John M. Zobitz

Exploring Modeling with Data and Differential Equations Using R

To my parents Joan and Francis,

who gave me strong roots from which to grow.

To my wife Shannon,

who provides me the support to keep my trunk from breaking.

To my children Colin, Grant, Phoebe,

who are the branches that support the emerald green leaves.

Kiitos. Amor a todos.

Contents

List of Tables

List of Figures

Welcome

This book is written for you: the student learning about modeling and differential equations. Perhaps you first encountered models, differential equations, and better yet, building plausible models from data in your Calculus course.

This book sits "at the intersection" of several different mathematics courses: differential equations, linear algebra, statistics, calculus, data science - as well as the partner disciplines of biology, chemsitry, physics, business, and economics. An important idea is one of *transference* where a differential equation model applied in once context can also be applied (perhaps with different variable names) in a separate context.

I intentionally emphasize models from biology and the environmental sciences, but throughout the text you can find examples from the other disciplines. I hope you see the connections of this content to your own intended major.

This book is divided into 4 parts:

1. Models with differential equations
2. Parameterizing models with data
3. Stability analysis for differential equations.
4. Stochastic differential equations

You may notice the interwoven structure for this book: models are introduced first, followed by data analysis and parameter estimation, returning back to analyzing models, and ending with simulating random (stochastic) models.

Unsure what about all these topics mean? Do not worry! The topics are presented with a "modeling first" paradigm that first introduces models, and equally important, how data are used to inform a model. This "conversation" between models and data are important to help build plausibility and confidence in a model. Stability analysis helps to solidify the connection between models and parameters (which may change the underlying dynamical stability). Finally the notion of *randomness* is extended with the introduction of stochastic differential equations.

Unifying all of these approaches is the idea of developing workflows for analysis, visualization results, and interpreting any results in the context of the problem.

Computational code

This book makes heavy use of the `R` programming language, and unabashedly develops programming principles using the `tidyverse` syntax and programming approach. This is intentional to facilitate direct connections to courses in introductory data science or data visualization. Throughout my years learning (and teaching) different programming languages I have found `R` to be the most versatile and adaptable. The `tidyverse` syntax, in my opinion, is transformed my own thinking about sustainable computation and modeling processes - and I hope it does for you as well.

There is a companion `R` package available called `demodelr` to run programs and functions in the text. Instructions to install this package are given in Section **??**. The minimum version of R Version 4.0.2 (2020-06-22) (R Core Team 2021) and RStudio is Version 1.4.1717 (RStudio Team 2020)

The `demodelr` package uses the following `R` packages:

- `tidyverse` (and the associated packages) (Version 1.3.1) (Wickham et al. 2019)
- `GGally` (Version 2.1.2) (Schloerke et al. 2021)
- `formula.tools` (Version 1.7.1) (Brown 2018)
- `expm` (Version 0.999-6) (Goulet et al. 2021)

Questions? Comments? Issues?

Feel free to file an issue with the `demodelr` package to my github[1]

About the cover

The cover art was taken by Shannon Zobitz during a hike at Orinoro Gorge[2] in Finland. The photo is indicative of several things: (1) the journey ahead as you commence learning about modeling, differential equations, and `R`, (2) the occasional roots in the path that may cause you to stumble (such as coding errors). Everyone makes them, so you are in good company. (3) the yellow

---

[1]https://github.com/jmzobitz/ModelingWithR/issues
[2]https://visitleppavirta.fi/en/service/orinoro-gorge

markings on the trees indicate the way forward. May this textbook be the guide for you as you progress over the hill and onward. Let's get started!

Acknowledgments

This book has been developed over the course of several years in a variety of places: two continents, between meetings, in the early mornings, at coffee shops, or while waiting for practices to end. Special thanks are to the following:

- **Augsburg University:** You have been my professional home for over a decade and given me the space and support to be intellectually creative in my teaching and scholarship. Special thanks to my Mathematics, Statistics, and Computer Science Department colleagues - it is a joy to work with all of you.

- **Augsburg University students:** Thank you for your interest and engagement in this topic, allowing me to test ideas in an upper division course titled (wait for it ...) *Modeling and Differential Equations in the Biological and Natural Sciences.* While the course title is a mouthful, you provided concise, honest, and insightful feedback, shaping this text. I am forever indebted to you.

- **My family:** Shannon, Colin, Grant, and Phoebe for humoring me (and my occasional grumpiness) while this project has been completed.

1

Modeling With Rates of Change

Section **??** provided examples for modeling with rates of change, and Section **??** introduced the computational and visualization software `R` and `RStudio`. how we can translate equations with rates of change to understand phenomena. The focus for this section will be on taking a contextual description and starting to develop differential equation models for them.

Oftentimes when we construct differential equations from a contextual description we bring our own understanding and knowledge to this situation. How *you* may write down the differential equation may be different from someone else - *do not worry!* This is the fun part of modeling: models can be considered testable hypotheses that can be refined when confronted with data. Let's get started

1.1 Competing plant species and equilibrium solutions

Consider the following context to develop a mathematical model:

> A newly introduced plant species is introduced to a region. It competes with another established species for nutrients (and is a better competitor). However, the growth rate of the new species is proportional to the difference between the current number of established species and the number of new species. You may assume that the number of established species is a constant E.

For this problem we will start by naming our variables. Let N represent number of new species and E the number of established species. We will break this down accordingly:

- *"the growth rate of the new species"* describes the rate of change, or derivative, expressed as $\dfrac{dN}{dt}$.
- *"is proportional to the difference between the current number of established species and the number of new species"* means $\alpha \cdot (E - N)$, where α is the proportionality constant. Including this parameter helps to avoid assuming we have a 1:1 correspondence between the growth rate of the new species and the population difference.
- *"and is a better competitor"* helps to explain why the term is $\alpha \cdot (E - N)$ instead of $\alpha \cdot (N - E)$. We know that the newly established species will start out in much smaller numbers than N. But since it is a better competitor, we would expect its rate to increase initially. So $\dfrac{dN}{dt}$ should be *positive* rather than negative.

Taking all these assumptions together, Equation (1.1) shows the differential equation to model this context:

$$\frac{dN}{dt} = \alpha \cdot (E - N) \tag{1.1}$$

You may recognize that Equation (1.1) is similar to Equation (**??**) in Section **??** for the spread of Ebola. It is not surprising to have similar differential equations appear in different contexts. We will see throughout this book that it is more advantageous to learn techniques to analyze models qualitatively rather than memorize several different types of models and not see the connections between them.

An interesting solution to a differential equation is the *steady state* or *equilibrium solution*. Equilibrium solutions occur where the rates of change are zero. For Equation (1.1), this means that we are solving $\dfrac{dE}{dt} = \alpha \cdot (E - N) = 0$. Granted, the expression $\alpha \cdot (E - N)$ may look like alphabet soup, it is helpful to remember that α and E are both parameters; the steady state occurs when the expression $E - N$ equals zero, or when $N = E$. We may consider the new species N being established when it reaches the level population level as E. Identifying steady states in a model aids in understanding the behavior of any solutions for a differential equation. Sections **??** and **??** dig deeper into steady states and their calculation.

1.2 The Law of Mass Action

Our next example focuses on how to generate a model that borrows concepts from modeling chemical reactions. For example let's say you have a substrate

A that reactions with enzyme *B* to form a product *S*. One common way to represent this process is with a reaction equation (Equation (1.2)):
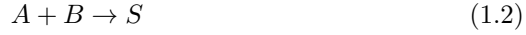
$$A + B \rightarrow S \tag{1.2}$$

Figure 1.1 is a schematic diagram of Equation (1.2):
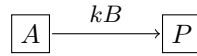
$$A \xrightarrow{kB} P$$

FIGURE 1.1 Schematic diagram of a substrate-enzyme reaction.

One key quantity is the rate of formation for the product *P*, which we express by Equation (1.3):

$$\frac{dP}{dt} = kAB, \tag{1.3}$$

where *k* is the proportionality constant or the rate constant associated with the reaction. Notice how we express the interaction between *A* and *B* as a product - if either the substrate *A* or enzyme *B* is not present (i.e. *A* or *B* equal 0), then product *P* is not formed. Equation (1.3) is an example of the law of mass action.

Modeling interactions (whether between susceptible and infected individuals, enzymes and substrates, or predators and prey) with the law of mass action is always a good first assumption to understand the system, which can be subsequently refined. For example, if we consider that the substrate might decay, we can revise Figure 1.1 to Figure 1.2:
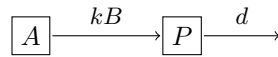
$$A \xrightarrow{kB} P \xrightarrow{d}$$

FIGURE 1.2 Revised schematic diagram of substrate-enzyme reaction with deacy of the product *P*.

In this instance the rate of change of *P* would then include a term *dP* (Equation (1.4)):

$$\frac{dP}{dt} = kAB - dP, \tag{1.4}$$

1.3 Coupled differential equations: lynx and hares

Another example is a *system of differential equations*. The context is between the snowshoe hare and the Canadian lynx, shown in Figure 1.3. Figure 1.4 also displays a timeseries of the two populations overlaid. Notice how in Figure 1.4 both populations show regular periodic fluctuations. One plausible reason is that the lynx prey on the snowshoe hares, which causes the population to initially decline. Once the snowshoe hare population declines, then there is less food for the lynx to survive, so their population declines. The decline in the lynx population causes the hare population to increase, and the cycle repeats.[1]

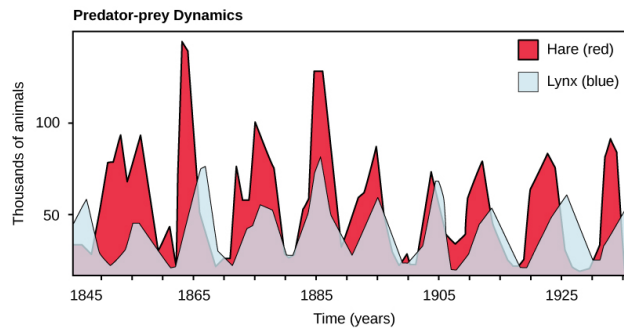**FIGURE 1.3** Examples of lynx and hare - aren't they beautiful?

FIGURE 1.4 Timeseries of the combined lynx and hare populations. Notice how the populations are coupled with each other.

In summary it is safe to say that the two populations are *coupled* to one an-

---

[1]There is a lot more nuance for reasons behind periodic fluctuations in these two populations, which includes more complicated food web interactions and climate variation. MacLulich (1937), Stenseth et al. (1997), and King and Schaffer (2001) are good places to dig into the complexity of this fascinating biological system. Image sources for Figure **??**: Kilby (2012) and Frank, Jacob W. (2021). Image source for Figure 1.4: OpenStax (2016)

other, yielding a coupled system of equations. But in order to understand how they are coupled together, first let's consider the two populations *separately.*

To develop the mathematical model we will make some simplifying assumptions. The hares grow much more quickly than then lynx - in fact some hares have been known to reproduce several times a year. A reasonable assumption for large hare populations is that rate of change of the hares is proportional to the hare population. Based on this assumption Equation (1.5) describes the rate of change of the hare population, with H is the population of the hares:

$$\frac{dH}{dt} = rH \tag{1.5}$$

Since the growth rate r is positive, so then the rate of change (H') will be positive as well, and H will be increasing. A representative value for r is 0.5 year^{-1} Brady and Butler (2021). You may be thinking that the units on r seem odd - (year^{-1}), but that unit on r makes the term rH dimensionally consistent to be a rate of change.

Let's consider the lynx now. A approach is to assume their population declines exponentially, or changes at the rate proportional to the current population. Let's consider L to be the lynx population, with the following differential equation (Equation (1.6)):

$$\frac{dL}{dt} = -dL \tag{1.6}$$

We assume the death rate d in Equation (1.6) is positive, leading to a negative rate of change for the Lynx population (and a decreasing value for L). A typical values of d is 0.9 yr^{-1} Brady and Butler (2021).

The next part to consider is how the lynx and hare interact. Since the hares are prey for the lynx, when the lynx hunt, the hare population decreases. We can represent the process of hunting with the following adjustment to our hare equation:

$$\frac{dH}{dt} = rH - bHL \tag{1.7}$$

So the parameter b represents the hunting rate. Notice how we have the term HL for this interaction. This term injects a sense of realism: if the lynx are not present ($L = 0$), then the hare population can't decrease due to hunting. We say that the *interaction* between the hares and the lynx with multiplication. A typical value for b is .024 lynx^{-1} year^{-1}. It is okay if that unit seems a little odd to you - it should be! As before, if we multiply out the units on bHL we would get units of hares per year.

How does hunting affect the lynx population? One possibility is that it increases the lynx population:

$$\frac{dL}{dt} = bHL - dL \tag{1.8}$$

Notice the symmetry between the rate of change for the hares and the lynx equations. In many cases this makes sense - if you subtract a rate from one population, then that rate should be added to the receiving population. You could also argue that there is some efficiency loss in converting the hares to lynx - not all of the hare is converted into lynx biomass. In this situation we sometimes like to adjust the hunting term for the lynx equation with another parameter e, representing the efficiency that hares are converted into lynx:

$$\frac{dL}{dt} = e\,bHL - dL \tag{1.9}$$

(sometimes people just make a new parameter $c = e\,b$, but for now we will just leave it as is and set $e = 0.2$). Equation (1.10) shows the coupled system of differential equations:

$$\begin{aligned} \frac{dH}{dt} &= rH - bHL \\ \frac{dL}{dt} &= e\,bHL - dL \end{aligned} \tag{1.10}$$

The schematic diagram representing these interactions is the following is shown in Figure 1.5:
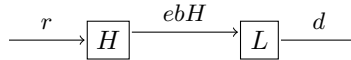

FIGURE 1.5 Schematic diagram Lynx-Hare system.

Equation (1.10) is a classical model in mathematical biology and differential equations - it is called the *predator prey* model, also known as the *Lotka-Volterra model* Volterra (1926).

1.4 Functional responses

In several examples we have seen a rate of change proportional to the current population, as in the rate of growth of the hare population is rH. This is one example of what we would call a functional response[2]. Another type of functional response assumes that the rate reaches a limiting value proportional

---

[2]https://en.wikipedia.org/wiki/Functional_response

to the population size, so $\dfrac{dH}{dt} = \dfrac{rH}{1 + arH}$. This is an example of a **type II functional response**. Finally, the type II response has also been generalized (a **type III functional response**) $\dfrac{dH}{dt} = \dfrac{rH^2}{1 + arH^2}$. Figure 1.6 shows all three functional responses together:
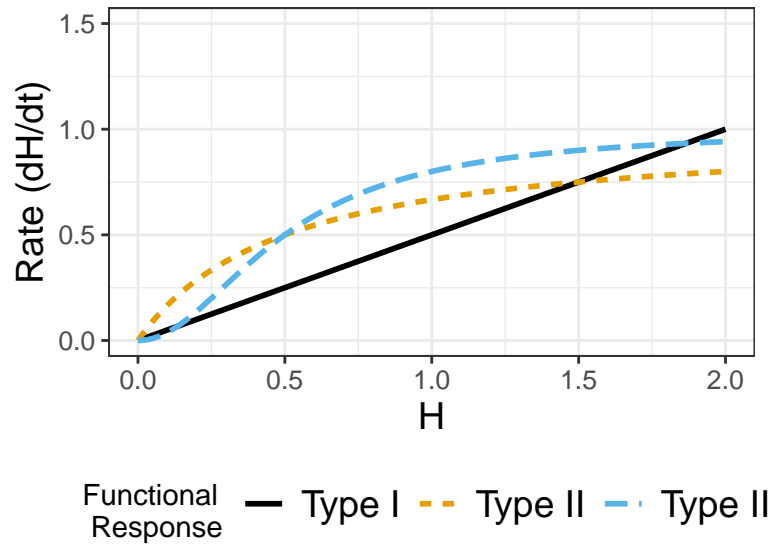

FIGURE 1.6 Comparison between examples of Type I - Type III functional responses. For a Type I functional response the Rate grows proportional to population size H, whereas for Types II and III the rate reaches a saturating value.

Notice the limiting behavior in the Type II and Type III functional responses. These responses are commonly used in ecology and predator-prey dynamics and in problems of how animals search for food.

1.5 Exercises

Exercise 1.1. Consider the following type of functional responses:

$$\text{Type I: } \frac{dP}{dt} = 0.1P$$

$$\text{Type II: } \frac{dP}{dt} = \frac{0.1P}{1 + .03P} \qquad (1.11)$$

$$\text{Type III: } \frac{dP}{dt} = \frac{0.1P^2}{1 + .05P^2}$$

For each of the functional responses evaluate $\lim\limits_{P \to \infty} \dfrac{dP}{dt}$. Since these functional responses represent a rate of change of a population, what are some examples (hypothetical or actual) would each of these responses be appropriate?

Exercise 1.2. A population grows according to the equation:

$$\frac{dP}{dt} = \frac{0.1P}{1 + .05P} - P = f(P) - g(P) \qquad (1.12)$$

a. On the same axis, plot the equations $f(P)$ and $g(P)$. What are the two positive values of P where $f(P)$ and $g(P)$ intersect?
b. Next algebraically determine the two steady state values of P, that is solve $\dfrac{dP}{dt} = 0$ for P. (*Hint:* factor a P out of the expression $f(P) - g(P)$.)
c. Does your algebraic solution match your graphical solutions?

Exercise 1.3. A population grows according to the equation:

$$\frac{dP}{dt} = 2P - \frac{4P^2}{1 + P^2} = r(P) - d(P) \qquad (1.13)$$

a. On the same axis, plot the equations $r(P)$ and $d(P)$. What are the two positive values of P where $r(P)$ and $d(P)$ intersect?
b. Next algebraically determine the two steady state values of P, that is solve $\dfrac{dP}{dt} = 0$ for P. (*Hint:* factor a P out of the expression $r(P) - d(P)$.)
c. Does your algebraic solution match your graphical solutions?

Exercise 1.4. A population grows according to the equation:

$$\frac{dP}{dt} = \frac{aP}{1 + abP} - dP, \qquad (1.14)$$

where a, b and d are all positive parameters. Determine the two steady state values of P, that is solve $\dfrac{dP}{dt} = 0$ for P.

Exercise 1.5. A chemical reaction takes two chemicals X and Y to form a substrate Z through the law of mass action. However the substrate can also disassociate. The reaction schematic is the following:

$$X + Y \rightleftharpoons Z, \tag{1.15}$$

where you may define a the proportionality constant k_+ is associated with the formation of the substrate Z and k_- the disassociation (Z decays back to X and Y).

Write down a differential equation that represents the rate of reaction $\dfrac{dZ}{dt}$.

Exercise 1.6. (Inspired from Thornley and Johnson (1990) and Logan and Wolesensky (2009)) For each of the following exercises consider the following contextual situations modeling rates of change. For each problem you will need to:

- Name and describe all variables and parameters.
- Determine a differential equation representing the context.
- Write a brief one-two sentence explanation of why your differential equation models the situation at hand.
- Hand sketch a rough graph of what you think the solution as a function of time, consistent with the context given.

a. The rate of change of an animal's body temperature is proportional to the difference in temperature between the environment.
b. A plant grows propritional to its current length L. Assume this proportionality constant is μ, whose rate also decreases proportional to its current value. You will need to write down a system of two equation with variables L and μ.
c. A patient undergoing chemotherapy receives an injection at rate I. This injection decreases the rate that a tumor accumulates mass. Independent of the injection, the tumor accumulates mass at a rate proportional to the mass of the tumor.
d. A cell with radius r assimilates nutrients at a rate proportional to its surface area, but uses nutrients proportional to its volume. Determine an equation that represents the rate of change of the radius.
e. A patient undergoing chemotherapy receives an injection at rate I. This injection decreases the rate that a tumor accumulates mass. Independent of the injection, the tumor accumulates mass at a rate proportional to the mass of the tumor.
f. The rate that a cancer cell divides (increases in amount) is proportional to the amount of healthy cells in its surrounding environment. You may assume that a healthy cell has a mortality δ_H and a cancer cell has mortality δ_C. Be sure to write down a system of differential equations for the population of cancer cells C and healthy cells H.
g. The rate that a virus is spread to the population is proportional to the

probability that a person is sick (out of N total sick and healthy individuals).
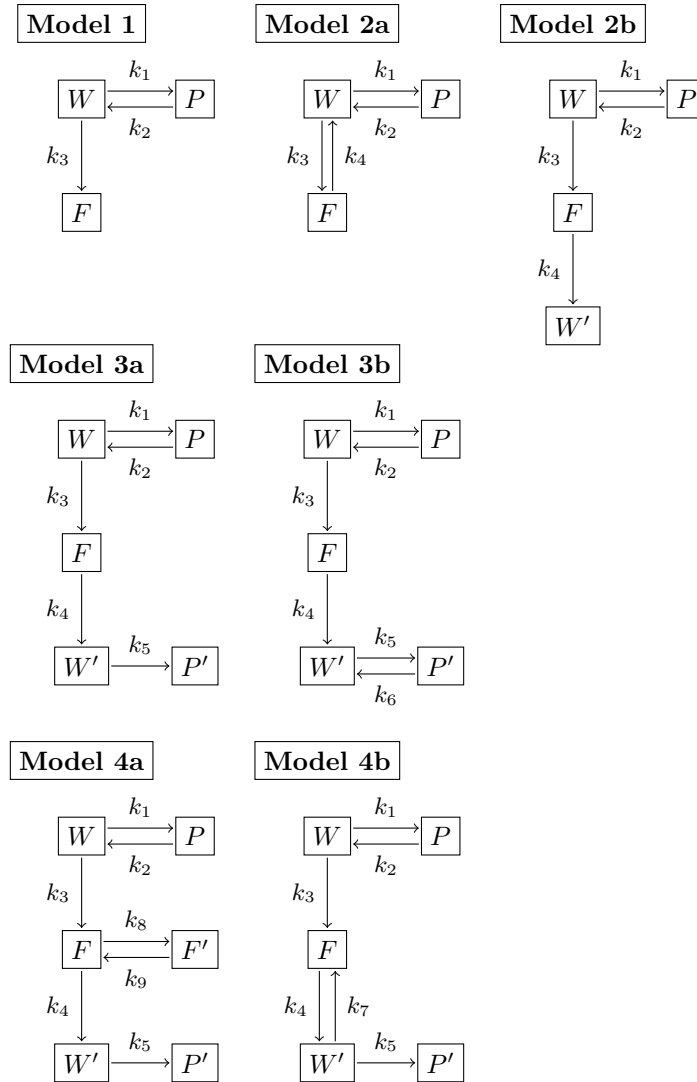

FIGURE 1.7 Modeled reaction schemes representing the potential effect of a pesticide on water quality.

Exercise 1.7. (Inspired by Burnham and Anderson (2002)) You are tasked with the job of investigating the effect of a pesticide on water quality, in terms of its effects on the health of the plants and fish in the ecosystem. Different

models can be created that investigate the effect of the pesticide. Different types of reaction schemes for this system are shown in Figure 1.7, where F represents the amount of pesticide in the fish, W the amount of pesticide in the water, and S the amount of pesticide in the soil. The prime (e.g. F', W', and S' represent other bound forms of the respective state). In all seven different models can be derived. For each of the model schematics, apply the Law of Mass Action to write down a system of differential equations.

References

Brady, R, and J Butler. 2021. "The Circle of Life: The Mathematics of Predator-Prey Relationships." *Frontiers for Young Minds* 9 (651131). https://kids.frontiersin.org/articles/10.3389/frym.2021.651131.

Brown, Christopher. 2018. *Formula.tools: Programmatic Utilities for Manipulating Formulas, Expressions, Calls, Assignments and Other r Objects.* https://github.com/decisionpatterns/formula.tools.

Burnham, Kenneth P., and David R. Anderson. 2002. *Model Selection and Multimodel Inference.* New York, NY: Springer New York.

Frank, Jacob W. 2021. "Snowshoe Hare." https://commons.wikimedia.org/wiki/File:Snowshoe_Hare_(6187109754).jpg.

Goulet, Vincent, Christophe Dutang, Martin Maechler, David Firth, Marina Shapira, and Michael Stadelmann. 2021. *Expm: Matrix Exponential, Log, Etc.* http://R-Forge.R-project.org/projects/expm/.

Kilby, Eric. 2012. "Canada Lynx." https://commons.wikimedia.org/wiki/File:Canada_Lynx_(8154273321).jpg.

King, Aaron A., and William M. Schaffer. 2001. "The Geometry of a Population Cycle: A Mechanistic Model of Snowshoe Hare Demography." *Ecology* 82 (3): 814–30. https://doi.org/10.2307/2680200.

Logan, J. David, and William Wolesensky. 2009. *Mathematical Methods in Biology.* 1st ed. Hoboken, N.J: Wiley.

Lotka, Alfred J. 1920. "Analytical Note on Certain Rhythmic Relations in Organic Systems." *Proceedings of the National Academy of Science* 6 (7): 410–15.

———. 1926. "Elements of Physical Biology." *Science Progress in the Twentieth Century (1919-1933)* 21 (82): 341–43.

MacLulich, D. A. 1937. *Fluctuations in the Numbers of the Varying Hare (Lepus Americanus).* Toronto, Canada: University of Toronto Press. https://doi.org/10.3138/9781487583064.

Mahaffy, Joseph. 2010. "Lotka-Volterra Models." https://jmahaffy.sdsu.edu/courses/f09/math636/lectures/lotka/qualde2.html.

OpenStax, C. N. X. 2016. `https://commons.wikimedia.org/wiki/File:Figure/_45/_06/_01.jpg`.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. `https://www.R-project.org/`.

RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Manual. Boston, MA: RStudio, PBC. `http://www.rstudio.com/`.

Schloerke, Barret, Di Cook, Joseph Larmarange, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg, and Jason Crowley. 2021. *GGally: Extension to Ggplot2*. `https://CRAN.R-project.org/package=GGally`.

Stenseth, Nils Chr, Wilhelm Falck, Ottar N. Bjørnstad, and Charles J. Krebs. 1997. "Population Regulation in Snowshoe Hare and Canadian Lynx: Asymmetric Food Web Configurations Between Hare and Lynx." *Proceedings of the National Academy of Sciences* 94 (10): 5147–52. `https://doi.org/10.1073/pnas.94.10.5147`.

Thornley, John H M, and Ian R Johnson. 1990. *Plant and Crop Modelling: A Mathematical Approach to Plant and Crop Physiology*. Caldwell, New Jersey: Blackburn Press.

Volterra, Vito. 1926. "Fluctuations in the Abundance of a Species Considered Mathematically." *Nature* 118 (2972): 558–60. `https://doi.org/10.1038/118558a0`.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. `https://doi.org/10.21105/joss.01686`.

Index