Theory-driven analysis of ecological data: a practical handbook

us

2024-11-06

Contents

Chapter 1

About Bookdown

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports; for example, a math equation $a^2 + b^2 = c^2$.

1.1 Usage

Each **bookdown** chapter is an .Rmd file, and each .Rmd file can contain one (and only one) chapter. A chapter *must* start with a first-level heading: `# A good chapter`, and can contain one (and only one) first-level heading.

Use second-level and higher headings within chapters like: `## A short section` or `### An even shorter section`.

The `index.Rmd` file is required, and is also your first book chapter. It will be the homepage when you render the book.

1.2 Render book

You can render the HTML version of this example book without changing anything:

1. Find the **Build** pane in the RStudio IDE, and

2. Click on **Build Book**, then select your output format, or select "All formats" if you'd like to use multiple formats from the same book source files.

Or build the book from the R console:

```
bookdown::render_book()
```

To render this example to PDF as a `bookdown::pdf_book`, you'll need to install XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.org/tinytex/.

1.3 Preview book

As you work, you may start a local server to live preview this HTML book. This preview will update as you edit the book when you save individual .Rmd files. You can start the server in a work session by using the RStudio add-in "Preview book", or from the R console:

```
bookdown::serve_book()
```

1.4 Here are some useful things for writing the book using bookdown

All chapters start with a first-level heading followed by your chapter title, like the line above. There should be only one first-level heading (`#`) per .Rmd file.

1.4.1 A section

All chapter sections start with a second-level (`##`) or higher heading followed by your section title, like the sections above and below here. You can have as many as you want within a chapter.

An unnumbered section

Chapters and sections are numbered by default. To un-number a heading, add a `{.unnumbered}` or the shorter `{-}` at the end of the heading, like in this section.

1.5 Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

1.5.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
   - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world` = `# Hello world {#hello-world}`.
   - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.

2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 1.5.
    - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

1.5.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 1.1.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

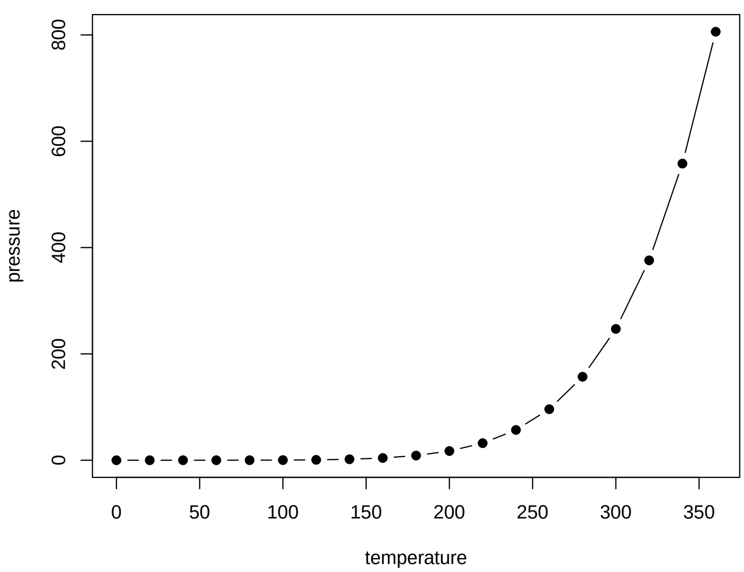

Figure 1.1: Here is a nice figure!

Don't miss Table 1.1.

```
knitr::kable(
  head(pressure, 10), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

Table 1.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

1.6 Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART\*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

1.7 Footnotes and citations

1.7.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one [1].

1.7.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2023] (check out the last code chunk in index.Rmd to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [Xie, 2015] (this citation was added manually in an external file book.bib). Note that the `.bib` files need to be listed in the index.Rmd with the YAML `bibliography` key.

[1] This is a footnote.

The RStudio Visual Markdown Editor can also make it easier to insert citations: https://rstudio.github.io/visual-markdown-editing/#/citations

1.8 Blocks

1.8.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (1.1).

1.8.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 1.1.

Theorem 1.1. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the* ***other*** *two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

1.8.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

1.9 Sharing your book

1.9.1 Publishing

HTML books can be published online, see: https://bookdown.org/yihui/bookdown/publishing.html

1.9.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

1.9.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book-all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/gitbook.html

Or use:

```
?bookdown::gitbook
```

Chapter 2

Preambule

Who is the textbook for?

Chapter 3

Why mathematical models? (Sonia)

To read: Deep Symbolic Regression for Physics Guided by Units Constraints: Toward the Automated Discovery of Physical Laws https://iopscience.iop.org/article/10.3847/1538-4357/ad014

Une intelligence artificielle retrouve des lois physiques à partir de données scientifiques https://www.insu.cnrs.fr/fr/cnrsinfo/une-intelligence-artificielle-retrouve-des-lois-physiques-partir-de-donnees-scientifiques#:~:text=La%20m%C3%A9thode%20d'intelligence%20artificielle,ces%20r%C3%A8gles%20d'analyse%20dimensionnelle

3.1 Big data

According to IBM, every day humanity generates 2.5 trillion (2.5 billion billion) bytes of text, image and sound data (https://www-01.ibm.com/software/fr/data/bigdata). Acceleration of information accumulation. We can learn from massive amount of data available.

When read by machines (analyses, they can reveal a wealth of unsuspected correlations (or that are difficult to identify). Success of Amazon or Netflix at learning what you like from studying your habits.

Machine learning (ML), a subset of AI that enables computers to learn from training data, has been highly effective at predicting various types of cancer, including breast, brain, lung, liver, and prostate cancer. In fact, AI and ML have demonstrated greater accuracy in predicting cancer than clinicians.

So, knowledge can emerge from this data analysis (even without any idea of the underlying laws).

This lead some people to suggest that we may not need theory anymore ("The end of theory: the data deluge, Wired). The availability of data and the development of methods to analyse them: does they mean that we're going through a major epistemological change? In other terms: do we still need models and theory?

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711825/

3.2 What is theory?

We define *ecological theory* broadly as an explanation of an ecological phenomenon. These explanations take the form of narratives that explain how an ecological process works, or why an ecological pattern is observed, and that becomes scientifically useful when expressed in a logical structure (Pickett et al. 2010; Rossberg et al. 2019).

Maris et al. 2017: "Scientific theories contain universal or general propositions regarding the system in question; they generally encompass a set of models, or rules to build models, as representations of the properties and dynamics of target systems. Models can then be expressed as hypotheses within a formal framework (van Fraassen 1980, Giere 2004)."

The transformation of an idea in narrative form into a logical, testable theory often, though not always, involves the use of *models* (Otto and Rosales 2020).

3.3 What's a model?

https://en.wikipedia.org/wiki/Ecosystem_model

The concept of a model is derived from scale models, referring to simplified replicas of larger structures such as, for instance, buildings or ships. This implies a structural similarity between a model and its original. Thus, **models are idealized representations of certain aspects of our study systems. They are idealized versions of reality**, just as architectural models represent key features of complex structures and model organisms represent a group of organisms that share common attributes (Kokko 2007).

Models can be as simple as a verbal statement about a subject or two boxes connected by an arrow to represent some relationship (conceptual models). Alternatively, models can be extremely complex and detailed, such as a mathematical description of the pathways of nitrogen transformations within ecosystems.

A mathematical model can be an equation or a set of equations (mathematical expressions) that describes how different aspects of a system relate to one another (Otto and Day 2007). They can be phenomenological or mechanistic (Grainger et al. 2022).

Theory and math are not inexorably linked. Indeed, many excellent theories do not involve math (e.g., the theory of evolution by natural selection; Darwin 1859), and many uses of math in ecology are not theory (e.g., practical applications of statistics).

Because models are an idealized, simplified version of the real world (e.g. maps), they are therefore not real. They're incomplete, they're wrong (intro Kokko's book). **"all models are false but some models are useful"** https://en.wikipedia.org/wiki/All_models_are_wrong

The systems we're interested in are complex. How can we hope capturing the complexity of reality? We can't. And if we did, it would not be helpful (if models are as complex as reality, e.g. map) because of:

- too many parameters to measure (would require too much time and resources),
- the equations are insoluble and require too long to run on even a good computer,
- even if soluble, we might not be able to understand the results.

Modeling involves a choice about what to include and what to leave out. We need to simplify systems in a way that preserves the essential features of the system (depending on our interest). This means that our big task as modelers is to decide what goes in the model and what doesn't. The art of modeling is to decide which aspect of reality one can sacrifice and which ones are crucial to retain.

We want to work with manageable models which max generality, realism and precision towards the goals of understanding and predicting….

These goals typically compete with each other so real models are mathematical descriptions that result from tradeoffs among these goals which depend on our needs: the tension between realism, generality, and precision: **Levins 1966** https://v4.chriskrycho.com/2016/realism-generality-and-precision-in-tension.html

- One can sacrifice generality to realism and precision. E.g. fisheries; good measurement of shot-time behavior, numerical solution, precise testable predictions applicable to a particular situation).

- One can sacrifice realism to generality and precision. E.g. physics-like models (general equations such as LV). The way in which nature deviates from theory will indicate where further complexity will be useful.

- One can sacrifice precision to realism and generality. E.g. MacArthur 1965. People concerned with the qualitative behavior (not quantitative). Graphical models.

3.4 Purpose of science?

Understand (and predict).

3.5 Can we do that with big data analysis?

Big data analysis reveals correlations. (Again, this can be useful) But those correlations are not causal. Famous examples of spurious correlations. Correlations can emerge by chance because you look at some many variables. Correlations can happen because of confounding factors, e.g. the example of shoe size and level of math in a school (due to age)

So we can find correlations, but we often don't know why. So we can't understand (understanding is finding associations that are causal).

This does not mean that non-causal correlations can't help predict! They can have great predictive power. Amazon, Netflix. So we can learn from those data and their analysis, even without causal knowledge.

We can predict, but we can also fail at predicting. Predictions can fail because lacks of general rules, so extrapolations can fail; financial crisis, elections. It relies on what happened in the past, so unable to work in new situations.

Predict without understanding.... act without understanding? Risks (cf justice, police, discriminations)

Big data don't create theory, they need it to be exploited. (This last statement could evolve in the future? In particular, new analyses methods suggest they can successfully retrive causality from correlations; cg Zach, Stan's paper with Correlation Cross mapping).

Because of that, we can argue that we still do need theory and models! (models are one langage of theory, one way of expressing it)

3.6 The scientific method

Let's step back and reflect on the process by which science is carried out. The *scientific method* is an empirical method for acquiring knowledge that has characterized the development of science since at least the 17th century.

Science (through the scientific method) can build on previous knowledge and develop a more sophisticated understanding of its topics of study over time.

- **'Problem' identification**: It involves careful observation which leads to the formulation of a question.
- **Hypothesis**: A hypothesis is a conjecture (hypothetical explanations), based on the observation/the knowledge obtained while formulating the question, that may explain any given behavior. A scientific hypothesis

must be falsifiable, implying that it is possible to identify a possible outcome of an experiment or observation that conflicts with predictions deduced from the hypothesis; otherwise, the hypothesis cannot be meaningfully tested. Falsifiability is a deductive standard of evaluation of scientific theories and hypotheses, introduced by the philosopher of science Karl Popper in his book The Logic of Scientific Discovery (1934). A theory or hypothesis is falsifiable (or refutable) if it can be logically contradicted by an empirical test.

- **Prediction**: The prediction step deduces the logical consequences of the hypothesis before the outcome is known.
- **Testing**: Hypotheses are tested by conducting experiments or gathering observations. The purpose of the test is to determine whether observations agree with or conflict with the expectations deduced from a hypothesis.
- **Refinement (or elimination) of the hypotheses** based on the experimental findings.

Although procedures vary from one field of inquiry to another, the underlying process is frequently the same. In sum, the process is as follows: making conjectures (hypotheses), deriving predictions from them as logical consequences, and then carrying out experiments based on those predictions to determine whether the original conjecture was correct.

The scientific method is an iterative, cyclical process through which information is continually revised.

3.7 A feedback loop involving data and models

A scientific understanding of the biological world arises when ideas about how nature works are formalized, tested, refined, and then tested again.

Scientific inquiry should operate as a feedback loop in which theory that describes the natural world is developed, tested empirically through carefully articulated hypotheses, modified to better represent reality, and then tested again.

When this feedback loop works, **theory** provides a framework to guide inquiry, experimental design, and the interpretation of observed patterns, supplies mathematical tools to harness information from collected data, and connects individual experiments to general ideas about how nature operates.

In turn, **empirical research** can be used to support, refute, or revise theoretical predictions, indicate which theoretical assumptions are consistent with the natural world, and point theoreticians to overlooked processes that can be integrated into models.

However, there is currently a disconnection between theoretical and empirical research.

Although the benefits of feedback between theoretical and empirical research are widely acknowledged by ecologists, this link is still not as strong as it could be in ecological research.

Indeed, up to 45% of articles on empirical ecology make no mention of any theory whatsoever (Scheiner 2013), and fewer than 10% of ecologists and evolutionary biologists agree with the statement that 'theoretical findings drive empirical work' in their fields (Haller 2014).

Further references about the disconnect between empirical and theoretical work in ecology: Lomnicki 1988; Kareiva 1989; Fawcett and Higginson 2012; Scheiner 2013; Haller 2014; Rossberg et al. 2019.

Why this disconnection?

- A lack of theoretical training in ecology (Rossberg et al. 2019).
- A lack of motivation on the part of some theoreticians to engage with the language of empiricists (Grimm 1994) or with the elements of nature that empiricists focus on (Krebs 1988).
- A general lack of mutual appreciation between empiricists and theoreticians (Haller 2014).
- Persistent communication barriers between these two groups (Servedio 2020), in particular that theory is expressed in the langage of math (some ecologists may not have formal background in math and theoretical papers may not be written with a general audience, not always explaining asumptions, terminology and notations), some aspects of the theory may seem inaccessible by some, the more equations an ecology and evolutionary biology article contains, the fewer citations it receives (Fawcett and Higginson 2012).

This barrier presents a major challenge to the full integration of theoretical and empirical work in ecology.

A better integration of theory into empirical work is needed (Caswell 1988; Pickett et al. 2010; Marquet et al. 2014; Servedio et al. 2014; Servedio 2020).

This is especially important in the context of global change.

This is why this course/book!

Chapter 4

What types of theoretical models in ecology? (Isa)

4.1 What system? What question? What hypotheses (no maths shown here)

- Question is related to a **specific level of organization** / scale and determine what can be ignored
- Examples of questions and processes of interest for each scale based on our recurrent example plant-herbivore (P-H)
- **Updated Levin's triangle** (Bullock 2014) illustrated by different (P-H) models (conceptual, system-specific, generic) with related questions => define variables
- **What can be ignored?** The differences in time scale between processes allow to do some approximations: Upper level is slower than lower ones and its dynamics may be ignored depending on the question. Similarly, lower levels can be considered at equilibrium if much faster than the focal dynamics, or averaged (macroscale averaged view of microscale processes => physics statistics).
- **Link between assumptions, model and predictions**, with the risk of overinterpretation: always keep in mind that what mathematical methods is truth only in the context of initial assumptions.
- **The different types of assumptions** (Servedio 2014):
  - Critical assumptions: the ones necessary to answer the question (structure of the model).
  - Exploratory assumptions: the ones to go deeper in the exploration, assess the robustness of the results.
  - Logistical assumptions: the ones that have to deal with tractability.
- **Illustration of assumption types** with the Grazing-optimisation exam-

ple (De mazancourt et al. 1998):
  - Critical assumptions: two paths of recycling with different rates for plants and herbivores.
  - Exploratory assumptions: change in the functional response.
  - Logistical assumptions: use of deterministic ODEs and continuous time with no space (we'll see all that in more depth in next section).

4.2 What model formalism? (no maths shown here)

4.2.1 Deterministic – stochastic processes

- **What does stochasticity come from?**
  - Environmental (eg: perturbations): does not scale with population size and can vary in space and time.
  - Demographic (eg: more or less offsprings, trait variability that will impact biological rates): scales with population size.
- **When should we account for it** (illustration with pop logistic growth (no math) with addition or not of gaussian noise)
  - for questions related to output variability
  - when output distribution is skewed and mean no more a good predictor (small populations) => when stochastic processes are dominant (ecological or genetical drifts, Allee effect, large demographic stochasticity)
- **When can we use deterministic**: when processes can be summarised with averaged parameters (large populations)
- NB: mention of master equations (deterministic equation summarizing stochastic model)

4.2.2 Time: discrete – continuous

How do the processes are structured along time? when using discrete vs continuous time

- **What are the underlying assumptions?**
  - *Discrete-time*:
    - Fixed generation times
    - Synchronization of processes (e.g., seasonal forcing)
    - Sequential processes with specific order (e.g., complex life-cycle) => more transparent
  - *Continuous time*:
    - Events can happen at any time
    - Generation overlap
- **Illustration with logistic pop growth** in continuous and discrete time:
  - discrete-time with small t boils down to continuous

  - discrete time have richer dynamics due to a lag
- **Appropriate systems / questions**
  - discrete-time:
    * questions related to phenology
    * seasonal dynamics and synchronized generations
    * complex life cycled
  - continuous-time:
    * generation overlap
    * when processes can happen at any time
- Both can be either stochastic or deterministic

4.2.3 Accounting for space

- **None vs spatial**:
  - TIB, example of space implicit; In none spatial models, space is assumed homogenous and the scale can be integrated within the units (example: pop density in ind/km2)
  - When does the question need it explicit space?
    * When spatial processes impact local dynamics of interest
    * if we are interested in patterns emerging at higher spatial scale from local processes
- **Topology versus Distances explicit**: does geographical positions matter or only topology?
  - Topology (metapop model): space is represented in the structure of space and strength of connectivity. Adapted to study fragmented landscapes or problems related to structure of connectivity
  - Explicit: distances are explicit; in 1 (Fisher KPP) or two dimensions (grid): adapted to represents fronts, gradients, edge effects or emerging spatial patterns.
- **Discrete within the distance explicit case**: grids vs Continuous (PDE):
  - *Continuous* adapted to model processes occurring progressing locally (diffusion) but not processes with large distance / heterogeneous effects
  - *Grids* with small mesh size can boil down to continuous (sometimes, discretisation is used for a technical approximation of continuous space) => question of technical relevance.
  - As continuous can be more difficult (math) this can prevent its use (mention of lab inheritance)

4.3 What technical choices?

4.3.1 Analytical – Numerical

What do we gain in making models simple? => Principle of parcimony

- **Analytical solution**: we know the state of the model at any time point given $\frac{dN}{dt} \rightarrow f(t)$
- **Tractable model**: We are able to express equilibrium with parameters and study local stability $\frac{dN}{dt} \rightarrow N^* = f(parameters)$ => we can say generalities about the long term dynamics
- **Intractable models**:
  - we are obliged to do simulations using numerical algorithm or integration to approximate the dynamics step by step.
  - the results depend on the parameter values and initial conditions.
  - The dimension of the parameter space to explore is exponentially linked to the number of parameters
  - Simulations allow to investigate transient dynamics

4.3.2 Rules versus Maths

- **What is an agent-based model?** Algorithm which represent processes with a series of rules applied to each agent at each time step using proba: example with birth, death, interactions
- **What is the alternative**: Math => equations: example of ODE, SDE, DE, PDE: what does it mean?
  - ODE: change over time of our variable of interest (SDE; with some stochasticity) => give
  - DE: State at the next time step
  - PDE: Change over time and space
  - Overall idea: Balance between inputs and outputs. See with exponential model?
- **Advantages / Disadvantages**:
  - *ABM*:
    * The dynamics emerge from elemental processes
    * Simpler to build from empirical knowledge
    * But high computation consumption
    * Rarely tractable
  - *Equations*
    * Use of math and approximation for simplifications
    * Have large analysis power for extreme cases
    * Fast computation
    * Easier to fit to data
    * The relations between variables are imposed: processes synthetized
- **To what questions / system/ conditions is it most adapted?**
  - *ABM*:
    * When stochastic processes are dominant (ex: conservation of small pops)
    * When there are not too many parameters
    * For some questions where processes are difficult to synthesise (behaviour)

        * When there is no math skills around!
    - *Equations*:
        * Whenever processes can be synthesised with average parameters
        * => Large populations
- Such choices are often an inheritance from lab practices and habits.

4.4 Some classical models used in ecology and seen next days

- Systems of differential equations, no space :
    - Verhulst (logistic growth), Lotka-Volterra predator-prey (Day 2)
    - Lotka-Volterra, food web niche model (Day 4)
- Spatial systems of differential equations: TIB, Levins' occupancy (Day3)
- Spatial IBM: Neutral model of biodiversity from Hubbel (Day 3)

A Rmarkdown file is available to provide the code of the illustrations of section 2.2

Chapter 5

How to build a model? (Isa)

(1h) Lecture with interactivity (and Rmarkdown file for some parts)

- I will present different pictures of different plant-herbivore systems with associated research questions and ask how they will sketch the system (on board) to do a mode.
- I will ask then which formalism they would choose and why (in terms of determinism, time, space) [and we can try writing rules or equations?]
- Then we will focus on a question for which the Rosenzweig-MacArthur model (ODE) is appropriate, I will write the model on board and ask all the assumptions they could see that are behind the model formulation. We will discuss in which regards they are appropriate for our question.
- Then we will study line by line a code in R to explain how this could be implemented in R, while also explaining the principle of numerical integration.
- I will let them 5-10min to run the model and plot some dynamics and ask them how they would use it to answer the question (to introduce the next section).

Chapter 6

How to analyze a model? (Isa)

(1h) Interactive lecture + based on a script for some parts

(Rmarkdown file with the example of Rosenzweig-MacArthur model)

- Analyze the behavior of the model long-term (Box 3 Grainger et al Am-Nat)
  - Local stability (when possible for long term dynamics) (script)
    - Calculating equilibria
    - What are the Jacobian matrix and eigenvalues (meaning but no math derivation) ?
    - How to interpret the eigenvalues to infer the long term behavior type
- Isoclines / graphical *Phase plane (script)
- Bifurcation diagrams (long term dynamics) (script)
- When results depend on initial conditions => finding all the equilibria of intractable ODE systems (loop to run large series of initial conditions, function searchZeros of package nleqslv)
- Model Usages to answer the question (will try to show some concrete examples of questions to answer with these usages for the model of section 3a)
  - Parameter variation (see bifurcation diagrams + example DeMaz Grazing-Optimization)
  - Comparison with a null model to assess the impact of a process (with or without the process, with different formulation of the process. Example with functional responses (script)

    - Generation of synthetic data on which to run in silico experiments. - Example of food webs under different regimes of perturbations.

- Parameter exploration and robustness of conclusions
    - Exhaustivity is possible (tractable)
    - We know the values of parameters from empirical data => allows us to fix or restrain the range of some parameters.
    - Sensitivity analysis: effect size when varying 10% each parameter + look if conclusions are modified if the most sensitive parameters are varied.

Chapter 7

Temporal data (Day 2)

7.1 Introduction

7.2 Conclusion

Chapter 8

Spatial data (Day 3)

8.1 Introduction

Comapre to Day 2, we will increase scale to consider

Definition of a community and examples

The typical spatial dataset: a community matrix (sites*species), with either presence/absence, or detection/non detection, or abundance, or presence only, data We might also have additional data: species traits, environmental variables, phylogenetic relationships

The typical questions we want to address

8.2 Theoretical frameworks toattack these issues

We present three general frameworks used in this context: – The habitat filtering approach – The metacommunity framework – Vellend's community ecology theory framework We list strengths/advantages for each framework We point to this cool website for reflections from Vellend and Leibold on the fundamental publications introducing frameworks (2) and (3)

8.3 DATA-DRIVEN (STATISTICAL) APPROACHES

8.3.1 Community level methods

Unconstrained do not directly use environmental variables: a posteriori associations (PCA…) Constrained methods simultaneously consider species composition and environmental descriptors (RDA…) Short focus on Cottenie's approach of partitioning variance (partialRDA + PCNM) with examples from mollusk communities in the French Antilles ### Species level methods Permutation approaches, the Cscore, etc.. ==> make a connection with Day 4 here (networks and permutations) We mention the fact that even without interactions, several processes may yield spurious species associations (e.g. Calcagno et al. 2022) Species Distribution Models Occupancy models: their hallmark is to add a detection layer (probability to detect a species when it is present). We will talk more about this in the next section (mechanistic approaches)

8.4 PROCESS-DRIVEN (MECHANISTIC) APPROACHES

8.4.1 Patch-occupancy models

We start by introducing the two simplest and historically important models! Levins (1969) and McArthur & Wilson (1967) models. We highlight the mathematical connection between Levins' model and the logistic model seen on the previous Day There is an intrinsic covariation of r and K in this formulation! We compute the equilibrium occupancies in the two models We introduce the MW model of island biogeography: why are the curves non linear? We make them reflect on this, and ton the role of species trait variation in (c,e) We present the example of fitting this model to island presence/absence data: data from Manne et al. (1998) J. Anim. Ecol.

We then discuss possible extensions: – explicit space (IFM models from Hanski and followers) – competition/colonization trade-offs, spatial networks, trophic webs…

Models with explicit abundance dynamics in patches These models quickly become complicated and need to be simulated. One exception is Hubbel's neutral model, which has a dynamics for species abundances in patches and remains reasonably tractable One example of simulation framework for metacommunities is Thompson et al. (2020) Ecology Letters. This is the one we will use to simulate data in the afternoon practical.

8.5 Conclusion

Chapter 9

Networks (Day 4)

9.1 Introduction

9.2 Conclusion

Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL http://yihui.org/knitr/. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2023. URL https://github.com/rstudio/bookdown. R package version 0.37, https://pkgs.rstudio.com/bookdown/.