaRange: A framework to build mechanistic range models. *Methods in Ecology and Evolution*, *16*(1), 49–56. https://doi.org/10.1111/2041-210X.14461

Fernandes, R. F., Scherrer, D., & Guisan, A. (2018). How much should one sample to accurately predict the distribution of species assemblages? A virtual community approach. *Ecological Informatics*, *48*, 125–134. https://doi.org/10.1016/j.ecoinf.2018.09.002

Hagen, O., Flück, B., Fopp, F., Cabral, J. S., Hartig, F., Pontarp, M., Rangel, T. F., & Pellissier, L. (2021). gen3sis: A general engine for eco-evolutionary simulations of the processes that shape Earth's biodiversity. *PLOS Biology*, *19*(7), e3001340. https://doi.org/10.1371/journal.pbio.3001340

Hijmans, R. J. (2026). *terra: Spatial Data Analysis* (Version R package version 1.8-94) [Computer software]. https://rspatial.org/

Jarnevich, C. S., Stohlgren, T. J., Kumar, S., Morisette, J. T., & Holcombe, T. R. (2015). Caveats for correlative species distribution modeling. *Ecological Informatics*, *29*, 6–15. https://doi.org/10.1016/j.ecoinf.2015.06.007

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*(7), 601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x

Leibold, M. A., Rudolph, F. J., Blanchet, F. G., De Meester, L., Gravel, D., Hartig, F., Peres-Neto, P., Shoemaker, L., & Chase, J. M. (2022). The internal structure of metacommunities. *Oikos*, *2022*(1), oik.08618. https://doi.org/10.1111/oik.08618

Leibold, M., Barbier, M., Bittleston, L., Clark, A. T., Cuellar-Gempeler, C., D'Andrea, R., Frans, V. F., Khattar, G., Miller, Z., Peres-Neto, P. R., & Wisnoski, N. I. (2025). *Linking Pattern to Process in Metacommunities: Challenges and Opportunities*. https://ecoevorxiv.org/repository/view/9021/

Lin, J.-H., Quan, Y.-J., & Han, B.-P. (2024). MetaIBM: A Python-based library for individual-based modelling of eco-evolutionary dynamics in spatial-explicit metacommunities. *Ecological Modelling*, *492*, 110730. https://doi.org/10.1016/j.ecolmodel.2024.110730

Malinowska, K., Markowska, K., & Kuczyński, L. (2023). Making virtual species less virtual by reverse engineering of spatiotemporal ecological models. *Methods in Ecology and Evolution*, *14*(9), 2376–2389. https://doi.org/10.1111/2041-210X.14176

Markowska, K., Malinowska, K., & Kuczyński, L. (2025). rangr: An R package for mechanistic, spatially explicit simulation of species range dynamics. *Methods in Ecology and Evolution*, *16*(3), 468–476. https://doi.org/10.1111/2041-210X.14475

Meynard, C. N., Leroy, B., & Kaplan, D. M. (2019). Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? *Ecography*, *42*(12), 2021–2036. https://doi.org/10.1111/ecog.04385

Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B. J., Peet, R. K., Ott, J. E., Violle, C., Enquist, B. J., Jørgensen, P. M., & Svenning, J. (2016). A network approach for inferring species associations from co-occurrence data. *Ecography*, *39*(12), 1139–1150. https://doi.org/10.1111/ecog.01892

Mouquet, N., & Loreau, M. (2003). Community Patterns in Source-Sink Metacommunities. *The American Naturalist*, *162*(5), 544–557. https://doi.org/10.1086/378857

Ovaskainen, O., Rybicki, J., & Abrego, N. (2019). What can observational data reveal about metacommunity processes? *Ecography*, *42*(11), 1877–1886. https://doi.org/10.1111/ecog.04444

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, *20*(5), 561–576. https://doi.org/10.1111/ele.12757

Seoane, J., Estrada, A., Jones, M. M., & Ovaskainen, O. (2023). A case study on joint species distribution modelling with bird atlas data: Revealing limits to species' niches. *Ecological Informatics*, *77*, 102202. https://doi.org/10.1016/j.ecoinf.2023.102202

Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, *10*(12), 1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Thompson, P. L., Guzman, L. M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., Viana, D. S., & Chase, J. M. (2020). A process-based metacommunity framework linking local and regional scale community ecology. *Ecology Letters*, *23*(9), 1314–1329. https://doi.org/10.1111/ele.13568

Tilman, D. (1994). Competition and Biodiversity in Spatially Structured Habitats. *Ecology*, *75*(1), 2–16. https://doi.org/10.2307/1939377

494 Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J.,
495 Reineking, B., Schröder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and
496 observers. *Oikos*, *119*(4), 622–635. https://doi.org/10.1111/j.1600-0706.2009.18284.x

497