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 and ecological plausibility of the package through two proofs of concept: (i) generating synthetic spatiotemporal arrays to quantify the scale-dependent effects of dispersal on α, β, and γ diversity; and (ii) constructing an *in-silico* sandbox to isolate the competition-colonisation trade-off. By providing a complete, end-to-end generative data pipeline, mrangr enables computational ecologists to rigorously benchmark statistical models, optimise sampling geometries, and 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 (Seoane et al., 2023), rigorously validating their predictive accuracy requires a "ground truth". Since empirical datasets lack a known generative mechanism, researchers need synthetic data in which the processes driving population dynamics and biotic interactions are explicitly defined. Metacommunity theory (Leibold et al., 2004; Thompson et al., 2020) provides the ideal mathematical framework for generating these datasets, as it explicitly links density-independent abiotic filtering, density-dependent biotic interactions, and dispersal across spatial scales.

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. To resolve this, we conceptualise these niches strictly through a Hutchinsonian framework, defining the fundamental niche as the abiotic multi-dimensional hypervolume. In `mrangr`, this theoretical construct is explicitly spatialised as georeferenced carrying capacity maps. These baseline maps act strictly as an abiotic boundary condition and can be generated by applying known response curves to environmental rasters. The realised niche is never provided as an input parameter. Instead, it emerges dynamically through spatial simulation as the metacommunity evolves. Furthermore, whereas classic theory posits that the realised niche is strictly a restricted subset of the fundamental niche, our implementation of an asymmetric interaction matrix allows the realised niche to both contract via competition and expand via facilitation.

In ecological modelling, generative frameworks are classified as either phenomenological or mechanistic (Dormann et al., 2012). While phenomenological simulators generate spatial data by

reconstructing observed patterns from sampling distributions derived from statistical models (e.g., Malinowska et al., 2023), `mrangr` is strictly mechanistic. Its input parameters represent explicit biological traits and constraints, such as *per capita* growth rates, interaction coefficients, and dispersal distances. Because the resulting spatiotemporal population arrays are emergent, i.e. calculated iteratively from local demographics and migration balances, `mrangr` serves as an *in silico* experimental platform for computational ecology. Unlike empirical field studies, where parameters such as carrying capacity and fundamental niche are unknown and need to be estimated, the simulator treats these theoretical constructs as exact, user-defined inputs. This provides a known "ground truth", which fulfils two key analytical purposes. Firstly, it provides a controlled environment to validate and benchmark statistical models, allowing scientists to test the effectiveness of algorithms in reconstructing true parameters from noisy data. Secondly, it acts as a theoretical laboratory to simulate complex spatiotemporal dynamics, helping researchers to identify plausible ecological mechanisms before initiating logistically demanding empirical surveys. 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 a strict separation between fundamental and realised niches. 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 a pre-configured, population-level computational framework 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. Building upon the single-species metapopulation framework developed in the `rangr` package (Markowska et al., 2025), `mrangr` elevates simulations

to the metacommunity scale. Just as individual metapopulations interact to form complex communities in nature, `mrangr` dynamically couples multiple virtual populations through an asymmetric, generalised Lotka-Volterra (GLV) interaction matrix within the `terra` spatial ecosystem. This architecture leverages vectorised matrix operations across environmental rasters, enabling computationally efficient simulations where complex spatial co-occurrence patterns and community structures arise as emergent properties of explicitly parametrised dispersal, demography, and interspecific interactions. Crucially, the package couples this generative engine with a "Virtual ecologist" module, enabling researchers to apply realistic observation models to the simulated ground truth. This dual architecture – generating mechanistically pure data and simulating targeted sampling biases – provides a comprehensive toolkit for generating synthetic spatial datasets to benchmark ecological algorithms and assess the robustness of macroecological inferences. Finally, to validate the simulator, we present two proofs of concept that replicate well-documented metacommunity phenomena, demonstrating the computational reliability and ecological plausibility of the generative engine.

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 directly 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. These carrying capacity grids can be provided as static maps or as

multitemporal raster stacks, allowing the abiotic baseline to shift during the simulation to model global change scenarios. The generative engine then iterates across discrete time steps, computing the emergent realised metacommunity through the integration of:

1. Demography and interaction: Within each raster cell, local community abundances 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. Spatial connectivity: Following local demographic updates, species-specific dispersal kernels are executed across the simulation grid, linking isolated cells into a metacommunity.
3. Observation error: The "Virtual ecologist" module applies probabilistic error functions and spatial masking. This mimics the specific sampling geometries and observation biases of real-world biological surveys.

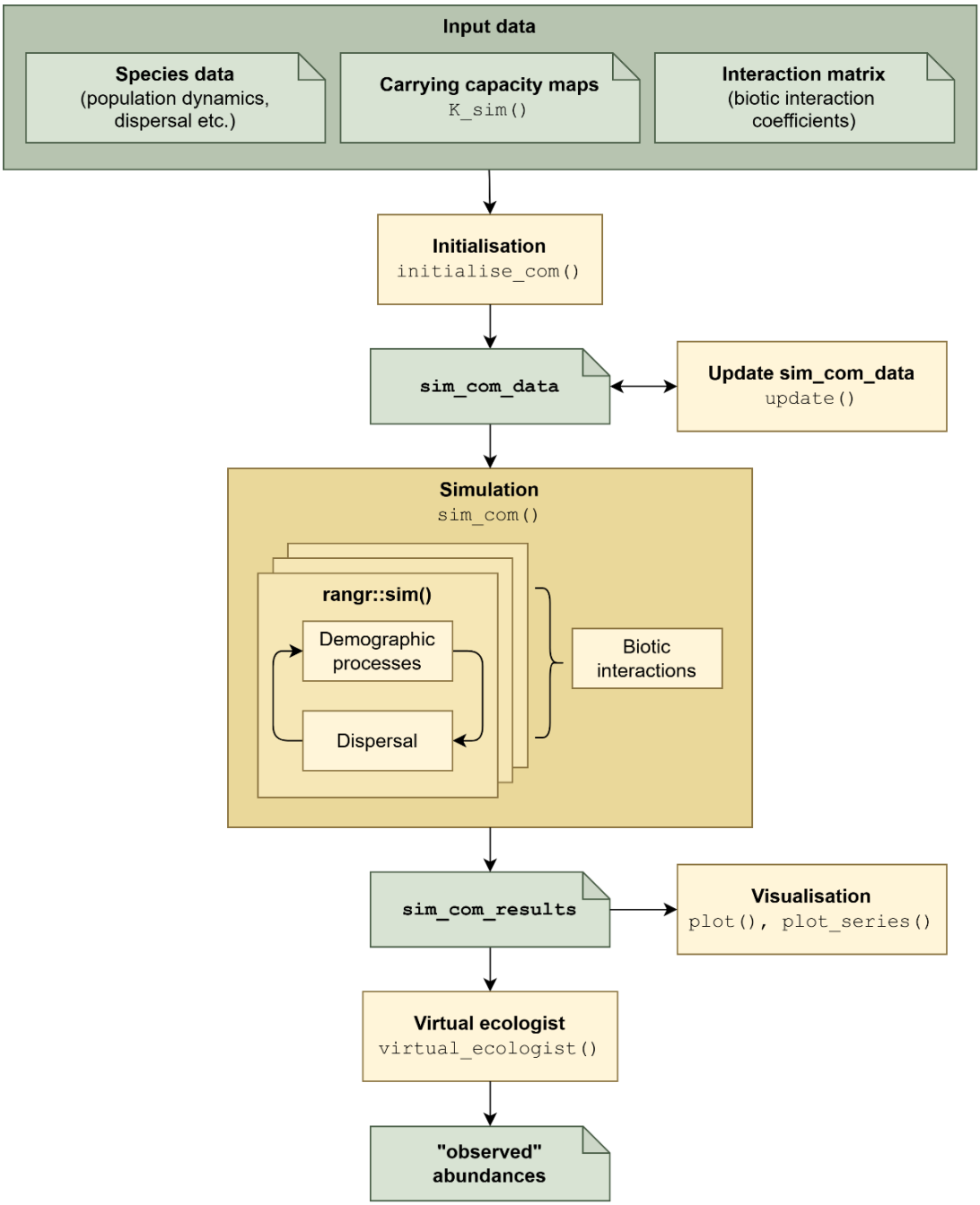

Figure 1. Conceptual framework and operational workflow of the `mrangr` package. The schematic illustrates the mechanistic decoupling of drivers: the fundamental niche is strictly defined by input carrying capacity maps, while the realised niche emerges dynamically from the integration of biotic interactions, 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 frameworks (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 Table 1.

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 simulate multispecies networks. The following features highlight the package utility as both a generative 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 local community.

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

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

where S is the total number of interacting species, $K_{fund,i}$ is the abiotic baseline carrying capacity, representing the fundamental niche of species i, and $N_{j,t-1}$ is the local abundance of interacting

156 species from the previous time step. The $max(\ldots, 0)$ function ensures that carrying capacity remains
157 non-negative. Unlike classical applications of the GLV, which modify intrinsic growth rates, our
158 implementation specifically scales the local carrying capacity based on the interaction matrix. This
159 parameterisation mathematically isolates the fundamental niche (the baseline capacity determined
160 purely by abiotic constraints) from the realised niche (the effective capacity modulated by biotic
161 interactions).

162 ### 3.2. Simulation workflow

163 The computational workflow relies on two primary functions: `initialise_com()` prepares the
164 simulation space by linking spatial maps with demographic parameters, and `sim_com()` executes the
165 spatial iterations over time. This streamlined architecture allows users to loop over parameter grids to
166 conduct sensitivity analyses or generate synthetic datasets. Crucially, this programmatic control over
167 all input variables allows users to design rigorous *in silico* experiments. Researchers can establish
168 baseline null models (e.g., communities assembled purely by environmental filtering with no biotic
169 interactions, or panmictic populations with no dispersal limitation) against which specific theoretical
170 treatments can be systematically compared.

171 ### 3.3. Simulating invasion dynamics

172 The package offers functionality to simulate species invasions. Users can designate specific species as
173 invaders and schedule their introduction at defined time steps, rather than initialising them at the start
174 of the simulation. This temporal flexibility allows for the simulation of biological invasions, priority
175 effects, and community assembly dynamics.

176 ### 3.4. The "Virtual ecologist" module

177 To bridge the gap between theoretical simulations and empirical data, the package includes a
178 `virtual_ecologist()` module (Zurell et al., 2010). While a large body of literature in empirical
179 ecology is dedicated to estimating imperfect detection and abundance errors using hierarchical models,

`mrangr` provides the forward-simulation equivalent of this process. The module mimics real-world biodiversity surveys by applying a hierarchical observation layer over the true ecological state. Users can simulate specific spatial sampling designs (e.g., surveying a specified percentage of the landscape) and introduce probabilistic observation errors (e.g., using binomial distribution for incomplete detection or log-normal distribution for abundance estimation errors). By generating these degraded, "observed" datasets alongside the known, mechanistically derived ground truth (Malinowska et al., 2023), this dual-output framework allows researchers to benchmark statistical methods and quantify how sampling biases affect ecological inference (Barbosa & Alves-Souza, 2025; 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 (GRFs) to create spatially autocorrelated landscapes. This functionality allows users to generate complex, cross-correlated environmental gradients.

3.6. GIS integration

The `mrangr` package is built directly on the `terra` spatial ecosystem (Hijmans, 2026). Because `terra` interfaces with foundational geospatial C++ libraries (GDAL, GEOS, and PROJ), `mrangr` natively reads and writes all standard GIS formats (e.g., GeoTIFFs, ESRI grids, shapefiles) and handles geographic coordinate reference systems. Consequently, users can directly import empirical spatial data from desktop GIS software (e.g., QGIS or ArcGIS) and export simulation raster stacks for external spatial analysis.

Figure 2 illustrates the standard spatial outputs generated by the simulation engine. As the software handles multidimensional spatial arrays natively, researchers can track complex demographic shifts, such as local competitive exclusion, at the pixel level across discrete time steps. As shown in the figure, the initial overlap of species distributions (*t = 1*) can be projected forward to reveal the emergence of landscape-scale spatial segregation (*t = 100*) directly within the simulated GIS environment.

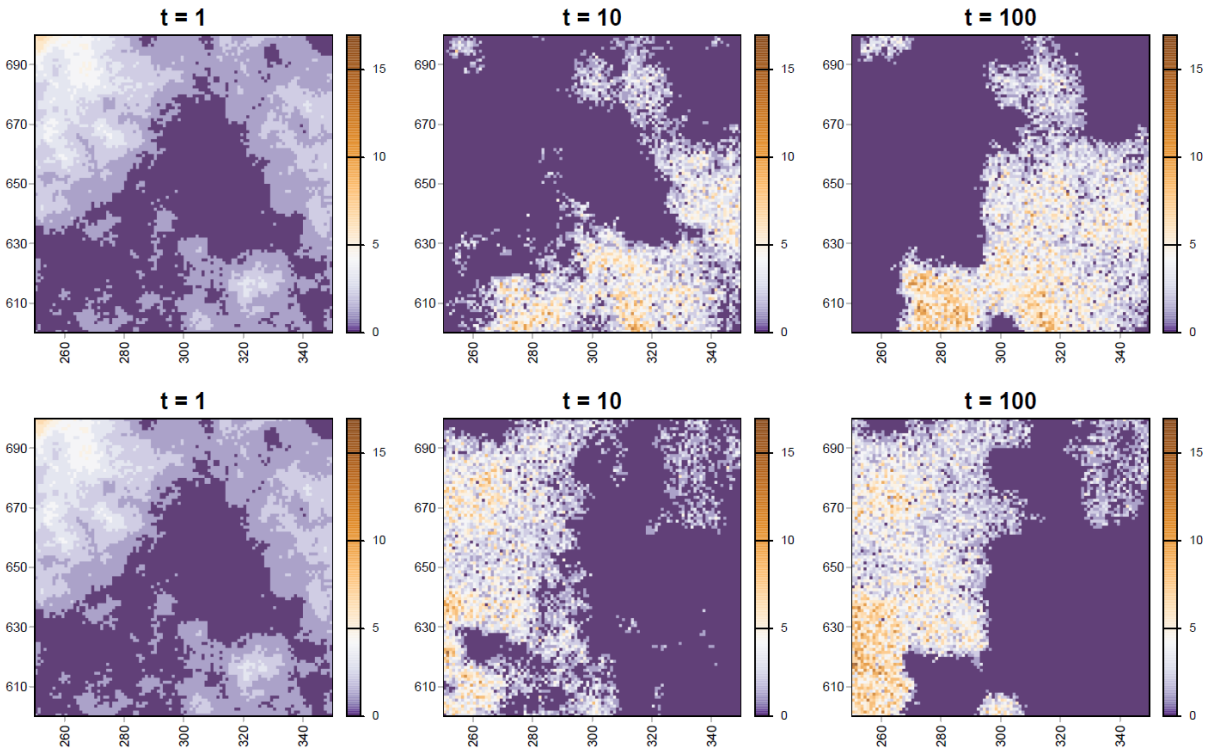

Figure 2. Spatiotemporal simulation of two competing species resulting in competitive exclusion. The sequence of maps illustrates the simulated abundance of a two-species community across three time steps (*t = 1, 10,* and *100*), with Species 1 and 2 shown in the top and bottom rows, respectively. Although both species have distinct and independent fundamental niches, they were initialised with identical spatial distributions. To simulate strong competition, both interspecific interaction coefficients (a_{12} and a_{21}) were set to *-0.8*. Consequently, the initially identical spatial distributions (*t = 1*) gradually diverge (*t = 10*) as a result of local competitive exclusion. By the final time step (*t = 100*), the spatial distribution of the two species are distinctly allopatric, demonstrating that each species has displaced its competitor from areas where it holds an environmental advantage. Axes represent spatial coordinates in kilometres (EPSG:2180 projection), and colour gradients indicate local population abundance.

3.7. Computational efficiency

Spatially explicit simulations can be computationally demanding, particularly when scaled to high-resolution landscapes or highly diverse communities. `mrangr` addresses this issue by delegating

spatial operations to the `terra` package, which is optimised using the `C++` programming language. This enables the processing of large landscape grids and facilitates replication within the `R` environment. Furthermore, the package supports parallel execution. Users can distribute simulation replicates across multiple processor cores using standard `R` parallelisation tools (e.g., `parallel`, `pbapply`) to conduct sensitivity analyses and quantify parameter uncertainty. This parallel architecture ensures that the simulation engine can be scaled from standard workstations to high-performance computing (HPC) clusters.

Consequently, the primary operational constraints of the package are strictly hardware-bound, dictated by the Random Access Memory (RAM) available to hold multidimensional raster arrays and by the number of logical cores allocated for distributed computing.

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 simulation object. This function binds the spatial rasters, demographic parameters, and interaction

matrix into a unified data structure. At this stage, the user also defines the dispersal kernel (the probability distribution of migration distances) and the starting spatial 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 translates the mechanistically derived "true" abundances into realistic, noisy survey data. Using the `virtual_ecologist()` function, users can apply spatial sampling designs (e.g., random or systematic plots) and observation models (e.g., imperfect detection or counting errors). This transforms the spatial grids into standard ecological data formats, such as presence/absence matrices or count data frames, ready for analysis by statistical tools like Species Distribution Models (SDMs) or Joint Species Distribution Models (JSDMs). By controlling sampling intensity and error rates, 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. Proofs of concept

We present two proofs of concept to validate the simulator against established ecological theory. These examples demonstrate capacity of the package to replicate known macroecological phenomena: specifically, the influence of dispersal on biodiversity scaling, and the dynamics of competition-colonisation trade-offs. Each proof of concept follows a standardised structure. First, we outline the theoretical expectations based on existing literature. Second, we detail the specific input parameters and computational implementation. Finally, we evaluate the resulting simulated trajectories. To ensure broad accessibility, the accompanying R scripts are configured for small spatial arenas, allowing users to execute the simulations independently, generate explicit raster maps, and verify the package's mechanics on standard personal computers.

Both proofs of concept were run in the same exemplary simulation environment, defined on a 20×20 grid (400 cells) with a 1 km^2 spatial resolution, utilising the EPSG:2180 coordinate system. Trends in the simulated parameters were quantified and visualised using Generalised Additive Models for Location, Scale and Shape (GAMLSS) to capture non-linear responses and heteroscedasticity.

As `mrangr` incorporates stochasticity into its demographic and dispersal mechanics, each individual simulation run generates a different trajectory. To illustrate this variability, our results show the empirical median and interquartile range (IQR) of the replicated simulations. Importantly, these bands represent the raw stochastic dispersion of the generative algorithm rather than the confidence intervals of the estimated unknown parameter.

5.1. Simulating the scale-dependent effects of dispersal on biodiversity

Dispersal is the fundamental process connecting local communities and shaping biodiversity patterns at multiple scales. According to metacommunity theory, varying dispersal rates create well-documented spatial pattern: dispersal promotes local coexistence through the rescue effect, yet potentially undermines regional diversity by homogenising 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 using the built-in `K_sim()` function. 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 (d_{mean}) from 100 m to 3000 m across 30 discrete intervals. Dispersal was modelled using an exponential kernel, where the rate parameter is defined as $1/\mathrm{d}_{mean}$.

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 3a). 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.
2. Spatial homogenization (β-diversity): Conversely, compositional turnover declined sharply as dispersal increased (Figure 3b). 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 3c). 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.

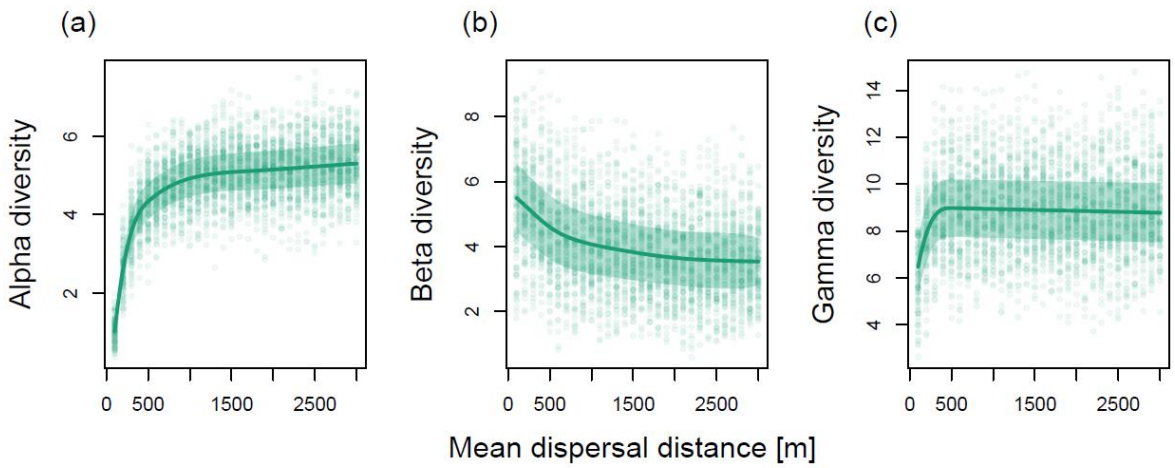

Figure 3. Response of metacommunity diversity components to mean dispersal distance. Scatterplots display (a) alpha, (b) beta, and (c) gamma diversity for metacommunities simulated with a regional pool of 20 species. Points represent individual simulation runs. Solid lines indicate the median and shaded regions represent the interquartile range, modelled using a Gaussian Location-Scale GAM (GAMLSS).

5.2. 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 dispersal capabilities, allowing them to colonise 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 (K_1) generated via a spatially autocorrelated log-normal distribution.
2. The inferior competitor (Species 2): Assigned a proportionally lower carrying capacity ($K_2 = 0.8K_1$) across the entire landscape.
3. Biotic interactions: We enforced strong, symmetric competition via the interaction matrix, setting the off-diagonal interspecific coefficients to ($a_{ij} = a_{ji} = -1$).

Under these conditions (identical fundamental niches and distinct fitness levels) the theory predicts the exclusion of Species 2. To test the trade-off, we fixed the mean dispersal distance of the inferior competitor at 1000 m and systematically varied the mean dispersal of the superior competitor 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× dispersal disadvantage).
3. Strong trade-off: Species 1 disperses 10 m (100× dispersal 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.

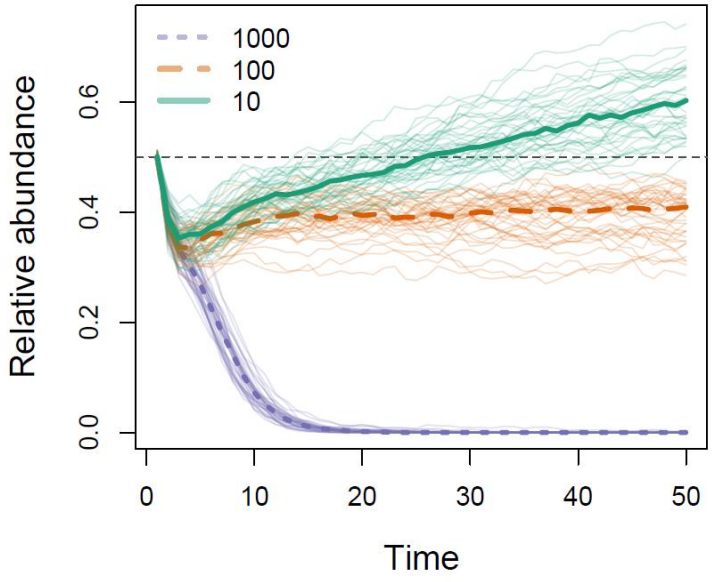

Figure 4. Testing the competition-colonization trade-off. Temporal dynamics of the inferior competitor's relative abundance over 50 simulation steps. The inferior competitor (Species 2) has a fixed high dispersal distance (1000 m) but lower competitive fitness ($K_2 = 0.8K_1$). Thin lines represent individual 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 confirm that a dispersal advantage can effectively counteract competitive exclusion (Figure 4). In the control scenario, when both species shared equal dispersal capabilities, the inferior competitor was rapidly driven towards extinction. However, as the strength of the trade-off increased, the persistence of the inferior competitor improved significantly. In the strongest trade-off scenario, in which the superior competitor was severely dispersal-limited, the inferior competitor successfully exploited vacant space, maintaining long-term coexistence and achieving numerical dominance despite its lower fitness.

6. Scope and limitations

Although `mrangr` offers a highly flexible architecture for spatial simulation, it is essential to define its limits. As a population-level metacommunity simulator, the package relies on abstraction to ensure computational tractability over large spatial extents. Consequently, individual-level behavioural rules and fine-scale physiological processes are beyond its intended scope. Researchers investigating these highly granular mechanisms would be better served by highly parameterised agent-based models (ABMs). However, for macroecological applications where the goal is to benchmark algorithms or investigate broad, landscape-level assembly rules, this level of abstraction is not a limitation, but a fundamental mathematical requirement.

Furthermore, relying on a generalised Lotka-Volterra (GLV) engine introduces specific biological assumptions. Currently, the package does not support demographic stage structure (e.g., age classes), instead assuming that all individuals within a local population are demographically equivalent. Similarly, the simulation strictly focuses on ecological timeframes. Population growth rates, dispersal kernels and interaction coefficients remain fixed, meaning the package does not model eco-evolutionary dynamics such as mutation, adaptation, or speciation. Users must also be aware of the inherent mathematical constraints of GLV equations: randomly assembled, highly connected interaction matrices can become mathematically unstable for very large species pools, requiring careful parameterisation (Allesina & Tang, 2012; May, 1972).

Finally, it is important to emphasise that `mrangr` primarily operates as a forward simulator, generating synthetic data from explicitly defined mechanistic processes. It is not an inverse modelling tool designed to fit empirical datasets. While the generative engine could, in principle, be coupled with Approximate Bayesian Computation (ABC) or similar inverse frameworks to estimate parameters for specific real-world systems, its out-of-the-box utility is calibrated for methodological benchmarking and theoretical hypothesis generation.

7. Conclusions

To address the challenge of accurately linking observed spatiotemporal patterns back to their underlying generative processes (Leibold et al., 2025), `mrangr` simplifies complex metacommunity dynamics into a computationally tractable framework driven by three axes: space, time, and species. By explicitly separating the fundamental and realised niches and decoupling ecological states from observation models, the software moves computational macroecology beyond traditional static null models.

Ultimately, this framework provides the infrastructure to explore the limits of ecological inference, particularly regarding the identifiability of biotic interactions from observational data (Barbier et al., 2025). As macroecologists increasingly rely on complex correlative approaches, such as Joint Species Distribution Models (Morueta-Holme et al., 2016; Ovaskainen et al., 2017; Terry et al., 2023), it is necessary to establish rigorous baselines to quantify parameter bias and algorithmic error margins. Through strict mechanistic control, `mrangr` serves as a benchmarking platform for testing these sophisticated analytical methods, enabling researchers to explore 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.

As an open-source package within the R ecosystem, `mrangr` is designed to be highly modular and extensible. Future developments may build on this foundation to incorporate eco-evolutionary dynamics or stage-structured demography. By providing a mechanistically derived ground truth, `mrangr` bridges the gap between theoretical metacommunity ecology and modern ecoinformatics, offering a computational sandbox to validate the statistical tools that will drive the future of biodiversity forecasting.

8. 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;

9. 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).

10. Conflict of interest statement

The authors declare no conflicts of interest.

11. 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.18641950).
- 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.18643289).
- The preprint of this manuscript is available at: https://ecoevorxiv.org/repository/view/11808/.

12. References

Allesina, S., & Tang, S. (2012). Stability criteria for complex ecosystems. Nature, 483(7388), 205–208. https://doi.org/10.1038/nature10832

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

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

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. Journal of Biogeography, 39(12), 2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.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

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/

Leibold, M. A., 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 with Process in (Meta)communities: Challenges and Opportunities. EcoEvoRxiv. https://doi.org/10.32942/X2504P

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

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

May, R. M. (1972). Will a Large Complex System be Stable? Nature, 238(5364), 413–414. https://doi.org/10.1038/238413a0

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., 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

Terry, J. C. D., 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

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

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