# A primer of community ecology using the $R$ language 

Kohmei Kadowaki ${ }^{1,2}$

${ }^{1}$ The Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan
${ }^{2}$ Graduate School of Agriculture, Kyoto University, Kyoto, Japan

## Correspondence

Kohmei Kadowaki, The Hakubi Center for Advanced Research, Kyoto University, Kyoto 606-8501, Japan.
Email: kinokomushi@gmail.com

## Funding information

KAKENHI, Grant/Award Number: 21H02233; The Hakubi Project, Kyoto University


#### Abstract

Community ecology beginners often struggle to understand theories expressed in complex mathematical formulas and to master computer programming. To remedy this situation, this article provides a practical, R-based introduction to community ecology by illustrating core concepts (vital rates, carrying capacity, and density dependence) and models that can be used to explore the patterns of species abundance and diversity. The structure of this article consists of three modeling exercises, each asking a general question that can be answered by a combination of theory and R programming: (1) what determines the abundance of species, and what makes a population persist and go extinct?; (2) what determines the distribution of species and species diversity?; (3) what determines the relative abundance of species and what allows species to coexist? Through the exercises, I discuss the following five concepts and ideas that provide valuable insights into the questions: (i) the tragedy of the commons, (ii) the theory of island biogeography, (iii) competitive exclusion, (iv) the neutral theory of biodiversity, and (v) frequency dependence. These materials are thus designed to guide the reader in developing an intuition for ecological thinking that will help capture the essence of the global environmental and biodiversity crisis. Although this article does not delineate the scope and depth of the vast field of community ecology, I hope that it will motivate the reader to step up to a more formal introduction to community ecology.


## KEYWORDS

density dependence, frequency dependence, the neutral theory, the theory of island biogeography, the tragedy of the commons

## 1 | INTRODUCTION

It is often said that studying community ecology is challenging (Mittelbach \& McGill, 2019; Vellend, 2017). There are probably at least two main reasons for this:

[^0]first, the various concepts of community ecology have not always been systematically organized in textbooks (Vellend, 2017), making theory-based learning challenging to complete. Second, understanding community research requires concerted efforts, including natural history, fieldwork, laboratory work, statistics, theoretical models, and computer programming. This suggests that performing modern community ecology research using

[^1]any single technique is nearly impossible. Among them, understanding theories/concepts and mastering computer programming are the two major requisites of community ecology research, yet surprisingly, these two have often been taught independently in the classroom. I believe that learning the two at the same time can create synergy and deepen understanding, as I will demonstrate below.

I demonstrate that one of the most effective ways to learn about community ecology is to do hands-on activities guided by ecological theories, rather than simply focusing on theory-based learning. Doing so allows students to immerse themselves in theoretical ideas expressed in mathematical formulas, particularly by breaking down a complex ecological theory into its component parts and assembling them in the R language to gain a deeper understanding of the nature of the ecological theory. Thus, this article is designed to provide an opportunity mainly for undergraduate and postgraduate students interested in studying community ecology to take their very first steps. With examples closely related to the fundamental concepts of community ecology, it is intended that the reader will have an understanding of the fundamental ecological concepts and theories as well as some basic skills of R computer programming.

The structure of this article consists of three modeling exercises, each asking a general question that can be answered using a combination of theory and $R$ programming.

- What determines the abundance of species? And what makes a population of species persist and go extinct?
- What determines the distribution of species and species diversity?
- What determines the relative abundance of species and what allows them to coexist?

In considering these questions, the following five concepts and ideas are discussed: (i) the tragedy of the commons, (ii) the theory of island biogeography, (iii) competitive exclusion, (iv) the neutral theory of biodiversity, and (v) frequency-dependence driven by plant-soil feedback. These five concepts were chosen for several reasons. First, they provide valuable insights into the three questions above. Next, they work well and spark curiosity about community ecology in the classroom. Finally, they are profound and broadly applicable concepts in practice.

## 2 | PREPARATION

Before getting to the main part, a few preparations and a computer are needed. When working on this article, I
would like the reader to have a notepad, a pen, and a computer with R installed ( R Core Team, 2023). To execute the $R$ code in this article, readers may install $R$ on the computer with the "tidyverse" package (Wickham et al., 2019). All the plots created in this article can be reproduced using Supporting Information "primerofCommEcol.R." To play with graphics and simulations, one might get some custom-made functions from "Rfunctions_primerofCommEcol.R" and save the file on the working directory.

```
#install package
install.packages("tidyverse")
#load library
library(tidyverse)
#load R code downloaded from Supplementary File
source("Rfunctions_primerofCommEcol.R")
```


## 3 | MODELING POPULATION ABUNDANCES: THE TRAGEDY OF THE COMMONS

Ecology, the study of biodiversity and the environment, is often seen as a discipline for people who study their favorite organisms/ecosystems. However, it is a discipline that can make a direct and long-term contribution to environmental issues, with practical applications in resource management and conservation. Local population extinction, which refers to the loss of a species from the local area, is a major concern in ecology and conservation biology due to its far-reaching effects on ecosystems. In this section, I examine overfishing (the extinction of a fish population) as an example for intuitively understanding how human use of natural resources causes population extinction, so that the reader will see this exercise as a significant and relevant problem.

The "tragedy of the commons" (Hardin, 1968) provides a valuable context in which to study the problem of overfishing. The tragedy of the commons is invoked when a group of people shares a common finite resource; when each member of the group acts independently and uses the resource for their self-interest, the group as a whole ends up with a tragic depletion of the resource, even if it is the consequence that none of them intended. The example below is borrowed in part from the fish game (http://eeinwisconsin.org/content/eewi/ 100323/FishGame.pdf, accessed January 24, 2023).


FIGURE 1 A fishing game to illustrate the tragedy of the commons can be played as follows: There are 20 fish in the pond (initial condition), and the simulation is played over 10 time steps. Prior to the game, three villagers independently choose one of three methods: (i) common fishing rod (the villager can catch one fish per time step); (ii) high-end fishing rod: the villager can catch two fish per time step; and (iii) trawling boat: the villager can catch three fish per time step. After each time step, villagers visit the pond and catch fish according to the selected fishing method; then, the uncaught fraction of fish reproduces at a rate of $25 \%$. [Color figure can be viewed at wileyonlinelibrary.com]

## 3.1 | The tragedy of the commons: Fixed catching scenario

I present a simulation of overfishing using R. The goal of the simulation is to see how the total number of fish changes over time and how fishing pressure affects fish stock depletion. Consider that three villagers A, B, and C manage a single fish stock pond and share the pond for food and business (Figure 1). The fish is a finite resource; suppose that the stock can contain up to 20 individuals of fish (i.e., carrying capacity) due to the limitation of food/ habitat for fish, and the fish population grows by $25 \%$ per time step (i.e., intrinsic rate of population increase), ignoring immigration, sex, and other biological details. Prior to the fishing, each villager decides which of the three fishing methods to use. The three fishing methods are: (i) common fishing rod (the villager can catch one fish per time step); (ii) high-end fishing rod (the villager can catch two fish per time step); and (iii) trawling boat (the villager can catch three fish per time step; Figure 1). At every time step, each villager visits the pond and catches a fixed number of fish defined by the fishing method. The uncaught fraction of the fish population grows by $25 \%$ at every time step. Note that, in the classroom, the tragedy of the commons is usually simulated by playing a game with fish-shaped tiddlywinks.

Here, the simulation starts with initial abundance $N_{0}=20$, with parameters defined as $a=1$ (villager A uses a common fishing rod), $b=2$ ( B uses a high-end fishing rod), and $c=2$ ( C uses a high-end fishing rod). The R code below allows the simulation to run under this constant catching scenario and the balance table as
output is shown in Table 1. Figure 2a displays a trajectory of fish abundance under this setting, which eventually leads to extinction at the eighth time step.

```
#tragedy of the commons
NO = 20 # initial number of fish (stock)
a=1 #number of fish A catches every round
b=2 #number of fish B catches every round
c=2 #number of fish C catches every round
D = a + b + c #total catch by villagers
r=1.25 # intrinsicgrowth rate of fish
population
N = NO # number of fish
counter = 0 #countingtime steps
balance.table = data.frame(counter, a, b,
c,N) #to keep track of balance
while (N > D && counter < 11) {
    counter = counter +1
    #number of individuals after fishing
    N_tmp = N - a - b - C
    # fish reproduction
    N = min(round (N_tmp *r), 20)
    balance.table = rbind(balance.table ,
        c(counter, a, b, c, N) ) #make table
}
colnames(balance.table) = c("counter", "a",
"b", "c", "N") # name column names
balance.table
```

TABLE 1 A balance table (balance. table) of the fish game, where each of the three villagers catches a fixed number of fish ( $a=1, b=2$, and $c=2$ ) until the fish population goes extinct at round 8 (see also Figure 2a).

| Counter | $\boldsymbol{a}$ | $\boldsymbol{b}$ | $\boldsymbol{c}$ | No. of fish |
| :--- | :--- | :--- | :--- | :--- |
| 1 | 1 | 2 | 2 | 20 |
| 2 | 1 | 2 | 2 | 19 |
| 3 | 1 | 2 | 2 | 18 |
| 4 | 1 | 2 | 2 | 16 |
| 5 | 1 | 2 | 2 | 14 |
| 6 | 1 | 2 | 2 | 11 |
| 7 | 1 | 2 | 2 | 8 |
| 8 | 1 | 2 | 2 | 4 |

Note: The column "counter" represents the number of rounds (time steps) played until extinction. The columns " $a$," " $b$," and " $c$ " represent the number of fish caught at each round by the villagers A, B, and C, respectively.




where $r$ denotes intrinsic growth rate ( 1.25 in this case) and $D$ is defined as $D=a+b+c$. Therefore, per capita rate of change yields:

$$
\begin{aligned}
& \text { per capita rate of change } \\
& =\frac{N_{t+1}-N_{t}}{N_{t}}=(r-1)-\frac{r D}{N_{t}} \\
& =0.25-1.25 \frac{D}{N_{t}} .
\end{aligned}
$$ and under what conditions they occur, let us consider expressing the dynamics of fish populations in mathematical equations. The number of fish at time step $t+1$ $\left(N_{t+1}\right)$ can be written as:

$$
N_{t+1}=r\left(N_{t}-D\right),
$$

FIGURE 2 Numerical simulation of the tragedy of the commons, where the number of fish in a stock pond ( $N$, the number of individuals) is plotted against the time step, where $N_{0}=20, r=1.25$. (a) a simulation with $a=1, b=2, c=2$ (the number of catch per time step by the villagers A, B, and C, respectively). (b) Per capita rate of population change plotted as a function of population abundance ( $N_{t}$ ). The horizontal dotted line represents the per capita birth rate of the fish population (in this case, a constant of 0.25 ). The curve is given by the second term in Equation (2) and, the population declines when the per capita mortality curve exceeds the birth rate of 0.25 . For example, the population can be maintained at 20 individuals when $D=4$, but cannot be maintained if the population declines below that number, and is unsustainable at $D=5$, regardless of the size of the fish population (i.e., extinction is inevitable). (c) If the villagers agree to a density-dependent fishery, the fish population can be reduced to 15 individuals, but the population can be maintained by adjusting the catch in a density-dependent manner as the population declines. (d) The plot shows how the balance between the per capita birth (horizontal dotted line) and mortality rate (straight line) changes with fish population by varying the value of the density dependence factor $k$ from 0.01 to 0.03 .

To visualize this outcome under the constant catching scenario, one can plot Equation (2) using ggplot. Figure $2 b$ shows that the rate of change per individual is divided into the first term, the birth rate (dotted line), and the second term, the mortality rate (shown in different tones of gray color depending on $D$ ), and these two terms are plotted as a function of the number of individuals. For example, at $D=5$, the per capita death rate always exceeds the per capita birth rate, making extinction inevitable. In contrast, at $D=4$, the birth rate and death rate intersect and balance at 20 individuals. A key point in Figure 2b is that the smaller the population size $N_{t}$, the more likely the per capita mortality rate will exceed the per capita birth rate, thus increasing the likelihood of extinction. One might wonder "why didn't anyone notice before fish became extinct?" or "what did villagers say when they caught the last fish?" Yet, by the time they realize that the fish number is going to dip, it might be already too late. Thus, extinction occurs very easily when each villager catches a fixed number of fish and makes the independent decision of how many to catch. Theoretically, if villagers are independently allowed to choose which fishing method to use (Figure 1), the logical decision for each villager is to catch more fish, and every villager reaches the same conclusion. The result, however, is a tragedy: no fish can ultimately survive due to overfishing. Individual actions lead to the depletion of a common resource, even though the depletion is not in the best interest of the village as a whole (Hardin, 1968).

## 3.2 | Density-dependent fishing averts the tragedy of the commons

In Section 3.1, I mentioned that fish populations can easily go extinct, but this is the case when each villager independently catches a fixed number from a fish population. Humans have an amazing capacity to communicate verbally with each other and to solve problems, even with potentially conflicting opinions. The next scenario I consider is that three villagers discuss possible sustainable fishery and come up with an idea to regulate catch per time in a densitydependent manner such that the total catch ( $D$ ) can be updated with the density. Modeling $D$ as a function of $N$ with a density-dependent factor $k$, that is, $D=k N^{2}$ yields:

$$
\begin{equation*}
N_{t+1}=r\left(N_{t}-k N_{t}^{2}\right)=r N_{t}\left(1-k N_{t}\right) \tag{3}
\end{equation*}
$$

where $k$ represents the density dependence factor. Importantly, Equation (3) has a similar form as
the one commonly known as the (discrete-time) logistic growth model presented in ecology textbooks (Gotelli, 2008):

$$
\begin{equation*}
N_{t+1}=N_{t}+r N_{t}\left(1-\frac{N_{t}}{K}\right) \tag{4}
\end{equation*}
$$

where $r$ is intrinsic growth rate, and $K$ carrying capacity. Note that the formulation of Equations (2) and (3) used here was tailored for the fish-catching game and is structurally similar to the formulas presented in ecology textbooks, yet strictly speaking, the way the parameters are defined and the equation is structured is very different from the formal presentation of population dynamics models. For example, the two $r$ s in Equations (3) and (4) have different meanings: by transforming Equations (3) and (4), one can easily identify that adding 1 to $r$ in Equation (3) gives $r$ in Equation (4); and also, $k$ in Equation (3) is the reciprocal of the carrying capacity $K$ in Equation (4).

A simulation of density-dependent fishing is given in Figure 2 c , d . By modeling future fish population as a function of density-dependent factor and current fish abundance, one can conceive a case where each villager agrees to set a reasonable level of catch per time that is sufficient for food and less than the level of catch that could lead to population collapse. Note that villagers can catch up to 1-3 fish in this simulation (Figure 1), and it is possible to calculate a range of fish populations such that the number of fish caught by each villager falls within this range (readers may refer to the R code in primerof CommEcol.R). Thus, the tragedy of the commons can be averted by the villager's collective commitment to operate density-dependent fishing.

Readers may play with the functions fishgame 1 and fishgame 2 to contrast the two scenarios (constant catching vs. density-dependent catching) and examine their outcomes by changing parameters. These R-functions can be used to explore questions such as (i) the consequences of altering carrying capacity or vital rates for the fish abundance, and (ii) the consequence of the strength of density dependence. The function fishgamel can be used by setting the parameters NO, initial abundance, Nmax, carrying capacity (fish number at full capacity in the pond), $a, b$, and $c$ (fixed numbers of catch by villager $A, B$, and $C$ ), $r$, intrinsic growth rate, and max_counter, the simulation time steps to execute. Similarly, the function fishgame 2 can be used by setting an additional parameter that is, densitydependent factor $k$.

```
#constant catching scenario
fishgame1 (NO=20, Nmax=30, a=2, b=2, c=1,
r=1.2,max_counter=6)
#density dependent catching scenario
fishgame2 (N0=120, Nmax=200, k=0.003, r=1.2,
max_counter=100)
```

Through the modeling exercise, I show that population abundance depends on birth rate, death rate, carrying capacity, and density dependence. By understanding the factors that influence the abundance of a population, one can develop strategies for conserving and managing ecosystems in a sustainable way. This exercise assumes that fishing is the only mortality factor, but in nature, many ecological processes can impose densitydependent mortality on populations, including crowding (e.g., competition for resources) and natural enemy (e.g., predator). This exercise offers a path toward understanding exponential versus logistic population growth (Gotelli, 2008), continuous-time versus discrete-time models (Gotelli, 2008), and stable versus unstable equilibria and bifurcation theory (Kot, 2008; Otto \& Day, 2007), and more advanced topics in population dynamics such as stochasticity (Shoemaker et al., 2020). References concerning the tragedy of the commons and other game theories can be found in Nowak (2006).

## 4 | PATTERNS OF SPECIES DIVERSITY: THE THEORY OF ISLAND BIOGEOGRAPHY

Islands offer a fascinating setting for the study of community ecology. Islands are unique environments, often with different geological and environmental conditions compared to the mainland. Specifically, islands differ in the abiotic and biotic environment, the area, and the distance from the mainland (degree of isolation). Therefore, it would be intuitively expected that these differences between islands would lead to differences in the way species assemble on islands. The research field that seeks to explain how species assemble in habitat areas is called "community assembly" (Mittelbach \& McGill, 2019).

The theory of island biogeography (MacArthur \& Wilson, 1963), one of the most well-known theories in community ecology, aims to understand patterns of the number of species based on the idea that patterns of species diversity on islands are formed by the balance between the establishment of species migrating from the
mainland to islands and the extinction of species on islands. I use this theory as a guide to explore the patterns of species diversity and show that rates of colonization and extinction can be estimated from the distribution of species on islands. I then show that much insight can be gained by flexibly incorporating a range of ecological features (trophic relationships and species traits) into the theory. Two scenarios are considered: the first where there is a predator-prey relationship between species and colonization and extinction rates vary with trophic level, and the second where colonization and extinction rates vary with species traits such as body size. To demonstrate, I use a toy dataset illustrated in Figure 3, where a handful of bird species live on the mainland and consider how many of them colonize the four islands. For the sake of simplicity, assume that every island has the same area and is equally isolated from the mainland (note the assumption can be relaxed later; see Section 4.4.).

## 4.1 | Derivation of the theory of island biogeography

The place to start is to model the dynamics of the number of species $S$ as a function of the colonization rate $c$, extinction rate $e$, and the size of the mainland species pool $S^{*}$ (the maximum number of species that could potentially become established on the island; Alonso et al., 2015; MacArthur \& Wilson, 1963):

$$
\begin{equation*}
\frac{d S}{d t}=c\left(S^{*}-S\right)-e S, \tag{5}
\end{equation*}
$$

where $S(0)=0$. The first term in Equation (5) describes the product of the colonization rate and the number of species currently absent on a given island, while the second term describes the number of current species times the extinction rate. $S^{*}-S$ means that the already established species cannot colonize again. Solving Equation (5) yields (see Supporting Information for the step-by-step derivation of Equation (6)):

$$
\begin{equation*}
S(t)=S^{*}\left(\frac{c}{c+e}\right)(1-\exp (-(c+e) t)) \tag{6}
\end{equation*}
$$

The term $S^{*}(c /(e+c))$ represents the expected number of species on the island at equilibrium and $(1-\exp (-(c+e) t))$ controls the rate at which the equilibrium number of species is reached. Importantly, $c /(e+c)$ quantifies the probability of occupancy of a species, which is the key parameter in the ensuing subsection. The probabilities of a species $y$ being present and absent

(b)


FIGURE 3 The schematic diagram shows the toy dataset of equilibrium distribution of bird species on the islands (colonizations indicated by the arrows), assuming that each of the four islands has the same distance from the mainland and the same island area. (a) The first case considers a simple trophic structure with one predator species and two prey species. (b) The second case considers that five bird species of different body sizes colonize the islands, and the probabilities of colonization and extinction on an island are assumed to be independent between species. [Color figure can be viewed at wileyonlinelibrary.com]
on the island are represented by $\mathrm{P}(y=1)=c /(e+c)$ and $\mathrm{P}(y=0)=e / e+c$, respectively.

The way species accumulate on the island over time through extinction-colonization dynamics can be visualized using Equation (6) (see the plot function, spaccumplot ). Figure 4 a shows the asymptotic curve of species accumulation using the parameter set: the colonization rate parameter $c=0.1$, the extinction rate parameter $e=0.1$, and $\operatorname{Smax}=S^{*}=100$.

```
#plot species accumulation over time
spaccumplot (c=0.1, e=0.1, Smax=100)
```

To gain insights into how colonization and extinction rates influence the occupancy probability, one could fix $c$ or $e$ and plot the occupancy probability as a function of $e$ (Figure 4b) or $c$ (Figure 4c), respectively. MacArthur and Wilson (1963) envisioned that overlaying the two plots
helps find the expected number of species at equilibrium and how characteristics of island habitats (i.e., area, degree of isolation) influence the expected number of species present. The lower the value of $c$ the more isolated (farther from the mainland) the island is, and the lower the value of $e$, the larger the area of the island is. This reasoning explains why large islands have more species than small islands and why isolated islands have fewer species than islands close to the mainland.

## 4.2 | Parameter estimation of island biogeography

The advantages of the island biogeography theory constitute not only that its model structure is theoretically simple, but also that its simplicity allows simple statistical analysis to estimate key parameters and thus explain how species assemble. I present how the parameters of the model can be estimated using the toy dataset presented in Figure 3. This exercise aims to deepen understanding of the basics of statistical methods and how they


FIGURE 4 The theory of island biogeography. (a) The accumulation of species migrating from the mainland, reaching to asymptote at 50 species (parameters $c=0.1, e=0.1, S^{*}=100$ ) as a function of time (years). (b) The occupancy probability as a function of extinction rate when colonization rate is varied from 0.1 to 0.9 (increasing from the brightest light to the darkest line). (c) The occupancy probability as a function of colonization rate when the extinction rate is varied from 0.1 to 0.9 (increasing from the brightest to the darkest line); by overlaying the two curves ( $b$ and $c$ ) the expected number of species on islands can be obtained. (d) Comparison of log-likelihoods between trophic (black curve) versus original (gray curve) models of the theory of island biogeography (Figure 3a). Vertical dotted lines denote the $\alpha$ values where the log likelihood functions are maximized. (e) Logistic regression fitted to the trait-based island community dataset (Figure 3b), where the occupancy probability is plotted as a function of body size. The observed species' presence/absence are plotted as transparent dots, and hence dots appear darker in color when overlapped. (f) Derivative of $\theta$ (Equation 17) finds the maximum rate of change in the occupancy probability given one unit increase in body size trait. The vertical dotted line denotes the $\theta$ value where the derivative attains a maximum.
can be applied to ecological problems, practically by checking that there is an agreement between the results obtained by hand calculation and those using R.

The theory of island biogeography assumes that all species share the same values for colonization and extinction parameters. For the sake of simplicity, the two-parameter model can be reduced to a one-parameter model by defining the diversity parameter $\alpha=c / e$, where a larger value represents a higher colonization rate, lower extinction rate, or both. For the moment, I follow this assumption ignoring the predator-prey relationship in Figure 3a, and consider how to estimate the parameter $\alpha$. To do this, let us first calculate the probability that the distribution of the three bird species will be generated given the parameter. Let predator species be A, prey 1 B , and prey 2 C . R code for making the data set is as follows:

```
#make dataset of island bird community
#(Figure 3a)
spA=c(1,0,1,1) #predator
spB=c(1,0,0,1) #prey1
spC=c(0,1,1,0) #prey2
tibdat1= data.frame(Species=rep (c
("A", "B","C"), each=4), Occupancy=c (spA,
spB, spC))
```

Theory supposes that $\mathrm{P}(y=1)=\frac{c}{c+e}=\frac{\alpha}{1+\alpha}$ and $\mathrm{P}(y=0)=\frac{e}{c+e}=\frac{1}{1+\alpha}$. The occurrence probability of a species can be described using the Bernoulli distribution. By counting the number of presence/absence of each bird
species on the islands, the probabilities of the occupancy for species $\mathrm{A}, \mathrm{B}$, and C in Figure 3a are given as follows:

$$
\mathrm{P}\left(y_{\mathrm{A}}\right)=\left(\frac{\alpha}{1+\alpha}\right)^{3}\left(\frac{1}{1+\alpha}\right)^{1}, \mathrm{P}\left(y_{\mathrm{B}}\right)=\left(\frac{\alpha}{1+\alpha}\right)^{2}\left(\frac{1}{1+\alpha}\right)^{2}, \mathrm{P}\left(y_{\mathrm{C}}\right)=\left(\frac{\alpha}{1+\alpha}\right)^{2}\left(\frac{1}{1+\alpha}\right)^{2} .
$$

Hence, the joint probability of species occupancy for all the three species yields:

$$
\begin{equation*}
f(y \mid \alpha)=\prod_{i} \mathrm{P}\left(y_{i}\right)=\left(\frac{\alpha}{1+\alpha}\right)^{7}\left(\frac{1}{1+\alpha}\right)^{5} \tag{7}
\end{equation*}
$$

Reframing this, it describes the likelihood of the parameter $\alpha$, given the observed data $y$, that is, seven presences and five absences in the community.

$$
\begin{equation*}
L(\alpha \mid y)=\left(\frac{\alpha}{1+\alpha}\right)^{7}\left(\frac{1}{1+\alpha}\right)^{5} \tag{8}
\end{equation*}
$$

Note that $f(y \mid \alpha)$ is a function of $y$, whereas the likelihood function $L(\alpha \mid y)$ is a function of $\alpha$. Here I use the maximum likelihood method, a statistical method used to estimate the parameter(s) of a statistical model (Bolker, 2008). In principle, the observed data are most likely to have been generated by a set of parameter values that maximize the likelihood function (i.e., the maximum likelihood principle). To estimate $\alpha$ that maximizes the likelihood function given the data, one can take logarithms on both sides to get log-likelihood:

$$
\begin{equation*}
\log L(\alpha \mid y)=7 \log \alpha-12 \log (1+\alpha) \tag{9}
\end{equation*}
$$

Solving Equation (9) set to 0 gives $\alpha=7 / 5$, and this can be converted into the occupancy probability (i.e., $\alpha /(1+\alpha)): 7 / 12=0.5833$.

Next, I demonstrate the parameter estimation using R. Parameter $\alpha$ can be estimated using glm function, $g l m$ is a function for fitting a generalized linear model to data, which is the standard function in modern statistical analysis in R. In this case, the response variable is the binary data of whether a given bird species is present or absent on the island (i.e., the occupancy probability), and the explanatory variable is a constant (intercept), the only parameter in the model. The important thing is to choose a probability distribution to fit the data, here I use a binomial distribution and a link function "identity," because it allows to model the response variable using the explanatory variable without any transformation.

Following glm nomenclature:

```
#fit generalized linear model
mod=glm(Occupancy~1, data=tibdat1,
family=binomial(link="identity"))
#get loglikelihood
logLik(mod)
#get a summary of the glm analysis
summary(mod)
```

The logLik function results in $-8.150(\mathrm{df}=1)$. The detailed statistics can be output using summary below. The summary table shows that the estimated coefficient for the intercept is 0.5833 . I confirm that for the simple theory of island biogeography, the results of maximum likelihood estimation calculated by hand agreed with those produced by R programming.

```
Call:
glm(formula = Occupancy ~ 1, family=
binomial(link = "identity"), data = tibdat1)
Deviance Residuals:
\begin{tabular}{lcrcl} 
Min & 1Q & Median & 3Q & Max \\
\(-1.323-1.323\) & 1.038 & 1.038 & 1.038
\end{tabular}
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.5833 0.1423 4.099 4.15e-05
***
Signif. codes:0'***'0.001 '**'0.01'*' 0.05`.'
0.1'`1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 16.301 on 11 degrees of freedom Residual deviance: 16.301 on 11 degrees of freedom AIC: 18.301
```

Number of Fisher Scoring iterations: 2

## 4.3 | Trophic theory of island biogeography

It is time to relax the simplified assumption that all bird species have the same rates of colonization and
extinction. For example, a predator cannot establish itself unless their prey species are present on the island; a predator also necessarily goes extinct when the last prey species goes extinct. Therefore, the distribution of bird species on islands may be better explained by considering the trophic relationships between species (i.e., predators and their prey). One way to do this is to compare model fits with and without such trophic structure using the maximum likelihood approach (Gravel et al., 2011). Let us go back to the toy dataset Figure 3a; among three species, species A is a predator species that feeds on the other two species B and C.

Let $p_{g}$ the occupancy of a species with diet breadth $g$, and assume (1) a species can only invade the focal island if at least one species already present on the island is a natural prey of the focal species; (2) a species that loses its last prey on the island (due to extinction processes) also goes extinct. I follow to define $q_{g}$ as the probability a prey species has one or more of its prey species present in the island when it colonizes, and $\varepsilon_{g}$ the rate at which a species with diet breadth $g$ loses its last prey item species. Therefore, the trophic theory of island biogeography can be modeled by incorporating $q_{g}$ and $\varepsilon_{g}$ into Equation (5) (Gravel et al., 2011):

$$
\begin{equation*}
\frac{d p_{g}}{d t}=c\left(1-p_{g}\right) q_{g}-\left(e+\varepsilon_{g}\right) p_{g} . \tag{10}
\end{equation*}
$$

Solving Equation (10) below yields the following equilibrium occurrence probability of species:

$$
\begin{equation*}
p_{g}^{*}=\frac{c q_{g}}{c q_{g}+e+\varepsilon_{g}} . \tag{11}
\end{equation*}
$$

For basal species (species B and C) the colonization rate and extinction rate are the same as the one given in the original theory of island biogeography: $\mathrm{P}(y=1)=$ $c /(c+e)$ and $\mathrm{P}(y=0)=e /(c+e)$. Next, I derive $q$ and $\varepsilon$ for predator species A using the probabilistic statement:

$$
\begin{gather*}
q=1-\mathrm{P}\left(y_{\mathrm{B}}=0\right) \mathrm{P}\left(y_{\mathrm{C}}=0\right)=1-\left(\frac{e}{c+e}\right)^{2} . \\
\varepsilon=e_{\mathrm{B}} \mathrm{P}\left(y_{B}=1\right) \mathrm{P}\left(y_{\mathrm{C}}=0\right)+e_{\mathrm{C}} \mathrm{P}\left(y_{\mathrm{B}}=0\right) \mathrm{P}\left(y_{\mathrm{C}}=1\right)=\frac{2 c e^{2}}{(c+e)^{2}}, \tag{13}
\end{gather*}
$$

where $e_{\mathrm{B}}$ and $e_{\mathrm{C}}$ are the extinction probabilities of species $B$ and C. Plugging Equations (12) and (13) into Equation (11) and rewriting with the diversity parameter $\alpha$, one obtains the occupancy probability of predator A (see Supporting Information for derivation):

$$
\begin{equation*}
\mathrm{P}\left(y_{\mathrm{A}}=1\right)=\frac{\alpha\left(1-\left(\frac{1}{1+\alpha}\right)^{2}\right)}{1+\alpha\left(1+\left(\frac{1}{1+\alpha}\right)^{2}\right)} . \tag{14}
\end{equation*}
$$

R code for maximum likelihood is therefore given below. It calculates log-likelihood for the whole dataset (three species combined including predator A and prey species B and C). By using optimize function that performs numerical optimization, one can estimate $\alpha=$ 2.451 and get a log-likelihood of -7.842 .

```
#log likelihood function for predator
predp = function (x){c(c(c(x^2)*c(x+2))/c(c(1
+x)^}3+x))
logLpredp= function (x){log(c(predp (x)^3)*C
(1-predp (x))^^1)}
#log likelihood function for prey 1 & 2
logLprey1p= function(x){log(c(c(x/c(x+1))^2)
*c(1/c(x+1))^2)}
logLprey2p= function(x){log(c(c(x/c(x+1))^3)
*c(1/c(x+1))^1)}
```

\#overall log-likelihood
$\log L=f u n c t i o n(x)\{\log \operatorname{Lpredp}(x)+\log \operatorname{Lprey} 1 p(x)$

+ logLprey2p (x) \}
\#find maximum alpha value optimize ( $\mathrm{f}=\mathrm{log} \mathrm{L}$, interval $=\mathrm{C}(0,100)$, maximum=T)

I find that the log-likelihood is greater when the trophic structure is explicitly considered compared to when it is not (trophic model vs. original model $=-7.842$ vs. -8.150 ; Figure 4 d ), suggesting that trophic position influences the pattern of species distribution on the islands. This section demonstrates through likelihoodbased goodness-of-fit measures the ability of the trophic theory of island biogeography to predict species-specific occupancies and compared them with the null expectation of the original theory of island biogeography. For those who are interested in more complex trophic interactions in the food web may refer to Gravel et al. (2011). For those who wish to study statistical inference based on likelihood and model selection, Hilborn and Mangel
(1997) and Bolker (2008) provide empirical views of the link between theory and data.

## 4.4 | A trait-based theory of island biogeography

If a species trait predicts why a particular species is present in one place and absent in another, then it can be empirically tested whether this holds true for many other species or ecosystems. The measurement of traits is therefore a promising approach to explaining community assembly across taxa and ecosystems, given that traits can be easily measured and linked to specific aspects of species performance. Ecologists have measured and compiled databases of a wide range of traits in a large number of taxa, and by linking species traits with the distribution and abundance of species, there is huge potential to provide a testable and predictive explanation of community patterns (McGill et al., 2006). To start, the dataset in Figure 3b I use can be created in R:

```
# trait-based theory of island biogeography
#(Figure 3b)
spA=c(1,0,0,0)
spB=c (0,0,0,1)
spC=c (0,1,0,1)
spD=c(1, 1, 1,0)
spE=c (1, 1, 1, 1)
bodysize=c(12,14,18,20,30)
tibdat2=data.frame(species=rep (c)(LETTERS
[1:5]),each=4),bodysize=rep(bodysize,
each=4), occopancy=c (spA,spB, spC,spD,spE))
```

A simple formulation of the trait-based approach describes the occupancy probability as a function of the body size of bird species. Note that a logit link is adopted here instead of an identity link (used in parameter estimation Section 4.2.), for the response variable has to be logit-transformed beforehand to avoid the problem that the left side can be binary outcomes ( 0 or 1 ), while the right side can take any real number ( $+\infty$ to $-\infty$ ):

$$
\begin{equation*}
\log \left(\frac{\theta}{1-\theta}\right)=\beta_{0}+\beta_{1} x, \tag{15}
\end{equation*}
$$

where $\theta=\alpha /(1+\alpha)$. This can be transformed to:

$$
\begin{equation*}
\theta=\frac{e^{\beta_{0}+\beta_{1} x}}{1+e^{\beta_{0}+\beta_{1} x}} . \tag{16}
\end{equation*}
$$

Since this cannot be solved in closed form, "glm" can be used to estimate the parameter (Figure 4e)

```
#fit glm to the data
logreg = glm(occupoancy~bodysize,
data=tibdat2, family=binomial(link="logit"))
summary(logreg)
```

The summary output shows:

```
Call:
glm(formula = occupancy ~ bodysize, family=
binomial(link = "logit"), data = tibdat2)
Deviance Residuals:
Min 1Q Median 3Q Max
-1.5945-0.82420.1885 0.81141.8540
Coefficients:
\begin{tabular}{lrccc} 
& Estimate & Std. Error & \(z\) value & \(\operatorname{Pr}(>|z|)\) \\
(Intercept) & -5.2159 & 2.6399 & -1.976 & 0.0482 * \\
bodysize & 0.3079 & 0.1549 & 1.987 & 0.0469 *
\end{tabular}
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05'.'
0.1`'1
(Dispersion parameter for binomial family
taken to be 1)
Null deviance: 27.526 on 19 degrees of freedom
Residual deviance: 19.461 on 18 degrees of
freedom
AIC:23.461
```

Number of Fisher Scoring iterations: 5

The estimated coefficient of body size is statistically significant (estimated coefficient $=0.308, p=0.047$ ), justifying that body size can be effectively used to explain the probability of occupancy in the island bird community. Figure 4 e visualizes the model fit based on the estimated coefficient to
the observed data, where the observed and predicted occupancies are plotted as a function of body size of the bird species. The dots are the observed presence (1) and absences (0) drawn in transparent color so that where many dots overlap, the color appears darker.

In logistic regression, it is not intuitively clear to what extent a unit increase in body size increases the occupancy probability since the probability changes nonlinearly as a function of body size (Figure 4e). One way to gain insight into this question is to differentiate Equation (16) by body size $x$ and find the possible maximum increase in probability for a unit change in body size.

$$
\begin{equation*}
\frac{d \theta}{d x}=\frac{\beta_{1} e^{-\left(\beta_{0}+\beta_{1} x\right)}}{\left(1+e^{-\left(\beta_{0}+\beta_{1} x\right)}\right)^{2}} \tag{17}
\end{equation*}
$$

By solving Equation (17) set to 0, one finds Equation (16) is maximized at $x=-\frac{\beta_{0}}{\beta_{1}}=16.940$ and the maximum value is $\frac{\beta_{1}}{4}=0.077$. This means that a 1 cm increase in the body size of a bird species increases the probability of its occurrence by a maximum of $7.7 \%$ (Figure 4f). The effect that a unit change in the explanatory variable has on the response variable leads to the concept "effect size" used by statisticians.

Although it is assumed that each island is of equal size and is equally isolated, it is theoretically possible to estimate (i) how island area affects extinction rate as well as (ii) how island isolation affects colonization rate, using real-world datasets. This can be done by explicitly modeling that the colonization rate is inversely proportional to the distance from the mainland $D$ (fewer species on more isolated islands) $c \approx D^{\gamma}$ and the extinction rate is also proportional to the island area $A$ (fewer extinctions on larger areas) $e \approx A^{\delta}$. To statistically estimate the effect of area and distance from the mainland (or more precisely, their logarithmic values) on the probability of species occupancy $\theta$, one can build a logistic model with two parameters $\gamma$ and $\delta$ :

$$
\begin{equation*}
\log \left(\frac{\theta}{1-\theta}\right)=\delta \log A-\gamma \log D \tag{18}
\end{equation*}
$$

Last, I mention that the theory of island biogeography paves way for understanding metapopulation theory, a general theory that describes the distribution and dynamics of populations of a species in a landscape made up of multiple, isolated habitats or patches (Hanski, 1998). Unlike the mainland-island architecture of theory of island biogeography, the metapopulation theory models the probability of colonization/extinction in a focal patch as a function of a collection of the surrounding
patches as a whole and explores if species can persist in such landscapes when there is sufficient dispersal among the isolated patches, allowing for the colonization of new patches and the rescue of declining populations. Metapopulation theory provides a framework for understanding how landscape structure influences the distribution of species across a landscape. Specifically, one can estimate how much habitat is needed to maintain the current diversity of a community, and examine how many species are threatened with extinction if large habitats are destroyed and fragmented into smaller areas (Hanski, 1998).

## 5 | MODELING COEXISTENCE: COMPETITIVE EXCLUSION, NEUTRAL THEORY, AND FREQUENCY DEPENDENCE

Until now we have looked at the abundance of one species or the distribution of species, and here let us move on to examine both aspects, that is, the relative abundances of species or, more narrowly, the question of species coexistence. To begin with, I consider how competitive exclusion occurs, that is, when competition drives a species to extinction in a deterministic model of competition. I will then consider the stochastic version of a competition model, which is based on the simplified assumption that per capita birth and death rates are equivalent among individuals (and hence neutral). Stochastic simulation is an effective tool for an intuitive understanding of the neutral theory of biodiversity and has the advantage of easily incorporating the frequency dependence necessary to achieve coexistence (Shoemaker et al., 2020; Vellend, 2017). Through the exercise, I show how the coexistence of a tree community is made possible by incorporating frequency dependence, such as the accumulation of host-specific pathogens in the soil, into the neutral model.

## 5.1 | Competitive exclusion

Why are there so many plant species coexisting on the same resource? This is one the oldest questions that still remains as an active topic of discussion in modern ecology (Chase \& Leibold, 2003; Hutchinson, 1957, 1961; Mittelbach \& McGill, 2019). Competitive exclusion refers to the idea that two species competing for the same resources in an ecosystem cannot coexist in equilibrium. This concept is often demonstrated using a model of plants competing for light and nutrients (Tilman, 1982). Tilman (1982) presented a mechanistic model of plant
populations and communities that explicitly includes consumer-resource interactions. To understand how one species will outcompete the other and drive it to extinction in the presence of resource competition, here I start by modeling the dynamics of a single plant species population and its resource.

The plant abundance $N$ and its resource abundance $R$ (such as nitrate) can be modeled as follows:

$$
\begin{gather*}
\frac{d N}{d t}=N(a R-d) .  \tag{19}\\
\frac{d R}{d t}=r-a R N \tag{20}
\end{gather*}
$$

where $a$ is the rate at which plant species consume the resource, $d$ is the mortality rate of plant species, and $r$ is the rate at which resource is supplied. The resource abundance at equilibrium can be obtained by setting Equations (19) and (20) each being equal to $0: R^{*}=\frac{d}{a}$. Figure 5a shows a numerical simulation example of the plant-resource dynamics based on the discretized version of Tilman's model (see source code for details about how to perform numerical simulation). This suggests that the system reaches the equilibrium state at $R^{*}=\frac{d}{a}$.

Then, let us consider the scenario in which $n$ plant species compete for the shared resource. The mathematical formulae are as follows:

$$
\begin{align*}
& \frac{d N_{j}}{d t}=N_{j}\left(a_{j} R-d_{j}\right) .  \tag{21}\\
& \frac{d R}{d t}=r-R \sum_{j}^{N} a_{j} N_{j}, \tag{22}
\end{align*}
$$

where $a_{j}$ is the rate at which plant species $j$ consume the resource, $d_{j}$ is the mortality rate of plant species (but for now assume all plant species have the same death rate $d$ ). Under this setting, the equilibrium resource abundance yields $R^{*}=d / a_{j}$. To examine the outcome of multi-species plant competition, let us consider the numerical simulation of two plant species (competitors). Numerical simulation (Figure 5b) shows that the species with the lower $R^{*}$ wins the competition. Note that, in this case, the plants do not interfere with each other, but simply consume resources, thereby indirectly influencing the other plant species. The reason why species with lower $R^{*}$ win the competition is that the species is able to maintain its populations even if the resource abundance is reduced to a lower level (known as " $R^{*}$ rule"). To explore the generality of the $R^{*}$ rule, one could test whether the $R^{*}$ rule correctly predicts the outcome of exclusion by


FIGURE 5 Competitive exclusion predicted from consumerresource theory. (a) The results of a simulation of temporal changes in the amount of resource $R$ (e.g., nitrogen in the soil) and the abundance of a plant species $N$. The parameters used were: $N=10$ (initial condition), $R=10$ (initial condition), $a=0.02, d=0.2$, $r=1$. (b) The result of a simulation of temporal changes in the amount of resource and the abundances of two plant species $N_{1}$ and $N_{2}$ (Plant sp. 1 and sp.2) sharing the resource $R$. Species 2 (light gray curve), which can sustain its population at a lower resource level, displaces species 1 (dark gray curve).
reproducing the dynamics of consumer-resource interactions among 100 plant species (for each parameter, one can generate 100 values using the function "rnorm" generating the normal distribution and assign them to each species). Note, however, that the simulation examines only a part of the parameter space of a model, and in order to conduct a proper study, one must examine the vast parameter space and hence the generality of the conclusions obtained.

## 5.2 | The neutral theory of biodiversity

As seen in Section 5.1, if competition for resources makes it so difficult for plant species to coexist, why do so many plant species appear to coexist in nature? To bridge this
gap between theory and reality, ecologists have analyzed experimental and observational data using a variety of approaches. One notable approach, the neutral theory, changed the way we think about species coexistence, particularly in tree communities. Away from the overwhelmingly dominant idea at the time, which was to ask what species differences were necessary for communities to coexist, the neutral theory considers what would happen to tree communities if all tree individuals had the same per capita birth and death rates (and therefore there were no differences between species; the species equivalence assumption) (Hubbell, 2001). Hubbell (2001) argued that many species may coexist because of random processes such as dispersal, establishment, and extinction, rather than species trait differences involving competition for resources. This proposition has ignited one of the most intense debates about the structuring forces of communities (for those who are interested in the discussion in detail may refer to Mittelbach \& McGill, 2019).

There are several types of neutral models (Hubbell, 2001), but the simplest model discussed here considers the forest as a collection of cells, with one individual tree present in each cell and all the cells being fully occupied. To illustrate the model, below I modified Vellend's R code (Vellend, 2017). The neutral model simulation is carried out as follows: (1) consider a fixed, finite number of individuals in the entire tree community (community size), arising from unavoidable physical constraints on the total number of individuals that can be packed into a given space; (2) assume that the total individuals comprise two species, maple and cherry; (3) the initial condition is that $90 \%$ of the cells are occupied by maple trees; (4) at each time step, two individuals are randomly selected. The first individual dies and is replaced immediately by the offspring of the second individual. The newly colonized individual grows from seedling to adult until the next time step; and (5) the step 4 (death and replacement of individuals) is repeated over many time steps, and the relative occupancy of the two species is calculated each time. This model can be run using neutralmodel, with three parameters, n community size (total number of cells or trees), iniprop, initial proportion of maple tree, and nyears number of years to simulate:

[^2]By simulating tree community dynamics with two tree species (maple and cherry), the result shows stochastic dynamics of the tree populations: if the simulations start with $90 \%$ maple trees and $10 \%$ cherry trees, the random walk often leads to the extinction of cherry trees (Figure 6, black line). However, in some cases, maple trees become extinct (Figure 6, light gray line) and/or the two species can exhibit transient coexistence (Figure 6, dark gray line). Thus, a characteristic feature of stochastic simulations is that the trajectory followed by forest dynamics changes with each simulation due to its stochastic nature. To ensure that the simulation result is sufficiently representative, the stochastic simulations can be repeated many times to get a sense of how changing a parameter might affect the trajectory of relative abundance. In Figure 6, I show the result of 100 runs of stochastic simulations and overlay their dynamic trajectories using transparent colors; thus, areas with more shading mean that more runs pass through the areas.

This example shows that stochasticity constitutes an important component of tree community and affects the outcome of coexistence vs. exclusion. As seen later, the neutral model is useful as a null model to assess how the relative abundances of tree species change in the absence of specific traits that would produce frequency dependence.

## 5.3 | Frequency dependence via plantsoil feedback

Negative frequency dependence refers to the concept that the relative abundance of a species decreases as its frequency in a community increases; species are more likely to be favored if they are rare and disfavored if they are common. Negative frequency dependence is a general hallmark of multi-species coexistence, and ecologists continue to explore what processes on earth contribute to producing negative frequency dependence. The available evidence suggests that there is some degree of negative frequency dependence in many natural systems (Mittelbach \& McGill, 2019). There are a number of processes that lead to negative frequency dependence. One of the most studied in recent years is plant-soil feedback. Plant-soil feedback has been recognized as a crucial factor in mediating plant community diversity and coexistence (Bever et al., 1997). Negative plant-soil feedback can promote the coexistence of tree communities, if pathogens accumulate in the soil, preventing conspecific seedlings from establishment and favoring heterospecific seedlings. Seedlings may fail to colonize near adult trees of the same species due to local pathogen accumulation and may survive through the repeated acquisition of pathogen-free sites.


A simulation incorporating negative frequency dependence via plant-soil feedback can be executed as follows. Up to step 3, it is the same as the neutral model (Section 5.2), but step 4 is different. In the neutral model, all individuals in the community are equally likely to colonize the cell of a dead individual, whereas, in the frequency-dependent model, conspecifics are less likely to replace the cell. The strength of the feedback "feedbackst" indicates the relative survival rate of seedlings, that is, the percentage of individuals that can potentially colonize the new site. A new individual to replace the dead individual is randomly drawn from the pool of individuals subject to this conspecific disadvantage defined by "feedbackst." Therefore, the stronger the negative feedback, the less likely for conspecific trees to replace the cell. One can compare neutralmodel and psfmodel to understand what frequency dependence brings about the two species tree community. One can visualize how negative frequency dependence creates a dynamic equilibrium in which the abundance of the two species fluctuates over time (Figure 7). These R-functions help to probe questions such as (i) the consequences of changing the initial condition for the outcome of coexistence, (ii) whether community size influences how quickly competitive exclusion occurs, and (iii) what happens as plant-soil feedback becomes stronger (Figure 7).

[^3]FIGURE 6 The results of the numerical simulations of the neutral theory of biodiversity, that is, 100 simulation runs of the relative species occupancy of two tree species (maple and cherry) over 10,000 years, are shown as hues of shades. The darker shading indicates the areas that were passed through by more runs of the simulated trajectories. Three representative examples (extinction of maple, extinction of cherry, and both species persisting after 10,000 years) are shown as bold lines.


FIGURE 7 Incorporating negative plant-soil feedback into the neutral model. The results of 10 simulations, each with three different levels of feedback strength, are shown for up to 500 years. PSF represents the strength of negative plant-soil feedback experienced by tree seedlings, a hallmark of negative frequency dependence. It can be seen that the stronger the feedback, the faster it approaches an equal community share of the two species and continues to fluctuate randomly.

The primary focus in this section has been on individual death-birth events in a single local community, but by considering a metacommunity (e.g., mainland), where speciation occurs and acts as a source of species for the local community, one can explore how biodiversity is determined by a combination of phenomena occurring at regional and local scales (the unified neutral theory of biodiversity and biogeography; Hubbell, 2001 and see also Leibold \& Chase, 2017 for metacommunity theory). Furthermore, although I have focused on negative frequency
dependence as a stabilizing force, positive frequency dependence (e.g., mycorrhizal mutualism; Kadowaki et al., 2018), community modules (e.g., indirect predator-mediated effects; Holt, 1977; Kondoh, 2008), and interaction networks (e.g., pollination networks; Bascompte \& Jordano, 2013) are also possible forms of the dynamic component of an ecological community. How these complex forms combine to allow species-rich communities to coexist is an open question in ecology (Mittelbach \& McGill, 2019). Nevertheless, ecologists are making progress, and here I highlight three significant insights. First, the actual interaction networks may have distinctive structures that are theoretically expected to enhance coexistence (Bascompte \& Jordano, 2013; Mougi, 2020). Second, not all species contribute equally to the stability of the community, with some species being overwhelmingly more important than others (keystone species; Paine, 1966). Third, stochasticity alters the frequency dependence and mediates the outcome of coexistence depending on how species are differentially affected by and respond to such stochastic fluctuations in the environment (Chesson, 2000; Mittelbach \& McGill, 2019).

## 6 | SUMMARY

There is still a long way to go before all the basic knowledge of community ecology is covered. However, when asked what I would do if I had to select a suite of topics from the vast number of important ones, I came up with the structure of this paper. As a consequence, the emphasis was on facilitating an intuitive understanding of community ecology through the use of theory, statistics, and numerical simulations. Indeed, many of the questions that community ecologists address are not the sort of questions that can be answered with a simple yes or no, but rather require quantitative assessment from a broad spectrum of perspectives. Therefore, the nature of the questions about community ecology calls for an approach that effectively combines theory and data. Integrating data and theory to unravel ecological phenomena has never been more relevant, as modern ecologists are challenged to solve pressing problems of the global environmental and biodiversity crises. To achieve this, a good understanding of concepts and theories, as well as the programming skills to express them, are essential. I hope that readers will consider the essence of community ecology by modifying, experimenting, and playing with the models with their own ideas, and immerse themselves in the fascination of community ecology.

## AUTHOR CONTRIBUTIONS

Kohmei Kadowaki conceived the study design, performed data analysis, and wrote the manuscript.

## ACKNOWLEDGMENTS

I thank Editor-in-Chief for inviting me to contribute an article to the journal. I thank Samuel Ross, Yuuya Tachiki, Yutaro Fujimoto, and an anonymous reviewer who commented on the manuscript. Thanks go to Tomoko Kadowaki for help with figure preparation.

## FUNDING INFORMATION

Funding was provided by the Hakubi Project (Kyoto University) and KAKENHI Project No. 21H02233.

## CONFLICT OF INTEREST STATEMENT

The author declares no conflict of interest.

## ORCID

Kohmei Kadowaki © https://orcid.org/0000-0003-05121621

## REFERENCES

Alonso, D., Pinyil-Gallemi, A., Alcoverro, T., \& Arthur, R. (2015). Fish community reassembly after a coral mass mortality: Higher trophic groups are subject to increased rates of extinction. Ecology Letters, 18, 451-461.
Bascompte, J., \& Jordano, P. (2013). Mutualistic networks. Princeton University Press.
Bever, J. D., Westover, K. M., \& Antonovics, J. (1997). Incorporating the soil community into plant population dynamics: The utility of the feedback approach. Journal of Ecology, 85, 561-573.
Bolker, B. M. (2008). Ecological models and data in R. Princeton University Press.
Chase, J. M., \& Leibold, M. A. (2003). Ecological niches: Linking classical and contemporary approaches. Chicago University Press.
Chesson, P. (2000). Mechanisms of maintenance of species diversity. Annual Review of Ecology and Systematics, 31, 343-366.
Gotelli, N. J. (2008). A primer of ecology (4th ed.). Sinauer.
Gravel, D., Massol, F., Canard, E., Mouillot, D., \& Mouquet, N. (2011). Trophic theory of Island biogeography. Ecology Letters, 14, 1010-1016.
Hanski, I. (1998). Metapopulation dynamics. Nature, 396, 41-49.
Hardin, G. (1968). The tragedy of the commons. Science, 162, 12431248.

Hilborn, R., \& Mangel, M. (1997). The ecological detective: Confronting models with data. Princeton University Press.
Holt, R. D. (1977). Predation, apparent competition, and the structure of prey communities. Theoretical Population Biology, 12, 197-229.
Hubbell, S. P. (2001). The unified neutral theory of biodiversity and biogeography. Princeton University Press.
Hutchinson, G. E. (1957). Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology, 22, 415-427.

Hutchinson, G. E. (1961). The paradox of the plankton. The American Naturalist, 95, 137-145.
Kadowaki, K., Yamamoto, S., Sato, H., Tanabe, A. S., Hidaka, A., \& Toju, H. (2018). Mycorrhizal fungi mediate the direction and strength of plant-soil feedbacks differently between arbuscular mycorrhizal and ectomycorrhizal communities. Communications Biology, 1, 196.
Kondoh, M. (2008). Building trophic modules into a persistent food web. Proceedings of the National Academy of Sciences, 105, 16631-16635.
Kot, M. (2008). Elements of mathematical ecology. Cambridge University Press.
Leibold, M. A., \& Chase, J. M. (2017). Metacommunity ecology. Princeton University Press.
MacArthur, R. H., \& Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. Evolution, 17, 373-387.
McGill, M. J., Enquist, B. J., Weiher, E., \& Westoby, M. (2006). Rebuilding community ecology from functional traits. Trends in Ecology \& Evolution, 21, 178-185.
Mittelbach, G. G., \& McGill, B. J. (2019). Community ecology (2nd ed.). Oxford University Press.
Mougi, A. (2020). Diversity of functional traits and interactions. Perspectives on community dynamics. Springer.
Nowak, M. A. (2006). Evolutionary dynamics: Exploring the equations of life. Belknap Press.
Otto, S. P., \& Day, T. (2007). A Biologist's guide to mathematical modeling in ecology and evolution. Princeton University Press.
Paine, R. T. (1966). Food web complexity and species diversity. The American Naturalist, 100, 65-75.

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing https:// www.R-project.org/
Shoemaker, L. G., Suivvan, L. L., Donohue, I., Cabral, J. S., Williams, R. J., Mayfield, M. M., et al. (2020). Integrating the underlying structure of stochasticity into community ecology. Ecology, 101, e02922.
Tilman, D. (1982). Resource competition and community structure. Princeton University Press.
Vellend, M. (2017). Theory of ecological communities. Princeton University Press.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4, 1686.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kadowaki, K. (2023). A primer of community ecology using the R language. Population Ecology, 65(4), 240-256. https://doi.org/10.1002/1438-390X. 12158


[^0]:    This forum paper is invited, based on a symposium held on September 27, 2019, at the 35th Population Ecology Annual Meeting in Kyoto.

[^1]:    This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
    (c) 2023 The Authors. Population Ecology published by John Wiley \& Sons Australia, Ltd on behalf of The Society of Population Ecology.

[^2]:    \# simulate the dynamics of the neutral model
    neutralmodel ( $\mathrm{n}=100$, iniprop=0.1,
    nyears=10000)

[^3]:    \# simulate the dynamics of frequency dependent
    \# model driven by plant-soil feedback
    psfmodel(n=100, iniprop=0.1, nyears=10000,
    feedbackst=0.1)

