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

Metacommunity theory unifies ecology by integrating local biotic interactions with regional dispersal and environmental filtering. However, testing theoretical predictions against empirical data remains challenging due to the difficulty of disentangling these processes in nature and the confounding effects of imperfect detection. Here, we introduce `mrangr`, an R package designed for the mechanistic, spatially explicit simulation of multispecies communities. Unlike correlative approaches, `mrangr` strictly distinguishes between the fundamental niche (determined by abiotic carrying capacity) and the realised niche (an emergent property of biotic interactions). The package implements a generalized Lotka-Volterra framework on a lattice grid (via the `terra` ecosystem), allowing users to simulate diverse interaction types — including competition, predation, and facilitation — alongside species-specific dispersal kernels. A defining feature is the 'virtual ecologist' module, which samples the simulated ground truth with user-defined observation errors and sampling designs, thereby mimicking the constraints of real-world

biodiversity surveys. We demonstrate the package's capabilities through three case studies: (i) quantifying the scale-dependent effects of dispersal on α, β, and γ diversity, (ii) testing the conditions under which the competition-colonization trade-off promotes coexistence in the presence of fitness inequalities, and (iii) assessing the recoverability of fundamental niches from imperfect observational data constrained by biotic interactions. By providing a flexible platform to generate synthetic spatial data with known underlying mechanisms, `mrangr` enables researchers to benchmark complex statistical models (e.g. JSDMs), assess sampling strategies, and rigorously test hypotheses at the interface of theoretical and empirical macroecology.

Key-words: metacommunity dynamics, fundamental and realized niche, community assembly, process-based modelling, virtual ecologist, spatially explicit

Running Head: Mechanistic metacommunity simulation in R

1. Introduction

One of the key goals of ecology is to understand the mechanisms generating and maintaining biodiversity. Traditionally, theory separated these mechanisms by spatial scale: regional frameworks emphasized ecological drift, selection, speciation, and dispersal (Vellend, 2016), whereas local models focused on competitive coexistence (Chesson, 2000), and priority effects (Adler et al., 2007; Ke & Letten, 2018). Metacommunity theory (M. A. Leibold et al., 2004, 2017) unifies these perspectives by identifying three spatially explicit mechanisms that operate across scales: density-independent responses to abiotic conditions, density-dependent biotic interactions, and dispersal (Thompson et al., 2020). While explicitly integrating these mechanisms is essential for understanding biodiversity patterns, a central debate in macroecology remains the extent to which local biotic interactions leave a signature at regional and continental scales (Araújo & Rozenfeld, 2014). Testing such spatial scaling theories is currently hindered by a critical limitation in many existing metacommunity tools: the conflation of the fundamental and realized niche. Frequently, simulators rely on input suitability maps that implicitly incorporate biotic constraints, rendering it impossible to disentangle environmental filtering from community processes. To overcome this, a mechanistic framework must strictly define the fundamental niche as a measure of environmental potential, allowing the realized niche to emerge purely as a dynamic property of biotic interactions and dispersal.

The R programming environment has become the standard for macroecological research, offering robust tools to evaluate empirical metacommunity structure from static distribution matrices (e.g., Dallas, 2014). However, although these analytical packages excel at quantifying spatial patterns, they cannot simulate the underlying mechanistic processes that generate them. To move from correlative pattern description to process-based hypothesis testing, researchers require generative models capable of explicitly simulating community assembly over time and space. However, despite the need for mechanistic clarity, the current landscape of process-based metacommunity simulators – namely `gen3sis` (Hagen et al., 2021) and `metaIBM` (Lin et al., 2024) - often necessitates trade-offs. While

these frameworks share a core principle of coupling dispersal and density-dependent regulation, they implement interspecific regulation indirectly by the magnitude of fundamental niche overlap. This reliance on overlap prevents the explicit modelling of metacommunities with asymmetric (e.g. predation) or positive (e.g. facilitation) interactions, effectively locking the simulation into the "conflated niche" paradigm. Conversely, `metaRange` (Fallert et al., 2025) provides a programming environment in which multiple processes, including all types of biotic interactions, can be modelled flexibly; yet users must code these functionalities themselves. Consequently, no off-the-shelf tool is currently capable of simulating metacommunities with flexible, asymmetric biotic interactions while maintaining a strict separation between abiotic and biotic drivers.

To address this gap, we have developed `mrangr`: an R package for flexible, mechanistic metacommunity simulation in which dispersal processes, demographic rates and biotic interaction coefficients can be explicitly parametrised. Built as a multi-species extension of the `rangr` package (Markowska et al., 2025), `mrangr` inherits that tool's accessible parameterisation of population dynamics, dispersal, and virtual ecology. By representing species relationships through an asymmetric interaction matrix, it enables the simulation of diverse biotic interactions - including competition, facilitation, and predation - within a spatially explicit virtual environment. Moreover, by defining the fundamental niche strictly through user-supplied carrying capacity maps, the package allows the realized niche to emerge dynamically. This separation enables researchers to quantify the ”biotic deficit” - the specific loss of range or abundance attributable solely to biotic interactions - by comparing the input carrying capacity maps against the simulated equilibrium state.

2. Package overview

The core architecture of mrangr is designed to mechanistically decouple the fundamental niche from the realized niche (Figure 1). Users define the environmental potential for each species via spatially explicit carrying capacity maps (K) and intrinsic growth rates (r), while biotic constraints are governed

by a user-supplied asymmetric interaction matrix (a). Consequently, the realized metacommunity emerges dynamically from the interplay of species-specific demography, dispersal constraints, and pairwise interactions. To bridge the gap between these theoretical mechanisms and empirical reality, the package includes a 'virtual ecologist' module that replicates both observation error and the specific sampling designs of biological surveys. Crucially, this module generates outputs that mimic the structure of empirical monitoring data, such as sparse spatiotemporal records rather than complete grids. This structural fidelity allows researchers to feed simulated datasets directly into standard analytical pipelines (e.g., SDMs or occupancy models), providing a rigorous platform for benchmarking statistical methods against a known ground truth. A comprehensive overview of the supported biological and observational processes is provided in Table 1.

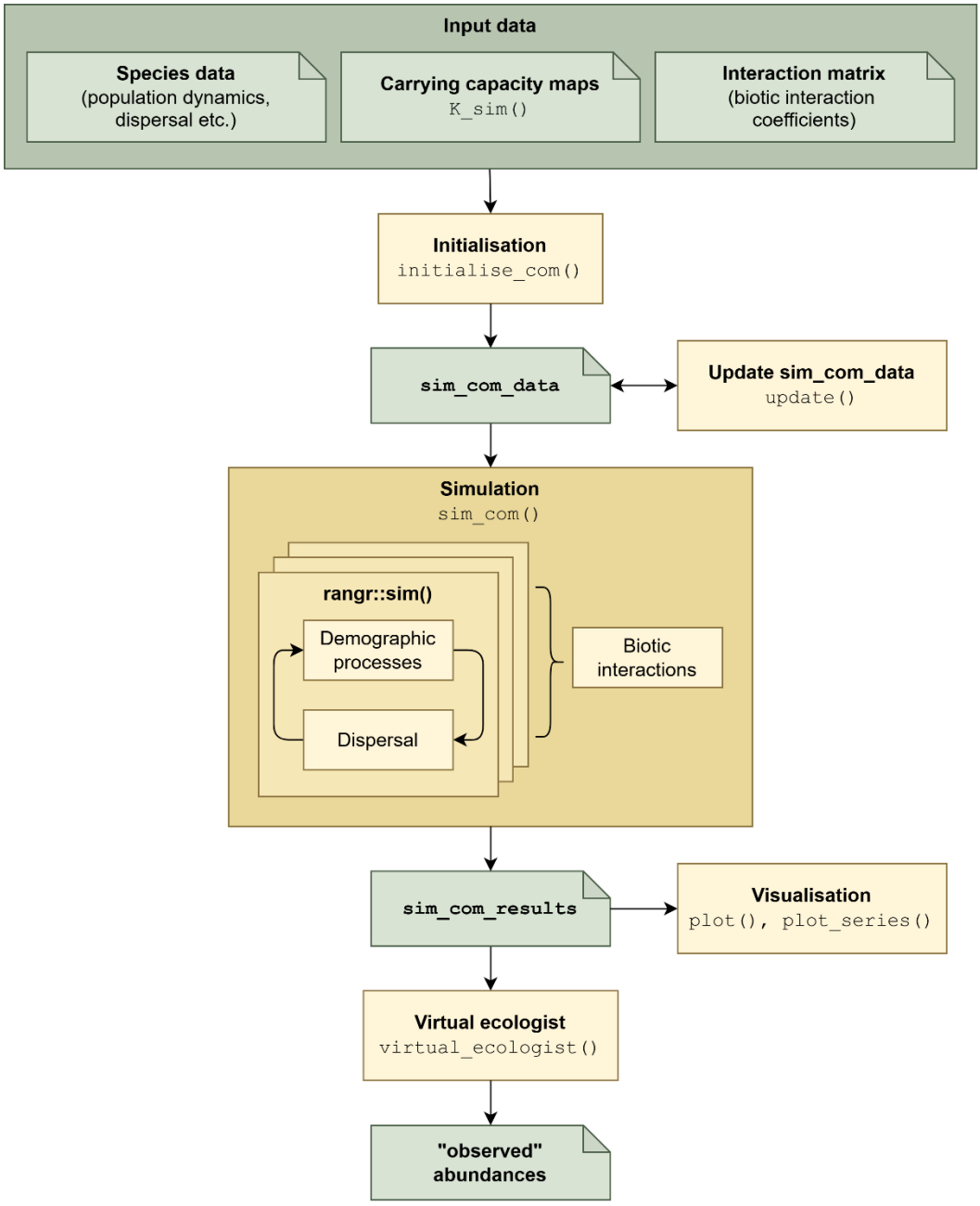

Figure 1. Conceptual framework and operational workflow of the `mrangr` package. The schematic illustrates the mechanistic decoupling of drivers: the fundamental niche is strictly defined by input carrying capacity maps, while the realized niche emerges dynamically from the integration of biotic interactions, demographic rates, and dispersal. The workflow progresses from initialization 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 metapopulation dynamics and sampling.

Table 1. Overview of the `mrangr` framework, distinguishing between the ecological state processes (mechanisms generating the true abundance) and the observation model (mechanisms generating survey data).

Simulation Component	Impact	Implementation in `mrangr`
PROCESS MODEL		
Abiotic constraints (Fundamental niche)	Defines the potential range and maximum abundance of a species based solely on environmental physiology, ignoring other species.	Users supply carrying capacity maps (K), either as static rasters or generated dynamically via `K_sim()` based on environmental variables.
Biotic filtering (Realized niche)	Modifies the fundamental niche by reducing abundance (competition, predation) or expanding it (facilitation), creating the realized distribution.	The interaction matrix (a) defines pairwise coefficients. The simulation solves for abundance at each time step, allowing the realized niche to emerge dynamically from the K maps and matrix a.
Dispersal	Regulates connectivity. 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.
Ecological drift	Stochastic changes in abundance, dominant in small populations.	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		
Observation process	Distorts biological patterns through sampling bias and imperfect detection. Essential for validating analytical methods against "known" truths.	The `virtual_ecologist()` function samples the simulated metacommunity. Users can specify sampling designs (e.g., random, systematic) and detection probability distributions (e.g., `obs_error`) to generate realistic "observed" datasets.

3. Key features of the package

`mrangr` inherits the core population dynamics of `rangr`, including spatially explicit growth models, flexible dispersal kernels, and non-monotonic regulation (e.g., Allee effects). As these fundamental mechanisms are detailed in Markowska *et al.* (2025), we focus here on the novel functionalities emerging from their integration into a multi-species context.

3.1. Interspecific regulation

Central to `mrangr` is a generalized interaction matrix that enables the simulation of diverse community dynamics. By parameterising both positive and negative coefficients in an asymmetric matrix (a), users can represent a full spectrum of ecological interactions, including competition, facilitation, and predation.

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 realized 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.

Formally, the effective carrying capacity for species i at time t in a given cell is calculated as:

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

where S is the total number of species, $N_{j,t-1}$ is the abundance of species j at the previous time step, and the $max(\ldots, 0)$ function ensures that carrying capacity remains non-negative. This formulation represents a specific implementation of the Lotka-Volterra framework where interactions expand or contract the available niche space (K) rather than acting directly on intrinsic growth rates (r).

3.2. Low entry level

The package is designed to minimize technical complexity, requiring only two primary functions to execute a complete simulation. First, a community object is established using the `initialise_com()`, which integrates spatial carrying capacity maps (K), the biotic interaction matrix (a) and species-specific life-history parameters. Subsequently, the `sim_com()` function executes the spatially explicit simulation. This streamlined workflow reduces the programming workload, allowing ecologists to study complex feedback loops and metacommunity dynamics without having to create custom simulation engines.

3.3. 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 enables the mechanistic investigation of invasion success. It allows researchers to explore how community composition, biotic resistance and arrival timing shape the settlement of new species within established metacommunities.

3.4. Virtual ecologist

A major challenge in spatial ecology is that theoretical models often assume perfect knowledge, whereas empirical data is inherently noisy and incomplete. To bridge this gap, `mrangr` includes a 'virtual ecologist' module designed to replicate the constraints of real-world biological surveys (Zurell et al., 2010). While the simulation inherently generates "true" abundances, the `virtual_ecologist()` function allows users to filter this output through imperfect observation methods. The module supports:

- Sampling designs: Users can define the sampling intensity (e.g., surveying only 5% of the landscape) and spatial configuration (random vs. systematic sampling).

- Detection error: It simulates imperfect detection and observation bias by applying error distributions (e.g. Binomial to mimic imperfect detectability or log-Normal to impose observation error on counts) to true abundance data.

Building on recent efforts to make virtual species more realistic (Malinowska et al., 2023), this module generates "observed" datasets alongside known, mechanistically derived ground truths. This dual-output framework allows researchers to rigorously benchmark statistical methods (such as species distribution models or occupancy models) and quantify how sampling limitations affect ecological inference (Meynard et al., 2019).

3.5. Virtual environment generator

To facilitate theoretical investigations, the `K_sim()` function allows for the generation of spatially explicit carrying capacity maps based on spatially autocorrelated Gaussian Random Fields (GRFs). This tool enables users to construct controlled synthetic landscapes by defining both the spatial structure (via the autocorrelation range) and the statistical properties (marginal distributions) of the environment. Furthermore, the function supports the specification of cross-correlations between different landscape layers, allowing researchers to simulate complex niche relationships — such as environmental trade-offs or positive associations — under precise experimental conditions. This offers a versatile framework for testing ecological hypotheses across a range of environmental configurations.

3.6. GIS integration

Unlike theoretical tools that rely solely on synthetic landscapes, `mrangr` is fully integrated with the `terra` ecosystem, the modern standard for spatial data analysis in R (Hijmans, 2026). This interoperability allows users to directly ingest empirical raster data — such as climate layers, land cover maps, or remote sensing outputs — to define simulation arenas. By enabling the use of real-world geographical data as boundary conditions, `mrangr` facilitates

the seamless transition from abstract theoretical exploration to data-driven macroecological modelling.

3.7. Computational efficiency

Spatially explicit simulations are often computationally expensive, particularly when scaling up to large landscapes or high species richness. `mrangr` addresses this by delegating intensive spatial operations to the `terra` package, which is optimized in C++. This allows the package to maintain the flexibility and readability of pure R code while achieving the performance necessary to handle 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 parallelization tools (e.g., `parallel`, `pbapply`), making it feasible to conduct extensive sensitivity analyses and robustly estimate parameter uncertainty.

4. **`mrangr`** workflow

The mrangr package provides a straightforward workflow consisting of 3 main steps.

4.1. Environment and community initialisation

The workflow begins by defining the simulation arena and the community structure. Users can integrate empirical spatial data by providing `SpatRaster` objects for carrying capacity maps (K), representing species fundamental niches, and initial abundance maps (N_1). Alternatively, for theoretical applications or sensitivity testing, the `K_sim()` function allows users to generate synthetic, spatially autocorrelated carrying capacity landscapes. Concurrently, interspecific dynamics are parameterised via an asymmetric interaction matrix (α), enabling the representation of complex biotic relationships. The `initialise_com()` function integrates the spatial data and interaction parameters into a `sim_com_data` object. At this stage, users define species-

specific traits, including intrinsic growth rates (*r*) and dispersal kernels (`kernel_fun`). This step validates the input maps and community parameters before the simulation begins, while also encapsulating all this data into a single `sim_com_data` object.

4.2. Simulation execution

Once the system is defined, the `sim_com()` function executes the spatially explicit simulation over discrete time steps. In each iteration, the model sequentially resolves dispersal and local population dynamics. First, the effective carrying capacity ($K_{i,t}$) of every grid cell is dynamically updated based on the local abundance of all interacting species (as defined by the interaction matrix a). Populations then grow according to their intrinsic growth rates (r), constrained by these dynamically updated realized niches. Simultaneously, individuals disperse across the landscape according to species-specific kernels. This cycle repeats for the specified duration, generating a complete spatiotemporal history of the metacommunity that captures the interplay between environmental forcing, biotic interactions, and dispersal.

4.3. Observation and analysis

Following the simulation, users can analyse the "biological truth" directly or, optionally, pass the results to the `virtual_ecologist()` function. This post-processing step applies the observation constraints described in Section 3.4 to the raw simulation output. By defining specific sampling protocols (e.g., plot number, detection probability) at this stage, users generate an "observed" dataset derived from the "true" state. This dual-output workflow allows researchers to seamlessly benchmark analytical methods by comparing statistical inferences drawn from the virtual samples against the known ground truth of the metacommunity. Following the simulation (and optionally an observation process), the resulting community state can be analysed directly. The package provides native plotting methods: `plot_series()` generates temporal

trajectories of total or mean abundance for all species, while `plot()` visualises the spatial distribution (i.e. realised niches) of the metacommunity at specific time points.

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-colonization 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. Example 1: Testing the effect of dispersal on species diversity

Dispersal is the fundamental process connecting local communities, shaping biodiversity patterns at multiple scales. In metacommunity theory, dispersal promotes local coexistence through the rescue effect, yet potentially undermines regional diversity by homogenizing distinct communities (Mouquet & Loreau, 2003). Consequently, the relationship between dispersal ability and diversity metrics is expected to vary across scales. 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 colonization 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 (α) 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 reorganize 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 are impoverished due to local extinctions and dispersal limitation. Increasing connectivity allows species to colonize and persist in suboptimal patches ('sink' habitats) via the rescue effect, thereby inflating local diversity.

2. Spatial homogenization (β-diversity): Conversely, compositional turnover declined sharply as dispersal increased (Figure 2b). High dispersal rates effectively mix 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 homogenization (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.

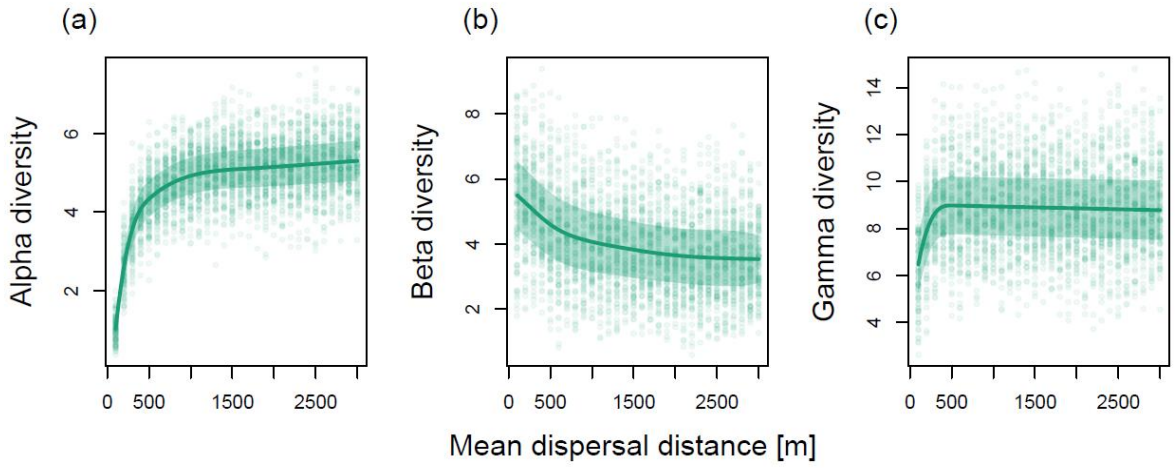

Figure 2. Response of metacommunity diversity components to mean dispersal distance. Scatterplots display (a) alpha, (b) beta, and (c) gamma diversity for metacommunities simulated with a regional pool of 20 species. Points represent individual simulation runs. Solid lines indicate the median and shaded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

5.2. Example 2: Competition–colonization trade-off

A fundamental puzzle in community ecology is explaining how inferior competitors avoid exclusion in landscapes dominated by superior species. The competition–colonization 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 colonization rates, allowing them to occupy vacant patches before slower-dispersing dominants arrive to displace them. Identifying this trade-off in empirical systems is often confounded by environmental heterogeneity and complex trait correlations. In this example, we use `mrangr` to simulate a test of this hypothesis by enforcing a strict constraint between competitive rank and dispersal distance, evaluating whether this trade-off alone is sufficient to maintain regional coexistence in a spatially explicit context.

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. Species 1 (superior competitor): Assigned a baseline carrying capacity generated via a log-normal distribution.
2. Species 2 (inferior competitor): Assigned a carrying capacity 20% lower than Species 1, across the entire landscape.
3. Biotic interactions: We applied strong, symmetric competition between the species ($a = -1$). Under these conditions — identical fundamental niches and distinct fitness levels — 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 per scenario over 50 time steps. We tracked the relative abundance of the inferior competitor to evaluate whether spatial niche partitioning (via colonisation ability) could prevent exclusion despite the lack of niche differentiation.

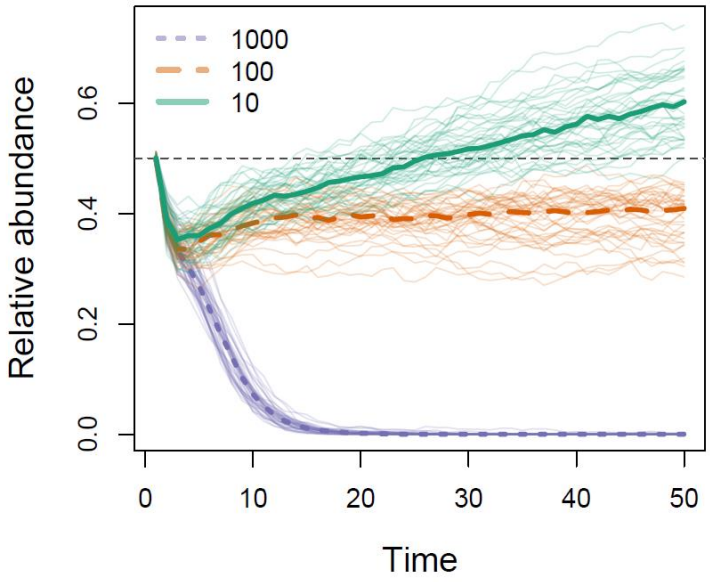

Figure 3. Testing the competition-colonization trade-off. Temporal dynamics of the inferior competitor's relative abundance over 50 simulation steps. The inferior competitor (Species 2) has a fixed high dispersal distance (1000 m) but lower competitive fitness ($K_2 = 0.8 \times K_1$). Thin lines represent 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 dispersal advantage can effectively counteract competitive exclusion (Figure 3). In the absence of a trade-off, when both species shared equal dispersal capabilities (violet dotted line), the inferior competitor was rapidly driven toward extinction. However,

as the trade-off strength increased, the inferior competitor's persistence improved significantly. In the strongest trade-off scenario (green solid line), where the superior competitor was severely dispersal-limited (10 m), the inferior competitor successfully exploited vacant space, achieving numerical dominance despite its lower fitness.

5.3. Example 3: Reconstruction of fundamental niches

Estimating the fundamental niche from field data is complicated by two filters: biotic interactions, which constrain the realized distribution, and observational errors, which distort detection. Consequently, ecological field data rarely reflect pure environmental potential (Soberón, 2007). Yet, 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 penetrate these biological and observational layers to 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 realized 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 initialized with abundances drawn from a Poisson distribution with expectations equal to the local carrying capacity ($\lambda = K$). The system was evolved for 50 time steps, with the first 10 steps serving as a burn-in period to allow the community to reach 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 realized 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 realized abundance from environmental potential (Figure 4c). Concurrently, the absolute bias in estimated environmental coefficients (β) rose (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), equivalent to random guessing. This implies that in highly interactive communities, standard correlative models frequently misidentify positive environmental associations as negative (and vice versa), yielding spurious niche estimates driven by community dynamics rather than abiotic suitability.

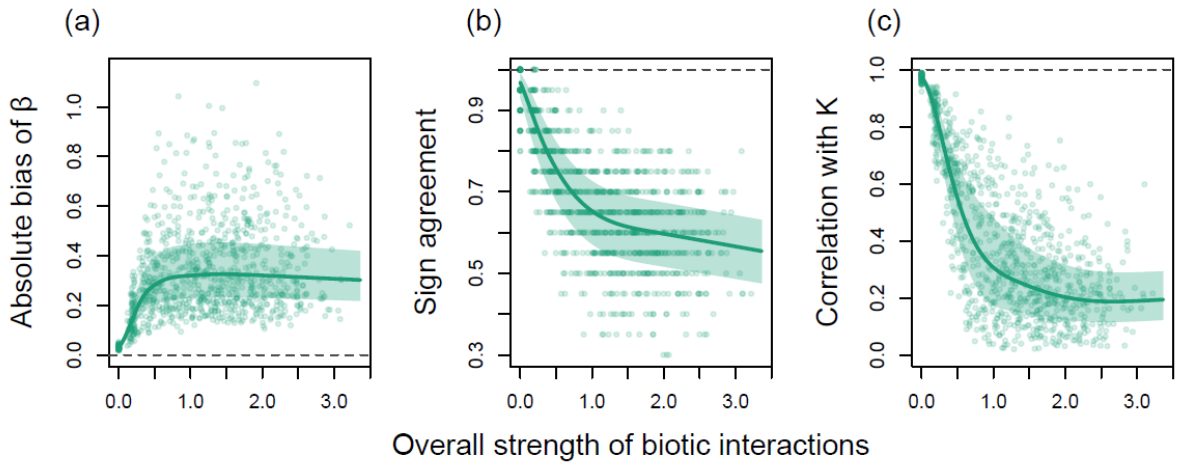

Figure 4. Influence of biotic interaction strength on the accuracy of fundamental niche estimation by the 'virtual ecologist'. Interaction strength is defined as the mean absolute value of off-diagonal elements in the interaction matrix. Estimation performance is evaluated via: (a) absolute bias of slope 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

The metacommunity concept has traditionally been framed within four major paradigms: species sorting, mass effects, patch dynamics, and neutral theory (M. A. Leibold et al., 2004). Recent work has shifted toward a unified process-based and prediction oriented framework (M. A. Leibold et al., 2022), yet operationalization of this synthesis and the ability to link observed patterns to underlying processes remains a key challenge (M. Leibold et al., 2025). Here, we address this issue by reducing metacommunity dynamics into three fundamental axes - space, time, and species - linked through core ecological processes.

Although this abstraction is necessarily simplified, `mrangr` captures both the biotic interactions that drive local species coexistence, and the spatiotemporal population dynamics that determine community assembly, while maintaining mechanistic flexibility. A central feature is the explicit separation of the fundamental and realized niche, ensuring that realised co-distribution patterns dynamically emerge from environmental potential and biotic processes parametrised as independent inputs. By incorporating stochasticity across all processes, `mrangr` enables the generation of parameter distributions from replicated simulations, facilitating robust statistical inference.

This provides a general and user-friendly framework to explore limits of inference, particularly the identifiability of species interactions in metacommunity dynamics (Barbier et al., 2025). Furthermore, being process-based, `mrangr` moves beyond null models, by enabling virtual experiments with arbitrary biotic interaction structures, fully spatially explicit environments, and an integrated observation process module. Therefore, it can serve as a platform for benchmarking of methods that encompass metacommunity dynamics into species co-distribution patterns (Christopher D. Terry et al., 2023; Morueta-Holme et al., 2016; Ovaskainen et al., 2017, 2019). This facilitates precise mechanistic control over the primary processes driving community dynamics, allowing researchers to replicate established patterns while exploring complex frontiers, such as the disentanglement of abiotic filtering from competition, the interplay between niche and fitness differences, or the spatiotemporal dynamics of species invasions.

7. Authors' contributions

KM and LK conceived the ideas and designed the methodology; KM and LK developed the algorithm; KM led the software development and R package implementation; KM, MW, and LK wrote the documentation and vignettes; MW contributed to the validation and testing of the software. All authors contributed equally to the writing of the manuscript and gave final approval for publication. LK acquired the funding and provided overall supervision for the project.

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

Adler, P. B., HilleRisLambers, J., & Levine, J. M. (2007). A niche for neutrality. *Ecology Letters*, *10*(2), 95–104. https://doi.org/10.1111/j.1461-0248.2006.00996.x

Araújo, M. B., & Rozenfeld, A. (2014). The geographic scaling of biotic interactions. *Ecography*, *37*(5), 406–415. https://doi.org/10.1111/j.1600-0587.2013.00643.x

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

Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, *31*(1), 343–366. https://doi.org/10.1146/annurev.ecolsys.31.1.343

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

Dallas, T. (2014). metacom: An R package for the analysis of metacommunity structure. *Ecography*, *37*(4), 402–405. https://doi.org/10.1111/j.1600-0587.2013.00695.x

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

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/

Ke, P.-J., & Letten, A. D. (2018). Coexistence theory and the frequency-dependence of priority effects. *Nature Ecology & Evolution*, *2*(11), 1691–1695. https://doi.org/10.1038/s41559-018-0679-z

Leibold, M. A., Chase, J. M., & Ernest, S. K. M. (2017). Community assembly and the functioning of ecosystems: How metacommunity processes alter ecosystems attributes. *Ecology*, *98*(4), 909–919. https://doi.org/10.1002/ecy.1697

462 Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D.,
463 Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity
464 concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*(7), 601–613.
465 https://doi.org/10.1111/j.1461-0248.2004.00608.x
466 Leibold, M. A., Rudolph, F. J., Blanchet, F. G., De Meester, L., Gravel, D., Hartig, F., Peres-Neto, P.,
467 Shoemaker, L., & Chase, J. M. (2022). The internal structure of metacommunities. *Oikos*,
468 *2022*(1), oik.08618. https://doi.org/10.1111/oik.08618
469 Leibold, M., Barbier, M., Bittleston, L., Clark, A. T., Cuellar-Gempeler, C., D'Andrea, R., Frans, V. F.,
470 Khattar, G., Miller, Z., Peres-Neto, P. R., & Wisnoski, N. I. (2025). *Linking Pattern to Process in*
471 *Metacommunities: Challenges and Opportunities*.
472 https://ecoevorxiv.org/repository/view/9021/
473 Lin, J.-H., Quan, Y.-J., & Han, B.-P. (2024). MetaIBM: A Python-based library for individual-based
474 modelling of eco-evolutionary dynamics in spatial-explicit metacommunities. *Ecological*
475 *Modelling*, *492*, 110730. https://doi.org/10.1016/j.ecolmodel.2024.110730
476 Malinowska, K., Markowska, K., & Kuczyński, L. (2023). Making virtual species less virtual by reverse
477 engineering of spatiotemporal ecological models. *Methods in Ecology and Evolution*, *14*(9),
478 2376–2389. https://doi.org/10.1111/2041-210X.14176
479 Markowska, K., Malinowska, K., & Kuczyński, L. (2025). rangr: An R package for mechanistic, spatially
480 explicit simulation of species range dynamics. *Methods in Ecology and Evolution*, *16*(3), 468–
481 476. https://doi.org/10.1111/2041-210X.14475
482 Meynard, C. N., Leroy, B., & Kaplan, D. M. (2019). Testing methods in species distribution modelling
483 using virtual species: What have we learnt and what are we missing? *Ecography*, *42*(12),
484 2021–2036. https://doi.org/10.1111/ecog.04385
485 Morueta-Holme, N., Blonder, B., Sandel, B., McGill, B. J., Peet, R. K., Ott, J. E., Violle, C., Enquist, B. J.,
486 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

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

Vellend, M. (2016). *The Theory of Ecological Communities*. Princeton University Press. https://doi.org/10.1515/9781400883790

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