**METHOD**

# MadingleyR: An R package for mechanistic ecosystem modelling

Selwyn Hoeks[1] [iD] | Marlee A. Tucker[1] | Mark A. J. Huijbregts[1] | Mike B. J. Harfoot[2] [iD] | Mike Bithell[3] | Luca Santini[1,4,5] [iD]

[1]Department of Environmental Science, Institute for Water and Wetland Research, Radboud University, Nijmegen, The Netherlands

[2]United Nations Environment World Conservation Monitoring Centre, Cambridge, UK

[3]Department of Geography, University of Cambridge, Cambridge, UK

[4]Department of Biology and Biotechnologies "Charles Darwin", Sapienza University of Rome, Rome, Italy

[5]National Research Council, Institute of Research on Terrestrial Ecosystems (CNR-IRET), Monterotondo, Rome, Italy

**Correspondence**
Selwyn Hoeks, Department of Environmental Science, Institute for Water and Wetland Research, Radboud University, Nijmegen, The Netherlands.
Email: s.hoeks@science.ru.nl

**Funding information**
ERC grant CoG SIZE, Grant/Award Number: 647224; Veni Grant, Grant/Award Number: 0.16.Veni.181.031

**Editor:** Amanda Bates

## Abstract

**Aim:** Mechanistic general ecosystem models are used to explore fundamental ecological dynamics and to assess possible consequences of anthropogenic and natural disturbances on ecosystems. The Madingley model is a mechanistic general ecosystem model (GEM) that simulates a coherent global ecosystem, consisting of photo-autotrophic and heterotrophic life, based on fundamental ecological processes. The C++ implementation of the Madingley model delivers fast computational performance, but it (a) limits the userbase to researchers that are familiar with the intricacies of C++ programming, (b) has limited possibility to change model settings and provide model outputs required to address specific research questions, and (c) has limited reproducibility of simulation experiments. The aim of this paper is to present an R package of the Madingley model to aid with increasing the accessibility and flexibility of the model.

**Innovation:** The MadingleyR R package streamlines the installation procedure and supports all major operating systems. MadingleyR enables users to combine multiple consecutive simulations, making case study specific modifications to MadingleyR objects along the way. Default input files are available from the package and study-specific inputs can be easily loaded from the R environment. MadingleyR also provides functions to plot and summarize MadingleyR outputs. We provide a comprehensive description of the MadingleyR functions and workflow. We also demonstrate the applicability of the MadingleyR package using three case studies: (a) simulating the cascading effects of the loss of mega-herbivores on food-web structure, (b) simulating the impacts of increased land-use intensity on the total biomass of different feeding guilds by restricting the total vegetation biomass available for feeding and (c) simulating the impacts of an intensive land-use scenario on a continental scale.

**Main conclusions:** The MadingleyR package provides direct accessibility to simulations with the mechanistic ecosystem model Madingley and is flexible in its application without a loss in performance.

**KEYWORDS**
accessibility, C++, general ecosystem model, Madingley model, process-based modelling, R package, reproducibility

## 1 | INTRODUCTION

The field of ecology focusing on patterns and processes over large temporal and spatial scales (broadly termed 'macroecology'; McGill, 2019) has traditionally embraced a strong observational and correlative approach (Cabral et al., 2017; Connolly et al., 2017). Although proven to be crucial for the development of macroecological research, thorough mechanistic understanding of empirical findings and correlative model outcomes is frequently lacking (Beck et al., 2012; McGill & Nekola, 2010). By modelling ecosystems through their core processes, mechanistic models may provide insights on causal processes (Connolly et al., 2017; Fitzpatrick & Hargrove, 2009; Wüest et al., 2020). As such, mechanistic modelling approaches can be considered a complementary and promising approach in the field of macroecology (Grimm et al., 2017). Methods to reduce ecosystem complexity and simultaneous improvements in computing power have increased the practical applicability of such approaches on a macroecological scale (Connolly et al., 2017; Grimm et al., 2017; Harfoot et al., 2014).

The Madingley model is one of the few mechanistic general ecosystem models (GEMs) that can simulate a coherent global ecosystem consisting of both photo-autotrophic and heterotrophic life (Harfoot et al., 2014; Purves et al., 2013). The biology and behaviour of the organisms are included in the model, allowing the fate of organisms as well as the interactions between them and the environment to be modelled. Additionally, patterns of ecosystem structure and function can be explored. Other mechanistic ecosystem models exist (e.g. Fitz et al., 1996; Krinner et al., 2005; White et al., 2000); however, the aim and purpose of these models vary, making them structurally different and applicable in different contexts. An advantage of the Madingley model is its flexibility and capacity to simulate and compare different regions across the world as well as run global-scale simulations. In doing so, the Madingley model is not directly constrained by location specific processes and/or a particular group of organisms within the ecosystem. Naturally, the generalization of ecosystems requires simplifications, making the Madingley model less suitable to address local context-, or taxa-, specific questions. Other terrestrial process-based ecosystem models are more suited for this purpose, like the SAVANNA model (Coughenour, 1993) or the EcoSim model (Lorek & Sonnenschein, 1999), which are tailored to specific locations, biomes and species. Additionally, dynamic global vegetation models (DGVMs) can be used to simulate plant communities globally and allow for the study of the interaction between vegetation and climate with greater detail than possible in the Madingley (Krinner et al., 2005; Kucharik et al., 2000), but do not simulate processes related to heterotrophs as done in the Madingley model.

Since the Madingley model was introduced, it has been used to examine human influence on ecosystem dynamics, such as the effect of land-use change on ecological patterns (e.g. Bartlett et al., 2016), the impact of bushmeat harvesting on ecosystems (Barychka et al., 2019, 2020), and the impact of megafauna loss for ecosystem functioning (Enquist et al., 2020; Hoeks et al., 2020). The Madingley model has also been used to gain deeper insights into the mechanisms underlying macroecological patterns such as the Bergmann's rule (Harfoot et al., 2020). These studies illustrate how the Madingley model can be used to make meaningful estimations as well as provide insights into important ecological mechanisms.

While the Madingley model reduces the number of ecological processes and interactions by categorizing life into functional groups and organizing similar organisms into cohorts, it is still a computationally intensive model. The first version of the Madingley model (Harfoot et al., 2014) was written in the C# programming language and later translated into C++, enabling model compatibility with Linux and Mac operating systems and partially retaining its compatibility with Windows machines. However, interpreting the C++ code, learning how to correctly setup a model run, compiling the C++ source code into an executable, and exporting and processing the desired model outputs is a highly challenging and time-intensive task. These steps require certain technical skills that may limit the number of potential users, as well as reducing the reproducibility of studies relying on the Madingley model. Especially considering the increasing popularity of the R programming language (R Core Team, 2020) amongst ecologists and the reality that most users are not familiar with the underlying C++ code of the R functions (Lai et al., 2019). This barrier makes the current version limited in its ability to change model settings to answer specific research questions.

Here, we introduce the MadingleyR R package to increase the flexibility and accessibility of the Madingley model. The current version of MadingleyR is only able to run simulations for the terrestrial realm. MadingleyR is set up as a transparent (open source), flexible and user-friendly R package. To be compatible with future Madingley model updates and achieve high model performance, the core code relies on the parallel (OpenMP) C++ version. This makes it as fast as the default C++ model and, after implementing minor modifications, is now also natively compatible with all operating systems. The integration between the C++ and R code allows input parameters, data files and output results to be easily transferred. The reduced complexity of setting up a simulation and visualizing the outputs also increases the transparency and reproducibility of future studies involving the MadingleyR package within the ecological community. The MadingleyR package lowers the number of hard-coded parameters, allowing the user to have control over the individual processes within the model. Finally, MadingleyR adds new output options to the model and provides functions for output visualization.

Here, we discuss the structure of the Madingley model, the implementation of the package and illustrate its capabilities using three example simulations: (a) simulating the cascading effects of large (>100 kg) herbivore loss on food-web structure, (b) simulating the impacts of increasing land-use intensity on the total biomass of different feeding guilds, and (c) simulating the impacts of an intensive land-use scenario on a continental scale.

## 2 | MADINGLEY

The Madingley model simplifies ecosystems by categorizing organisms into distinctive classes with unique functional roles (Harfoot et al., 2014). Furthermore, it aggregates similar organisms into cohorts, which act as a single entity, thereby reducing the computational requirements. Each cohort represents a collection of spatially co-existing organisms with identical categorical and quantitative functional traits, which results in a set of unique behavioural qualities. Heterotroph cohorts are divided into functional groups by their trophic level, metabolic characteristics and reproductive strategy. Additionally, body mass characterizes most cohorts' behaviours, such as feeding preference, growth rate, dispersal capabilities, reproduction and metabolic needs. The model simulates cohorts with body masses from 10 mg to 150,000 kg, including ectotherms and endotherms. The abundance is tracked to identify how many individuals are alive within each cohort. All cohorts have to undergo a number of predefined processes every time step. These include: (a) energy loss due metabolic processes, (b) feeding on autotrophs (herbivory) and/or on other cohorts (carnivory), (c) reproducing if certain conditions are met, (d) probability of dying from the background mortality, starvation or senescence, and finally (e) probability of dispersing to a neighbouring grid cell (see Figure 1). The outcome of these processes determines if a cohort grows in body mass (develop from juvenile to adult), if the abundance of the cohort is reduced (from mortality or predation) and if an adult cohort reproduces. The production, allocation and non-herbivory mortality dynamics of autotrophs are modelled as a function of climate using a terrestrial carbon model (Smith et al., 2013).

## 3 | MadingleyR PACKAGE

### 3.1 | Model updates

The C++ source used within MadingleyR is a translation of the C# version of the original model published by Harfoot et al. (2014) with updates made primarily to the flexibility of reading inputs and writing outputs. For example, the C# version and C++ translation of the Madingley model require a set of spatial inputs (stored as netCDF files) defining the environmental characteristics as well as separate spreadsheets (stored as csv files) defining respectively the ecological parameters, model setup details and scenario settings. To increase the integration of the Madingley model with the MadingleyR package, the C++ code was modified to accept and return outputs from and to R in raw text form, thereby removing the netCDF dependency in the Madingley model and allowing the model to be compiled natively on Windows. Compiling the netCDF C++ API on Windows is not officially supported and can be challenging; therefore the use of emulation (e.g. Cygwin) or virtualization would be the preferred options. However, these approaches reduce the performance of the C++ source code. Table 1 provides a complete overview of the alterations made. We note that the C++ source of the MadingleyR package only includes the ecology for the terrestrial realm. This is because the terrestrial component of the model is further developed and more extensively tested (e.g. Bartlett et al., 2016; Barychka et al., 2020; Enquist et al., 2020; Harfoot et al., 2020; Hoeks et al., 2020; Newbold et al., 2020) compared with the marine component. The Madingley model does not include any interactions between the marine and terrestrial realms, making terrestrial
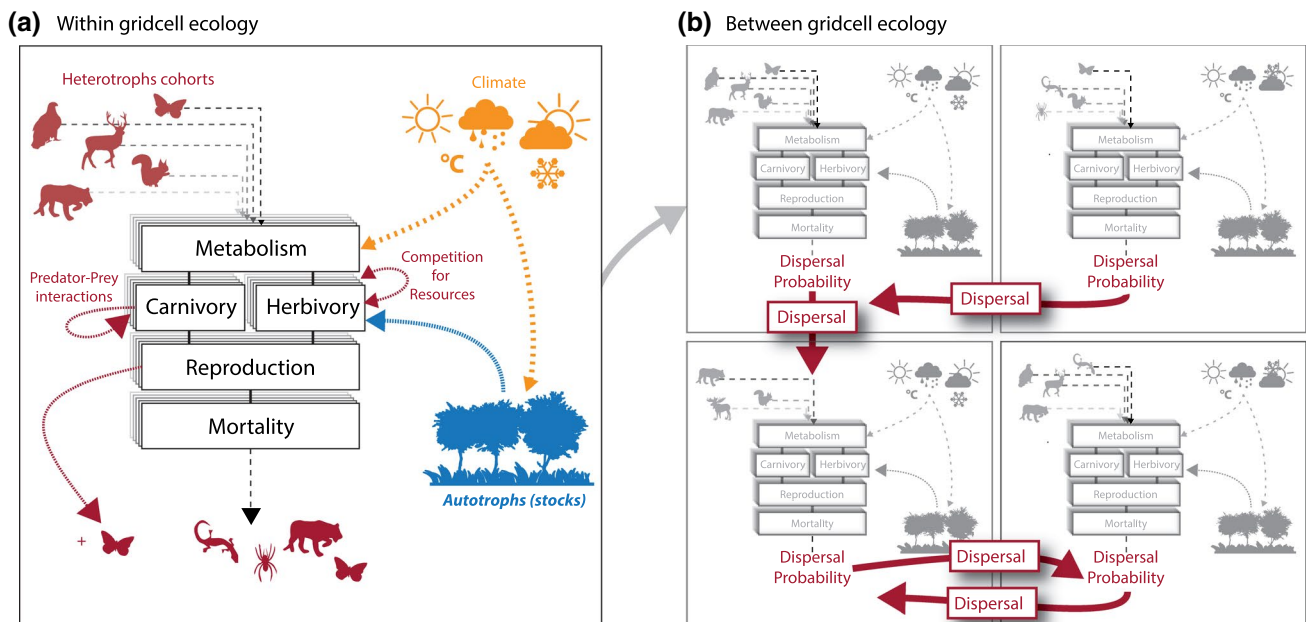


**FIGURE 1** Overview of the within grid cell (a) and between grid cell (b) ecological processes of the Madingley model. Animal silhouettes are used to depict the variety in cohort traits, but note that the Madingley model does not model specific species. Instead, cohorts are defined by a set of trait values, which makes one cohort functionally different from the other [Colour figure can be viewed at wileyonlinelibrary.com]

simulations completely independent of processes occurring within the marine realm.

## 3.2 | MadingelyR interface

The MadingleyR package acts as an interface to a pre-compiled C++ version of the Madingley model. The raw text files, used for communication between the C++ and R code, are created, compressed and stored dynamically in the temporary folder of R in order to run simulation experiments directly from R. Table 2 provides an overview of the MadingleyR features, comparing it to the pure C++ and C# versions of the model.

The MadingleyR package was built from three major components. The first component concerns the main model written in C++. The C++ code utilizes the OpenMP library for parallelization, which decreases the time required for a single simulation. The underlying C++ code was pre-compiled into three executables, specific to the targeted operating system (Linux, Mac and Windows). The executables, alongside the required dynamic-link library files, were embedded in the R package. This makes it possible to install MadingleyR using a single R command. The C++ source code as well as the method of compiling can be found on: https://madingleyr.github.io/MadingleyR/SourceCode/. On Linux the MadingleyR package provides the functionality to download and compile the C++ source code directly from R (https://madingleyr.github.io/MadingleyR/SourceCode/CompileLinux.html). All R functions included in the MadingleyR package make up the second component. MadingleyR acts as a wrapper around the included executable, handling all the input and output transfers between C++ and R and providing data visualization options. Because the output data are directly loaded as a standard R object after a simulation, custom plots or further analysis can be conducted in a straightforward way. The major functions in the MadingleyR package were tested using tinytest. In addition, Github Actions automatically deploy a Windows, Mac and Linux server when updates are published to the Github repository to test the functioning of the R package using tinytest. The third component of the MadingleyR package involves the default model parameters and the default spatial input layers. These can be loaded, altered or replaced with study-specific model parameters via the MadingleyR package.

## 3.3 | Model installation

The MadingleyR package can be directly installed from R. Because MadingleyR is hosted on Github (https://github.com/MadingleyR/MadingleyR), the devtools R package (Wickham & Chang, 2016) or the remotes R package (Hester et al., 2019) is required. The installation method is provided in the Github readme file as well as in the pre-compiled MadingleyR tutorial vignette (https://madingleyr.github.io/MadingleyR/Package/vignettes/MadingleyR.pdf). During the installation procedure the MadingleyR tutorial vignette is automatically compiled to test the functionality of the package. After

installation it can be viewed from R using vignette ("MadingleyR"). In addition to installing the MadingleyR dependencies (rgdal, sp, data.table and raster), the installation process also downloads the pre-compiled C++ executable, default spatio-temporal input layers and all other default input parameters and includes them in the installation folder.

## 3.4 | Model initialization

The function madingley_init() initializes a model run by generating a cohort and a stock data set. The cohort data set contains functional information for all cohorts (i.e. heterotrophs) needed to run a Madingley simulation. The stock data set holds the functional information concerning the stocks (i.e. photo-autotrophs). To generate stocks and cohorts using definitions other than the default options, user defined cohort and stock definitions can be supplied to the initialization function. When calling madingley_init() the required spatio-temporal input layers will be loaded automatically (e.g. monthly temperature). Note that all inputs can also be acquired manually, using the madingley_inputs() function. Every input returned from madingley_inputs() can be modified to fit the simulation scenario. The default spatial input layers have a resolution of 1 degree, but running MadingleyR at finer resolutions (e.g. 0.5 degree) is also possible as the resolution is automatically set by the resolution of the input layers. Figure 2 shows how the default inputs can be loaded (Step 1), modified (Step 2) and used to initialize the model (Step 3). Additional input arguments are shown in the R package function description (https://madingleyr.github.io/MadingleyR/Documentation/RFunctionDescription.pdf).

## 3.5 | Running the Madingley model

After generating cohorts and stocks using madingley_init(), a simulation can be started using the madingley_run() function. Although internally the model runs with monthly time steps, the MadingleyR package only allows the user to set simulation lengths in years. A typical Madingley simulation first requires a spin-up phase that allows ecosystem components to reach a stable state. This phase usually consists of a 100- to 1,000-year model simulation without any model user induced changes (Harfoot et al., 2014). The standard model input variables (cohort definitions, stock definitions and spatial inputs) can be changed for madingley_run() via the input arguments, similar to changing the input arguments for madingley_init(). In addition to these inputs, the madingley_run() function also enables users to set the model input parameters. These inputs quantify the strength, rates and constants of the ecological interactions in the model. In the original C++ code these model input parameters were hard-coded (defined directly within the source code). The madingley_inputs() function can be used to acquire the default model parameters.

**TABLE 1** Source code updates

| Update description | References |
|---|---|
| Food-web outputting, this allows for the tracking of feeding directions and rates | Implementation: Hoeks et al. (2020) |
| Large carnivore predator–prey ratio, large carnivores (>21 kg) are able to feed on larger prey | Values: Carbone et al. (1999) Implementation: Hoeks et al. (2020) |
| Spatial control of the initialization of cohort body mass range, maximum body masses can be set using the spatial input layers provided in the package | Values: Wilman et al. (2014); IUCN Red List (2017); Myhrvold et al. (2015); Roll et al. (2017) Implementation: Hoeks et al. (2020) |
| Direct control of all parameters included in the model equations that define the ecological behaviour of cohorts and stocks | This paper |
| I/O (input and output) handling for communication with R | This paper |
| Removal of netCDF dependencies, allowing for more straightforward compilation of the C++ source code on Windows | This paper |

**TABLE 2** Main differences between the published Madingley C# and Madingley C++ versions, and the MadingleyR package

| Feature description | Madingley C# Harfoot et al. (2014) | Madingley C++ Hoeks et al. (2020) | MadingleyR described here |
|---|---|---|---|
| Native operating system compatibility | Windows | Linux, Mac (Windows[a]) | All |
| High-performance computing support (single node in a cluster running on Linux) | – | Yes | Yes |
| Scenario setup: pre-processing of simulation input data in one single, streamlined workflow | – | – | Yes |
| Export a complete model state to initialize and run a consecutive simulation | Yes | – | Yes |
| Spatial control of the initialization of cohort body mass range | – | – | Yes |
| Direct control of all parameters included in the model equations that define the ecological behaviour of cohorts and stocks | – | – | Yes |
| Data visualization functions for various outputs statistics | – | – | Yes |

[a]Native Windows support not available because of netCDF libraries. Support for Windows only via Cygwin or Windows subsystem for Linux with lower performance.

Before the simulation starts, the madingley_run() function will show information concerning the input parameters used. During the simulation run, MadingleyR will print the simulation progress to the R console. Figure 2 shows how the object generated by madingley_init() can be used to run the model (Step 4). Similar to the madingley_init() function, madingley_run() returns a list containing all cohorts and stocks. This list containing the cohorts and stocks can be altered (e.g. changing the properties of the stocks and cohorts) and the spatial inputs (environmental variables) or model input parameters can be modified (Figure 2, Step 5), before they are used to continue the simulation run (Figure 2, Step 6A). Furthermore, the resulting output list can be used to visualize the simulation results using the MadingleyR plotting functions (see Figure 2, Step 6B). Figures 3–5 show the output plots for a simulation of 100 years using the default inputs; the spatial model domain was set from 31 to 35 degrees longitude and from −5 to −1 degrees latitude. The madingley_plot() function creates several plots (see Figures 3–5) from the outputs generated by madingley_run(); the individuals plots can also be produced one by one using the functions presented in the MadingleyR tutorial vignette.

## 4 | CASE STUDIES

We selected three case studies to showcase the MadingleyR package. The first case study demonstrates how MadingleyR can be used to assess the role of functional groups in ecosystems by removing herbivores >100 kg, following a similar procedure developed for large carnivores by Hoeks et al. (2020). In the second case study, we illustrate how the model parameters can be changed to reduce the amount of autotroph biomass available for herbivores and omnivores, mimicking the effect of land-use change on heterotrophs (Newbold et al., 2020). In the third case study, we simulate the reduction in autotroph biomass at a large spatial scale (across the African continent) and observe
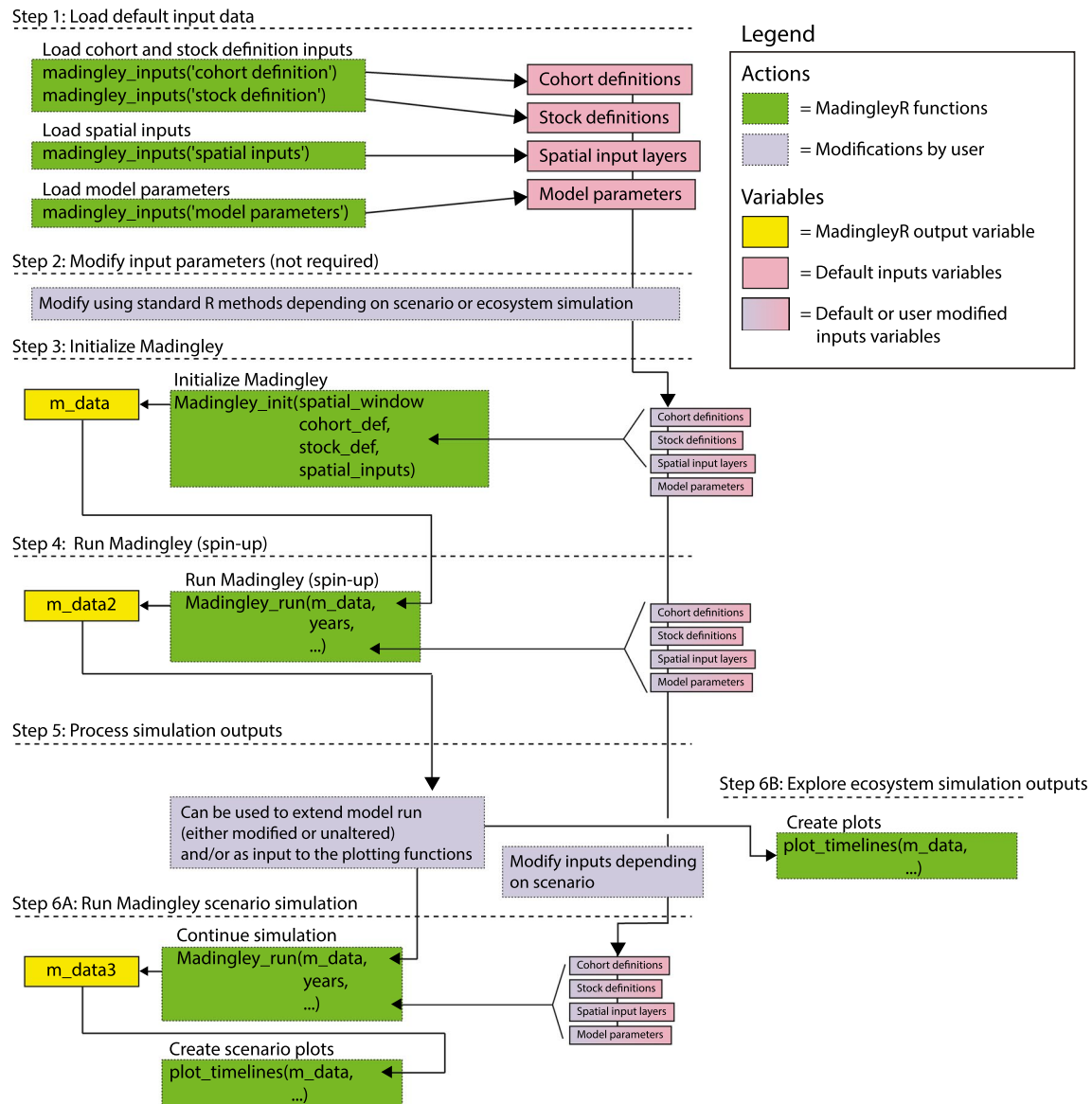
**FIGURE 2** Overview of the MadingleyR workflow [Colour figure can be viewed at wileyonlinelibrary.com]

how it impacts the biomass of herbivores spatially. More information on how to setup and run each of the case study simulations can be found in the case study specific documentations (https://madingleyr.github.io/MadingleyR/CaseStudies/).

## 4.1 | Large-herbivore removal case study

To investigate the role of large and megaherbivores in the Serengeti, we removed endothermic herbivores with a body mass >100 kg from the simulation. This was done by first initializing and running the model for a 100-year period (model spin-up) using the default cohort definitions, stock definitions, spatio-temporal input layers and model parameters. From the resulting MadingleyR object, we removed all endothermic herbivore cohorts with a body mass of >100 kg. The modified object was then used to run a consecutive simulation of 50 years. For the control

simulation, the same principle was applied, but without removing cohorts from the model spin-up object. Food-web results were plotted for means of comparison between the control and removal simulation.

Figure 6 depicts the $\log_{10}$-binned food-web structure of a control simulation (Figure 6a) compared to the simulation in which large and megaherbivores were removed (Figure 6b). The removal of large herbivores led to the disappearance of large endothermic carnivores, an increase in smaller herbivores and a consequent increase in meso-carnivores (Figure 6). Small herbivores indeed benefit from the loss of large predators and reduced competition with large herbivores, while meso-carnivores benefit from the increase in prey availability and simultaneous decrease in competition with larger carnivores (Hoeks et al., 2020). Future simulations could allow further exploration of triggers, tipping points, and mechanisms underlying trophic cascades (e.g. Estes et al., 2011; Knight et al., 2005; Pace et al., 1999; Svenning et al., 2016).
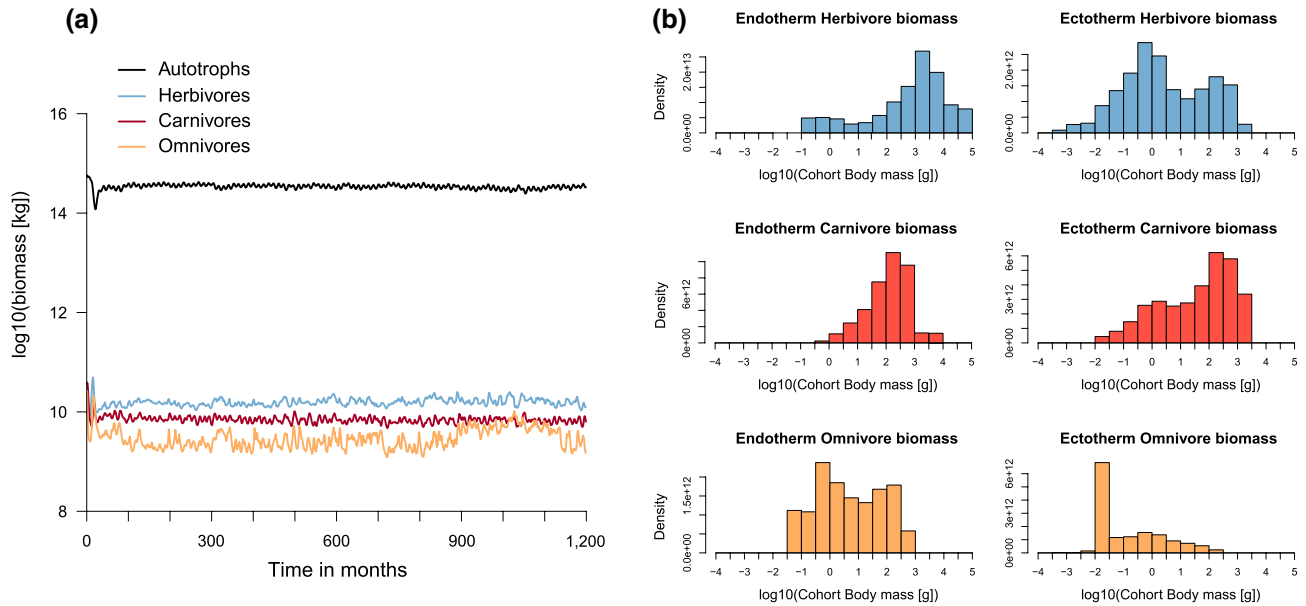
**FIGURE 3** Total $\log_{10}$(biomass in kg) per feeding guild over time (a) and the biomass density per $\log_{10}$-binned body mass category (in g) (b). The plots were made using the default plotting functions included in the MadingleyR package. The biomass density plots represent the yearly average biomass density extracted for the last simulation year. The total simulation time was set to 100 years, the selected spatial model domain included the area between 31 to 35 degrees longitude and −5 to −1 degrees latitude [Colour figure can be viewed at wileyonlinelibrary.com]
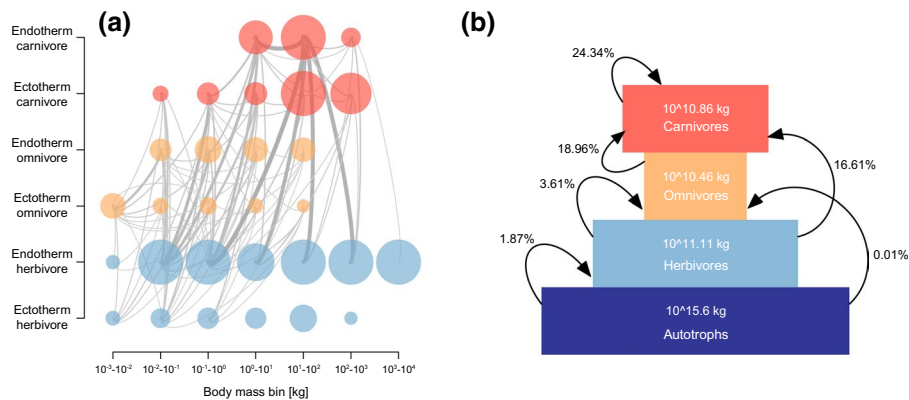
**FIGURE 4** $\log_{10}$-binned food-web (a) and trophic pyramid (b) plotted using the functions included in MadingleyR. Each circle in the first panel (a) represents a food-web node grouping all cohorts within the designated category. Node colour depicts feeding category: carnivores (red), omnivores (orange) and herbivores (blue). The size of the nodes represents the relative differences in total biomass. Grey lines connecting the nodes illustrate the flows. The thickness of the lines depicts the relative amount of biomass flowing between the food-web nodes. The trophic pyramid illustrated in the second panel (b) shows the total biomass (in kg) of four feeding guilds (coloured boxes) and the relative amount of biomass flowing between them. Plots are based on the outputs produced during the final year of a 100-year simulation using the same model domain used to create Figure 3 (31 to 35 degrees longitude and −5 to −1 degrees latitude) [Colour figure can be viewed at wileyonlinelibrary.com]

## 4.2 | Small-scale land-use intensity case study

In the second example, we reduced the biomass production of autotrophs (i.e. vegetation) and observed how it affected the biomass of endotherms. First, a 100-year spin-up simulation was run using the default MadingleyR input parameters (as described in Section 3.1). This spin-up simulation was then extended by an additional 50 years

without any reduction in available autotroph biomass and the end state was used as the control. The 100-year spin-up was then also used to run eight independent land-use intensity experiments where the fraction of autotroph production was reduced by 0.1 increments to test the effects over a gradient of land-use intensities. This was done by modifying the default model input parameters related to the terrestrial carbon model. Each land-use intensity experiment
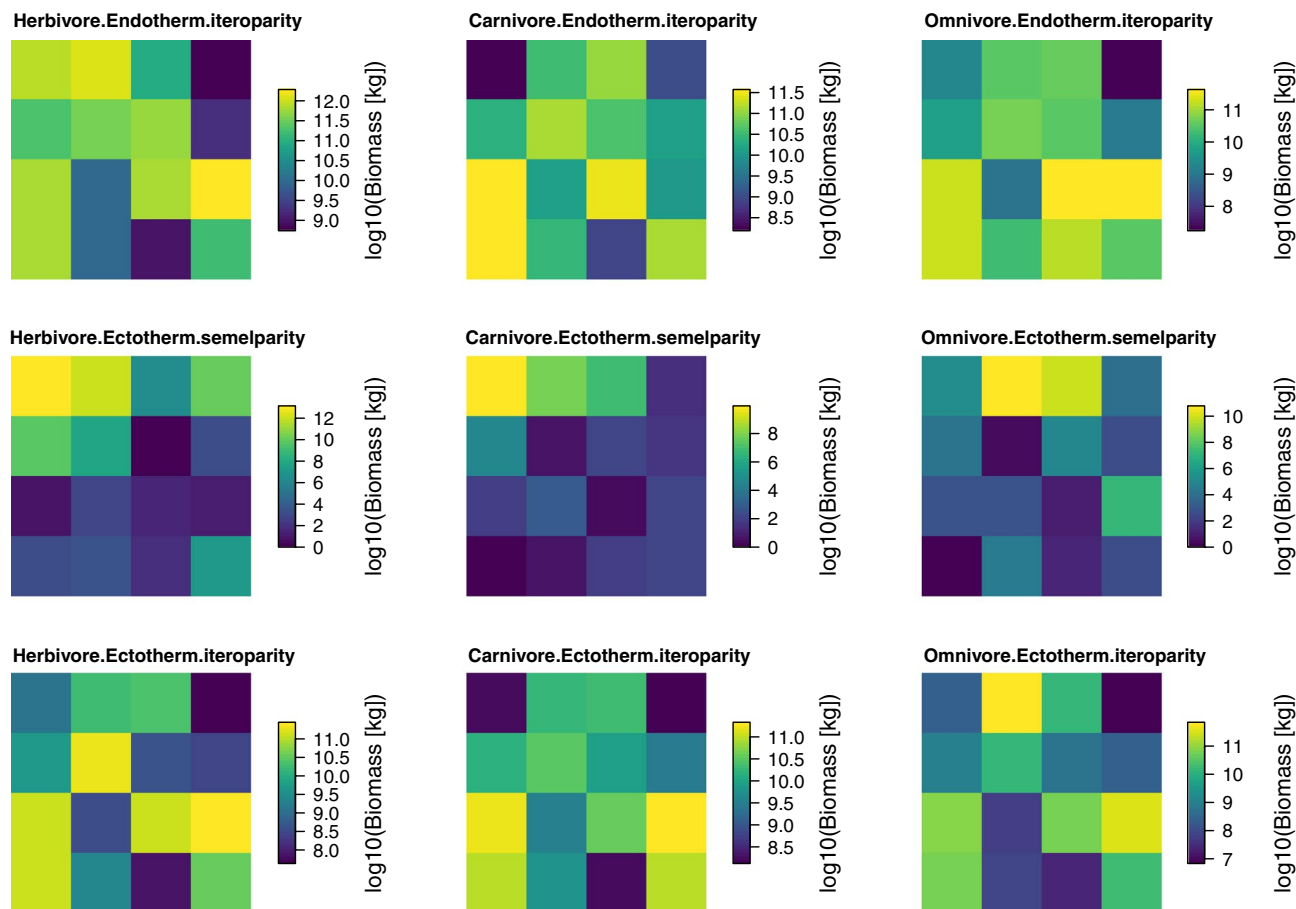
**FIGURE 5** The yearly average total biomass (in $\log_{10}$) plotted per grid cell and per included functional group. The yearly average biomass was computed over the last simulation year. The total simulation time (100 years) as well as the spatial model domain (31 to 35 degrees longitude and −5 to −1 degrees latitude) were set in the R code presented in the tutorial vignette. The plot was made using one of the default plotting functions included in the MadingleyR package [Colour figure can be viewed at wileyonlinelibrary.com]
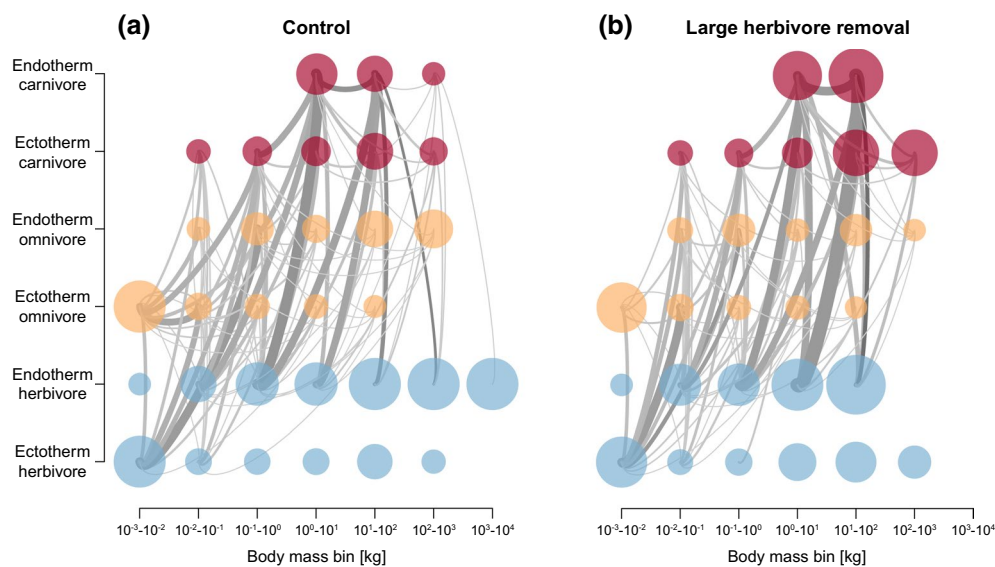
**FIGURE 6** $\log_{10}$-binned food-web plots constructed from a control simulation (a) and a simulation in which large and mega (>100 kg) endothermic herbivores were removed (b). Node colour depicts feeding category: carnivores (red), omnivores (orange) and herbivores (blue). Node size represents the relative differences in total biomass. Grey lines connecting the nodes illustrate the flows between grouped cohorts. The thickness of the lines depicts the relative amount of biomass flowing between the food-web nodes [Colour figure can be viewed at wileyonlinelibrary.com]

was replicated five times and run for 50 years. The end state of each land-use intensity scenario was compared to the control run.

The final plot (Figure 7) was made using base R functions (code provided in the case study specific documentation online). The linear decrease in vegetation biomass available to herbivores results in a nonlinear decrease of the total biomass of the endotherms (Figure 7). Although our method is a simplification of the method applied by Newbold et al. (2020) it is able to highlight a similar principle. In our example location (the Serengeti) the decrease in the proportion of endotherm biomass showed a concave down decreasing trend relative to the fraction of vegetation biomass made inaccessible (Figure 7). This means that the response in cohort biomass was not as strong as we would expect under a linear relationship between vegetation and cohort biomass.

## 4.3 | Large-scale land-use intensity case study

In the third example, we again reduced the biomass production of autotrophs and observed how it affected the biomass of endothermic herbivores on a continental scale. For this case study, we first initialized the model, setting the spatial window to fit the entire continent of Africa. By default, the Madingley model merges similar cohorts once the number of cohorts per spatial grid cell is over 1,000. However, because this case study is meant for illustrative
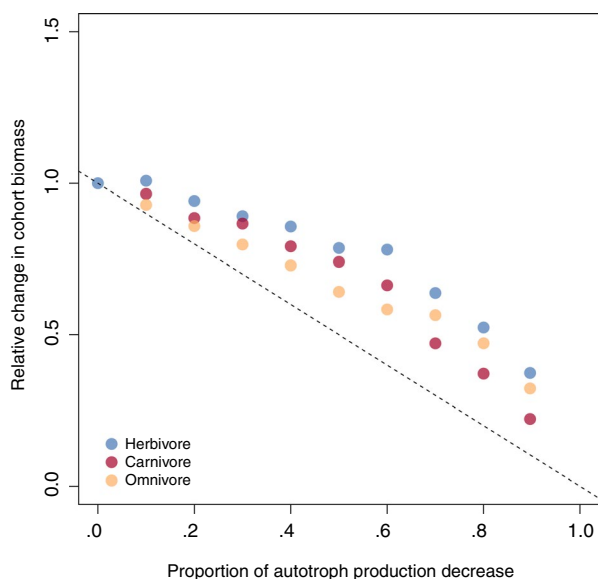


**FIGURE 7** Relative change in biomass of endotherm cohorts compared to the control simulation (biomass experiment/biomass control) plotted against the proportion of plant biomass reduction. A relative change in biomass of 1 indicates no change. Data points represent the relative change in biomass of endothermic carnivores (red), omnivores (orange) and herbivores (blue) averaged over five replicates extracted at the end of the 50-year simulation experiment. The dashed line indicates the impact expected if the biomass of endotherms decreases linearly with the proportion of autotroph production decrease (i.e. $y = -x$) [Colour figure can be viewed at wileyonlinelibrary.com]

purposes only, we lowered the cohort merging threshold to 350. By limiting the number of interactions in each grid cell the total simulation time is decreased. After the initialization, we ran a 200-year spin-up simulation using the default MadingleyR input parameters. Similar to case study 2, we altered the default model parameters in order to reduce autotroph biomass. Here, the vegetation biomass production was lowered by 5% every 5 years, until a 95% reduction was reached. This value (95%) was chosen as it is around the maximum intensity currently experienced by terrestrial ecosystems (Haberl et al., 2007; Newbold et al., 2020). Next, the simulation was continued for an additional 200 years, maintaining the 95% reduction autotroph production.

The biomass of endotherm herbivores was then summed in each grid cell and compared to the end state of the spin-up (Figure 8). The resulting spatial pattern shows that, although the fraction by which the autotroph biomass production was reduced was the same in each grid cell (95% reduction), herbivores are able to persist in high productive areas (see Figure 8). As explained by Newbold et al. (2020), systems characterized by low seasonality and high productivity can show a rapid degradation in ecosystem structure if a certain threshold of inaccessible vegetation biomass is surpassed. However, before reaching this threshold, these ecosystems may appear relatively stable.

## 5 | DISCUSSION

The Madingley model enables the study of ecosystem structure and function on a global scale (Harfoot et al., 2014). However, due to both its mathematical and technical complexity the original C++ and C# versions of Madingley remain inaccessible to many. The barrier for potential practitioners to apply these complex ecological models may be overcome by providing comprehensive explanation of its content and by producing open-source and user-friendly software (Rapacciuolo, 2019). The MadingleyR R package increases the accessibility and reproducibility of the Madingley model by means of a clearly structured, user-friendly R interface that communicates with the fast C++ code.

The major advantages of this R package over using the C++ or C# version Madingley model by itself include: (a) installation from R, (b) straightforward manipulation of input objects, (c) ability to run multiple consecutive simulations, thereby increasing simulation experiments' flexibility, (d) easy visualization and further processing of outputs, (e) compatibility with all operating systems. The MadingleyR package allows for small-scale test simulations to be executed on a local machine (e.g. Windows, Linux or MacOS laptop). If needed, the MadingleyR package makes it relatively straightforward to run the computations on a Linux powered server without any changes to the model setup, making use of parallelization and speeding up the simulation time. Storing all simulation outputs in a single R object simplifies the workflow, enables reproducibility, and makes collaboration easier. The ability to easily change the parameter settings in
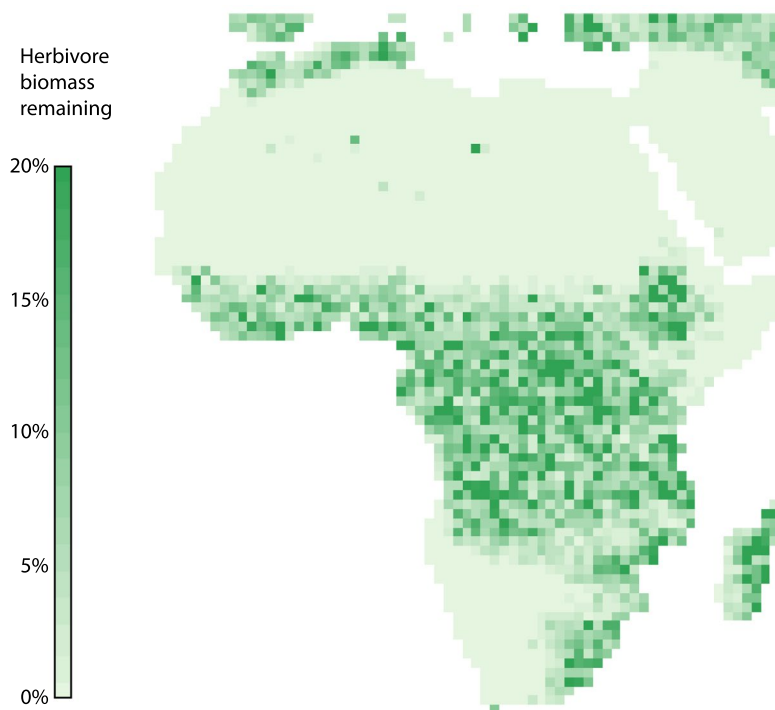
**FIGURE 8** Endotherm herbivore biomass remaining after reducing stock production (i.e. vegetation) by 95% in each grid cell. Biomass reduction calculated as the biomass before reduction/biomass after reduction * 100 [Colour figure can be viewed at wileyonlinelibrary.com]

the MadingleyR model enables extensive exploration of the model behaviour and testing of parameter sensitivity.

There are also limitations to the MadingleyR package that require further development. Most importantly, modifications of the fundamental process equations cannot be done from R. Changes in processes need to be directly implemented in the C++ source code, as is the case when using the full C++ implementation. We provide access to the fully commented C++ code and step by step documentation for compiling it for use in the MadingleyR package in order to make this option as user-friendly as possible. Additionally, the MadingleyR model was developed for the terrestrial realm only, so one of the next steps will be the inclusion of the marine realm.

The increased accessibility to the Madingley model offered by the MadingleyR package can help facilitate the exploration of the mechanisms underlying macroecological patterns (Beck et al., 2012; Enquist et al., 2020; Harfoot et al., 2020; Hoeks et al., 2020; Wüest et al., 2020). In addition, the use of Madingley may lead to generating new hypotheses and possibly new research avenues where different environmental scenarios can be examined. This is particularly interesting for exploring different scenarios and making baseline predictions of potential outcomes of human actions (Fitzpatrick & Hargrove, 2009).

Besides the potential for the MadingleyR package for researchers, it may also be relevant for education. By letting students interact with simulations of ecosystems and by letting them interpret the results (i.e. ecosystem responses) following perturbations could be a productive way to educate students on principles of ecology such as: prey–predator dynamics, ecosystem structure, top-down and bottom-up control, impacts of land-use change and the consequences of hunting (e.g. Bartlett et al., 2016; Harfoot et al., 2020; Hoeks et al., 2020; Newbold et al., 2020).

In conclusion, MadingleyR is designed to facilitate the use of the Madingley model for terrestrial ecosystem simulations and tackles multiple aspects associated with the flexibility and transparency of such complex and large-scale models (Peng, 2011; Rapacciuolo, 2019). The advantages of the MadingleyR package provide the opportunity for a wider use of mechanistic generic ecosystem modelling in future ecological research, conservation practice and education.

## AUTHOR CONTRIBUTIONS

S.H. conceptualized and developed the MadingeyR package. M.B. converted the original C# code into C++. S.H., L.S. and M.A.J.H. designed the case studies and S.H. ran the simulations. S.H. wrote the manuscript and all authors contributed to the manuscript and gave approval for publication.

## DATA AVAILABILITY STATEMENT

The MadingleyR R package is hosted on Github (https://github.com/MadingleyR/MadingleyR). The source code for the included (pre-compiled) C++ executable can be found at https://github.com/MadingleyR/MadingleyR/tree/master/SourceCode. All R

codes used in this publication can be found in the documentation on Github (https://madingleyr.github.io/MadingleyR/). A persistent copy of the MadingleyR package, documentation files and the C++ source code presented in this paper are archived on Zenodo (https://doi.org/10.5281/zenodo.4790806) as well as in a Github release (https://github.com/MadingleyR/MadingleyR/releases/tag/GEB).

## ORCID

*Selwyn Hoeks* (iD) https://orcid.org/0000-0001-5619-3233
*Mike B. J. Harfoot* (iD) https://orcid.org/0000-0003-2598-8652
*Luca Santini* (iD) https://orcid.org/0000-0002-5418-3688

## REFERENCES

Bartlett, L. J., Newbold, T., Purves, D. W., Tittensor, D. P., & Harfoot, M. B. (2016). Synergistic impacts of habitat loss and fragmentation on model ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20161027. https://doi.org/10.1098/rspb.2016.1027

Barychka, T., Mace, G. M., & Purves, D. W. (2019). Modelling variation in bushmeat harvesting among seven African ecosystems using the Madingley Model: Yield, survival and ecosystem impacts. *BioRxiv*, https://doi.org/10.1101/695924

Barychka, T., Mace, G. M., & Purves, D. (2020). The Madingley General Ecosystem Model predicts bushmeat yields, species extinction rates and ecosystem-level impacts of bushmeat harvesting. *BioRxiv*, https://doi.org/10.1101/2020.03.02.959718

Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M., & Dormann, C. F. (2012). What's on the horizon for macroecology? *Ecography*, *35*, 673–683. https://doi.org/10.1111/j.1600-0587.2012.07364.x

Cabral, J. S., Valente, L., & Hartig, F. (2017). Mechanistic simulation models in macroecology and biogeography: State-of-art and prospects. *Ecography*, *40*, 267–280. https://doi.org/10.1111/ecog.02480

Carbone, C., Mace, G. M., Roberts, S. C., & Macdonald, D. W. (1999). Energetic constraints on the diet of terrestrial carnivores. *Nature*, *402*, 286–288. https://doi.org/10.1038/46266

Connolly, S. R., Keith, S. A., Colwell, R. K., & Rahbek, C. (2017). Process, mechanism, and modeling in macroecology. *Trends in Ecology and Evolution*, *32*, 835–844. https://doi.org/10.1016/j.tree.2017.08.011

Coughenour, M. B. (1993). *The SAVANNA landscape model-documentation and users guide*. Natural Resource Ecology Laboratory, Colorado State University.

Enquist, B. J., Abraham, A. J., Harfoot, M. B., Malhi, Y., & Doughty, C. E. (2020). The megabiota are disproportionately important for biosphere functioning. *Nature Communications*, *11*, 1–11. https://doi.org/10.1038/s41467-020-14369-y

Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., Carpenter, S. R., Essington, T. E., Holt, R. D., Jackson, J. B. C., Marquis, R. J., Oksanen, L., Oksanen, T., Paine, R. T., Pikitch, E. K., Ripple, W. J., Sandin, S. A., Scheffer, M., Schoener, T. W., … Wardle, D. A. (2011). Trophic downgrading of planet Earth. *Science*, *333*, 301–306. https://doi.org/10.1126/science.1205106

Fitz, H. C., DeBellevue, E. B., Costanza, R., Boumans, R., Maxwell, T., Wainger, L., & Sklar, F. H. (1996). Development of a general ecosystem model for a range of scales and ecosystems. *Ecological Modelling*, *88*, 263–295. https://doi.org/10.1016/0304-3800(95)00112-3

Fitzpatrick, M. C., & Hargrove, W. W. (2009). The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, *18*, 2255. https://doi.org/10.1007/s10531-009-9584-8

Grimm, V., Ayllón, D., & Railsback, S. F. (2017). Next-generation individual-based models integrate biodiversity and ecosystems: Yes we can, and yes we must. *Ecosystems*, *20*, 229–236. https://doi.org/10.1007/s10021-016-0071-2

Haberl, H., Erb, K. H., Krausmann, F., Gaube, V., Bondeau, A., Plutzar, C., Gingrich, S., Lucht, W., & Fischer-Kowalski, M. (2007). Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proceedings of the National Academy of Sciences USA*, *104*, 12942–12947. https://doi.org/10.1073/pnas.0704243104

Harfoot, M. B., Abraham, A., Tittensor, D. P., Costa, G. C., Faurby, S., Feldman, A., Itescu, Y., Meiri, S., Morales-Castilla, I., Oliveira, B. F., & Purves, D. (2020). Mechanistic macroecology: exploring the drivers of latitudinal variation in terrestrial body size in a General Ecosystem Model. bioRxiv. https://doi.org/10.1101/775957

Harfoot, M. B. J., Newbold, T., Tittensor, D. P., Emmott, S., Hutton, J., Lyutsarev, V., Smith, M. J., Scharlemann, J. P. W., & Purves, D. W. (2014). Emergent global patterns of ecosystem structure and function from a mechanistic general ecosystem model. *PLoS Biology*, *12*, e1001841. https://doi.org/10.1371/journal.pbio.1001841

Hester, J., Csárdi, G., Wickham, H., Chang, W., Morgan, M., & Tenenbaum, D. (2019). remotes: R package installation from remote repositories, including 'GitHub'. R package version 2.4.0. https://cran.r-project.org/web/packages/remotes/index.html

Hoeks, S., Huijbregts, M. A. J., Busana, M., Harfoot, M. B. J., Svenning, J.-C., & Santini, L. (2020). Mechanistic insights into the role of large carnivores for ecosystem structure and functioning. *Ecography*, *43*, 1752–1763. https://doi.org/10.1111/ecog.05191

IUCN. (2017). The IUCN red list of threatened species. Ver. 2017-1. IUCN. www.iucnredlist.org

Knight, T. M., McCoy, M. W., Chase, J. M., McCoy, K. A., & Holt, R. D. (2005). Trophic cascades across ecosystems. *Nature*, *437*, 880–883. https://doi.org/10.1038/nature03962

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., & Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, *19*(1), https://doi.org/10.1029/2003GB002199

Kucharik, C. J., Foley, J. A., Delire, C., Fisher, V. A., Coe, M. T., Lenters, J. D., Young-Molling, C., Ramankutty, N., Norman, J. M., & Gower, S. T. (2000). Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure. *Global Biogeochemical Cycles*, *14*, 795–825. https://doi.org/10.1029/1999GB001138

Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., & Ma, K. (2019). Evaluating the popularity of R in ecology. *Ecosphere*, *10*, e02567. https://doi.org/10.1002/ecs2.2567

Lorek, H., & Sonnenschein, M. (1999). Modelling and simulation software to support individual-based ecological modelling. *Ecological Modelling*, *115*, 199–216. https://doi.org/10.1016/S0304-3800(98)00193-8

McGill, B. J. (2019). The what, how and why of doing macroecology. *Global Ecology and Biogeography*, *28*, 6–17. https://doi.org/10.1111/geb.12855

McGill, B. J., & Nekola, J. C. (2010). Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. *Oikos*, *119*, 591–603. https://doi.org/10.1111/j.1600-0706.2009.17771.x

Myhrvold, N. P., Baldridge, E., Chan, B., Sivam, D., Freeman, D. L., & Ernest, S. M. (2015). An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology*, *96*, 3109. https://doi.org/10.1890/15-0846R.1

Newbold, T., Tittensor, D. P., Harfoot, M. B., Scharlemann, J. P., & Purves, D. W. (2020). Non-linear changes in modelled terrestrial ecosystems subjected to perturbations. *Scientific Reports*, *10*, 1–10. https://doi.org/10.1038/s41598-020-70960-9

Pace, M. L., Cole, J. J., Carpenter, S. R., & Kitchell, J. F. (1999). Trophic cascades revealed in diverse ecosystems. *Trends in Ecology and Evolution*, *14*, 483–488. https://doi.org/10.1016/S0169-5347(99)01723-1

Peng, R. D. (2011). Reproducible research in computational science. *Science*, *334*, 1226–1227. https://doi.org/10.1126/science.1213847

Purves, D., Scharlemann, J. P., Harfoot, M., Newbold, T., Tittensor, D. P., Hutton, J., & Emmott, S. (2013). Time to model all life on Earth. *Nature*, *493*, 295–297. https://doi.org/10.1038/493295a

Rapacciuolo, G. (2019). Strengthening the contribution of macroecological models to conservation practice. *Global Ecology and Biogeography*, *28*, 54–60. https://doi.org/10.1111/geb.12848

R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria. https://www.R-project.org/

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., Böhm, M., Castro-Herrera, F., Chirio, L., Collen, B., Colli, G. R., Dabool, L., Das, I., Doan, T. M., Grismer, L. L., Hoogmoed, M., Itescu, Y., Kraus, F., LeBreton, M., … Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology & Evolution*, *1*, 1677–1682. https://doi.org/10.1038/s41559-017-0332-2

Smith, M. J., Purves, D. W., Vanderwel, M. C., Lyutsarev, V., & Emmott, S. (2013). The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties. *Biogeosciences*, *10*(1), 583–606. https://doi.org/10.5194/bg-10-583-2013.

Svenning, J.-C., Pedersen, P. B. M., Donlan, C. J., Ejrnæs, R., Faurby, S., Galetti, M., Hansen, D. M., Sandel, B., Sandom, C. J., Terborgh, J. W., & Vera, F. W. M. (2016). Science for a wilder Anthropocene: Synthesis and future directions for trophic rewilding research. *Proceedings of the National Academy of Sciences USA*, *113*, 898–906. https://doi.org/10.1073/pnas.1502556112

White, M. A., Thornton, P. E., Running, S. W., & Nemani, R. R. (2000). Parameterization and sensitivity analysis of the BIOME–BGC terrestrial ecosystem model: Net primary production controls. *Earth Interactions*, *4*, 1–85. 10.1175/1087-3562(2000)004<0003:PASAOT>2.0.CO;2

Wickham, H., & Chang, W. (2016). Devtools: Tools to make developing R packages easier. R package version 2.4.2. https://cran.r-project.org/web/packages/devtools/index.html

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals: Ecological Archives E095–178. *Ecology*, *95*, 2027. https://doi.org/10.1890/13-1917.1

Wüest, R. O., Zimmermann, N. E., Zurell, D., Alexander, J. M., Fritz, S. A., Hof, C., Kreft, H., Normand, S., Cabral, J. S., Szekely, E., Thuiller, W., Wikelski, M., & Karger, D. N. (2020). Macroecology in the age of Big Data-Where to go from here? *Journal of Biogeography*, *47*, 1–12. https://doi.org/10.1111/jbi.13633

## BIOSKETCH

**Selwyn Hoeks** is a PhD candidate at the Radboud University in The Netherlands. In his research he focuses on explaining ecosystem functioning using mechanistic modelling approaches. His main interest is the importance of megafauna, in particular the role of large carnivores in maintaining ecosystem structure. Besides a strong passion for ecology, he likes code, especially C++.