Problem Chosen	2023 MCM/ICM Summary Sheet	Team Control Number
A		**2315431**

Where's My Water?

Summary

With the rise of global warming, in the late-20th and 21st centuries, it has become readily apparent that the impact of human activity will affect previously longstanding ecosystems greatly in the short term. Per ecosystem, perhaps the single most important trophic level on the foodchain is the plant fauna, however, these are likewise the most susceptible organisms to even small-scale perturbations in precipitation, weather, and pollution patterns. In order to combat this, we aim to develop a greater understanding of factors driving plant evolution, with particular interest in symbiotic inter-population behavior and drought adaptability.

To accomplish this, we formulate a system of ordinary differential equations and time-dependent partial differential equations to simulate population dynamics over a specified domain. Specifically, we evolve plant populations through time via the **competitive Lotka-Volterra** equations and simulate population diffusion through time via a system of **coupled Fisher-KPP equations**.

These differential equations are coupled with several biologically-inspired environment models which simulate extreme weather events, the diffusion of precipitation throughout the domain, several kinds of pollution effects, and geographical features that determine a dynamic world with resources affecting plant growth. To simulate this world, we develop a numerical scheme on various evolution scenarios that are evaluated through a combination of adaptive time-stepping schemes (RK45), finite difference methods, and method of lines solvers. Some of these are implemented in `py-pde` and `sci-py`, but we developed our own heat equation solver as well.

This approach gives us the freedom to then model extreme environmental scenarios such as polluti on and irregular droughts by developing scaling transforms applied to the raw quantities of resources.

We quantify biodiversity using the **Hill numbers** and analyze their relationship with the number of species in the system. We discover an increase in biodiversity when the number of species is around 5 or 6, especially if there are a large quantity of droughts with medium to high severity. Through additional case studies and ablation tests, we verify the biological and theoretical robustness of our system, where we discover its ability to model unique types of species, agree with theoretical equilibria, and account for pollutants. Finally, we propose a data-inspired scheme for ensuring long-term plant viability.

Keywords: Competitive Lotka-Volterra, Fisher-KPP, Finite Differences, Interspecies Population Dynamics, Extreme Precipitation Events, Simulation

Contents

1 Introduction

We are interested in the development of a high fidelity, generalizable model of the large-scale ecological behavior of several interacting plant species. Globally, plant biomass dwarfs mammals – despite the impossibility of a sharp estimate, the numbers of trees alone on the planet is estimated to be in the trillions [Ehr15]. In light of global trends with respect to global warming and pollution, ecologically "rare events" occur with increasing frequency; in particular, the rise in average global temperatures comes coupled with an increase in heat-waves and extreme-precipitation phenomena (inclusive of both drought-like and monsoon-like)[1], both to which the planet's plant life will largely be vulnerable, though individually to widely varying degrees.

In lieu of counteracting global warming processes, we would like to investigate what can be done locally to preserve plant life, especially in regard to drought resilience; we conflate a success here with implications for success in preserving the long term resilience of ecosystems in their entirety in the face of global warming. This can be broken into the following components:

1. How many different plant species are required to achieve this effect of collective benefit, and how do the types of species matter?
2. What happens when extreme precipitation events become more frequent and severe?
3. How do other factors like pollution and habitat reduction impact the model?
4. What does the model imply can be done to ensure the long term well being and viability of the plant communities?

We chose to approach this problem with a focus on modeling dynamics between species through running numerical simulation. The analysis portion of this paper conducts multiple case studies, each varying different parameters of the model to see what changes in the simulations. Our framework is generalizable and provides a healthy amount of flexibility to simulate unique environments and interactions between species with high fidelity.

2 Global Assumptions

We make various assumptions and provide justification. Many of them are the cruxes of multiple points in our model. We also make implicit assumptions specific to the various subroutines, so here we only mention the explicit and global ones for conciseness and postpone discussion of specific assumptions to their respective sections.

1. **Up to small diffusive constants , an arbitrary domain Ω can be represented sufficiently with a unit square $\subset \mathbb{R}^2$.**

   Importantly, this allows us to greatly simplify the computation and parameterization of the meteorological and diffusive processes we want to simulate. In theory, modeling these as close to a real world understanding would also involve the prescription of a smooth surface $\subset \mathbb{R}^3$ and gradient computations along this manifold, but we shortcut this through this assumption. Generating terrains algorithmically is not feasible, so we would admit a hit to generalizability in doing so.

[1] https://www.c2es.org/content/heat-waves-and-climate-change/

2. **Plants only consume water and nutrients from the soil.**

   We assume it will be sufficient to only consider 2 resources under the geometric assumption. This prevents cluttering of the model with additional features, and allows us also to reduce concerns over the curse of dimensionality. Physically, the most notable loss is with respect to a plant species's consumption of light and dependence on 3d space to grow.

3. **Environmental effects from outside $\Omega \subset \mathbb{R}^2$ are negligible in net against the effects explicitly imposed on the interior.**

   We leverage this assumption to consider only the dissipation of internal resources into the outside environment, but can avoid worrying about inflow of resources.

4. **All quantities can be taken as non-dimensionalized at simulation time.**

   By a *dimensionless quantity*, we mean a quantity that has not been assigned units of physical measurement.

   Again in light of the first assumption, this is less of an assumption and more of a modeling decision to save the time and effort of computing transformations to remove the dimensions (since we need dimensionless variables to apply the differential equations). In a sense, we can capture greater generalizability to differing scales by not imposing model decisions onto a particular scale.

   The primary dimensionless quantities of interest are the population densities (interpreted as plant biomasses), time, the two resource quantities, and relevant derivatives.

5. **The time scale is sufficently small such that the effects of generational evolution are negligible.**

   Evolutionary dynamics is difficult to reasonably model in the continuum case. We provide a model framework in which evolutionary dynamics theoretically could be added, but they are not included in our implementations for fear of needing to overtune simulations in addition to time and compute constraints. Instead, we prioritize modeling intergenerational effects.

6. **Population growth can be modelled logistically.**

   We base one of our main dynamics (competitive Lotka-Volterra) on this assumption. See subsection 3.1 for a more in depth discussion.

7. **There are no inter-plant predatory interactions, only competitive interactions.**

   This again is an assumption that supports the validity of using the Competitive Lotka-Volterra equations.

3 Forward Problem Methodology

As a brief overview, we include a rough schematic of how the forward solver is pieced together in Figure 1.

3.1 Modelling Population Dynamics

Real populations do not exist in isolation but undergo inter-species competition, parasitism, mutualism, and other ecological interactions. For instance, as irrigation and nutrients in Ω are limited and subject to dramatic changes due to external influences, plant populations are naturally competing for survival. As the competitive exclusion principle states that two species competing for limited resources cannot co-exist at constant population values, we require a model of population dynamics.

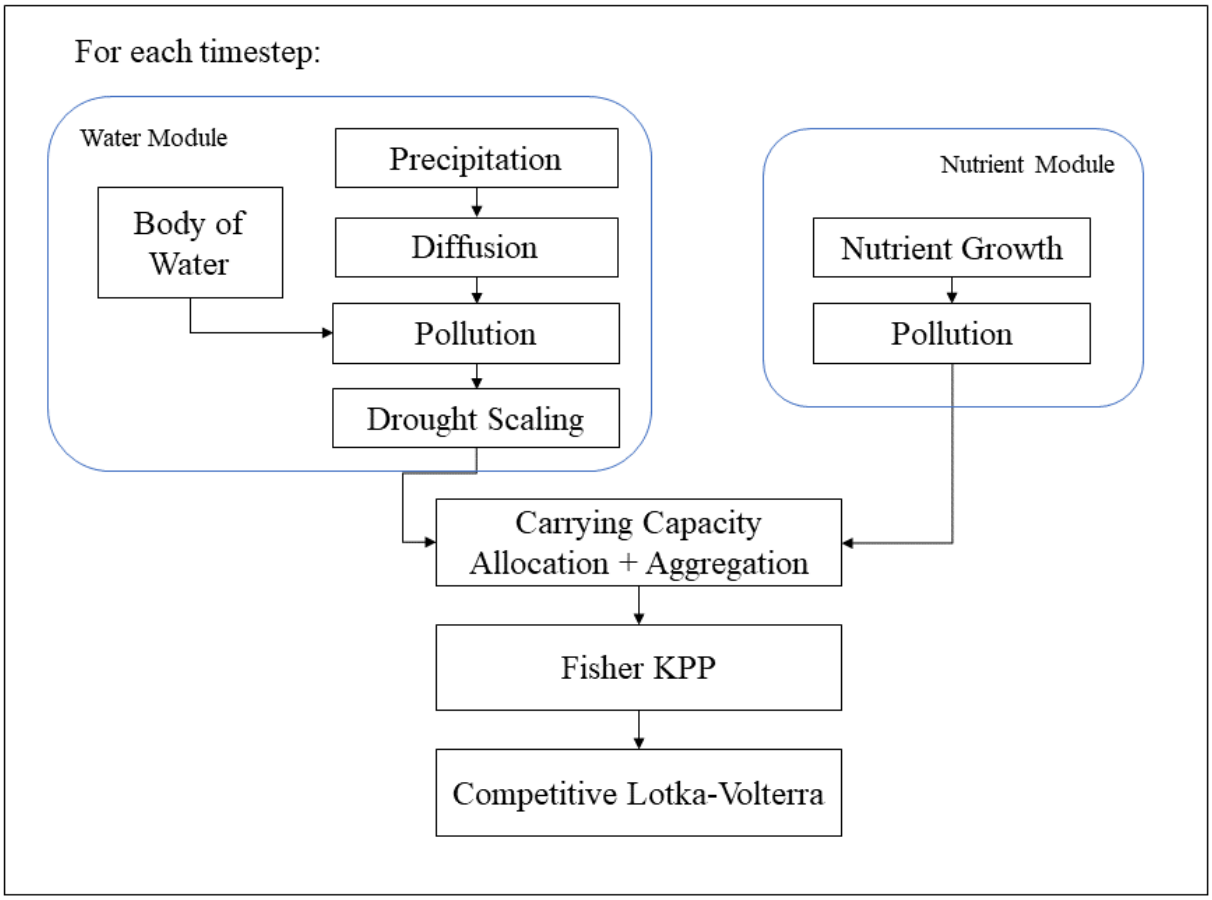


Figure 1: Flowchart of the forward methodology. The outer loop is run every timestep.

3.1.1 Competitive Lotka-Volterra Equations

Population models contain four critical demogrpahic processes of death, birth, immigration, and migration, which are all captured under Malthus' population principle of growth and represented mathematically in the logistic equation. As such, the base population model for competition is that of logistic population changes in time with terms that account for inter-species interactions. Explicitly, for N competing plant species the competitive Lotka-Volterra equations read

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = r_i x_i \left(1 - \frac{\sum_{j=1}^{N} \alpha_{ij} x_j}{K_i}\right) \tag{1}$$

where x_i is the population of species $i \in \{1, 2, \ldots N\}$, $\alpha_{i,j} \geq 0$ represents the influence that species j has on species i, K_i describes the carrying capacity for species i, r_i describes the inherent per-capita growth rate, and t is time. By convention, self-interactions are represented by the condition $\alpha_{ii} = 1$. As all species are modelled as competing, each α_{ij} represents an antagonistic interspecies effect with the severity determined by magnitude: $\alpha_{ij} \approx 1$ corresponds to a significant competitive effect while $\alpha_{ij} = 0$ represents no interaction.

The population counts, growth rates, and carrying capacities are considered as vectors while the α_{ij} terms determine a matrix. To ensure that our model represents the behavior of a non-chaotic system, we restrict possible species collections with $\{\alpha_{ij}\}_{1 \leq i,j \leq N}$ having all positive eigenvalues, a necessary condition for stability [AG04]. Thus, to initialize multi-species systems with varying interspecies effects, we require a method for constructing random symmetric positive definite matrices with unit diagonals, and all elements between $[0, 1]$. This is accomplished by computing $\alpha = XX^T$ where random unit vectors

comprise the rows of X (to see why these properties are guaranteed, one can prove it by definition of matrix multiplication and a quick invocation of the Cauchy Schwarz inequality).

While the competitive Lotka-Volterra equations represent population dynamics through time, we are also interested in the spatial propagation of species throughout Ω. To model this process, we couple the above with a reaction-diffusion equation.

3.1.2 Fisher Kolmogorov-Petrovsky-Piskunov (KPP) Equation

To model the species' diffusion throughout the domain Ω, we leverage the Fisher-KPP equation, a reaction-diffusion PDE that models population growth. Let $0 \leq u_i(x,t) \leq 1$ quantify the population density of species i. In 2D, the equation reads

$$\partial_t u_i - \mu_i \Delta u_i = \rho_i u_i \left(1 - \frac{u_i}{K_i}\right) \tag{2}$$

where $\mu_i > 0$ is the diffusivity coefficient, $\rho_i > 0$ is the growth rate, and $K_i > 0$ is the carrying capacity of the habitat for species i. We specify this equation for each species and subject the resulting system of PDE to the continuity condition

$$\sum_{i=1}^{n} u_i(x,t) = 1, \tag{3}$$

which ensures that each species is represented point-wise.

The dimensional variables are non-dimensionalized by the transformations $\tau_i = t/\rho_i^{-1}, \xi_i = x/\sqrt{\mu_i/\rho_i}$, and $\upsilon_i = u_i/K_i$ where ρ_i^{-1}, $\sqrt{\mu_i/\rho_i}$ and K_i are the time, length, and population scales for species i, respectively. In the dimensionless form, $\upsilon_i(\xi_i, 0) = (\upsilon_i)_0(\xi_i) \in [0,1]$ for $\xi_i \in \tilde{\Omega}_i$ as it is the ratio of the initial population to the carrying capacity. Here, $\tilde{\Omega}_i$ is the non-dimensionalized domain over which diffusion of species i occurs. As such, Equation 2 has the dimensionless form

$$\partial_{\tau_i} \upsilon_i - \Delta \upsilon_i = \upsilon_i(1 - \upsilon_i), \quad \xi_i \in \tilde{\Omega}_i, \tau_i \geq 0. \tag{4}$$

The Fisher equation is posed with no-flux conditions at infinity $\lim_{\xi_i \to \pm\infty}(\upsilon_i)_{\xi_i} = 0$. As the non-linearity presents a considerable numerical challenge, we linearize the PDE around $\upsilon_i(\xi_i, 0) = 0$ by setting $\upsilon_i(\xi_i, \tau_i) = \varepsilon\omega_i(\xi_i, \tau_i)$ for $0 < \varepsilon \ll 1$. This assumption implies a space equilibrium solution that satisfies the desired no-flux boundary conditions assuming the value of ω_i is small. Substituting, the resulting PDE is of the form $\varepsilon(\omega_i)_{\tau_i} - \varepsilon\Delta\omega_i = \varepsilon\omega_i - (\varepsilon\omega_i)^2$, which simplifies further to

$$(\omega_i)_{\tau_i} - \Delta\omega_i = \omega_i \tag{5}$$

when preserving terms of order $\mathcal{O}(\varepsilon)$. When this linearization is valid for small ω_i, the system is expected to attain a quasi-steady-state throughout the simulated domain. While several involved numerical schemes exist for this PDE, this linearization allows us to implement a simpler numerical scheme while implicitly assuming that each individual population has a small population density. We now discuss our approach to simulating resources in the environment, starting water.

3.2 Simulating Water

Our model accounts for the dynamics of 2 different sources of water: precipitation and large bodies of water. Modeling precipitation can be broken down into two stages: rainfall events (see subsubsection 3.2.1) and the diffusion of rainwater post-rainfall (see subsubsection 3.2.2).

3.2.1 Precipitation Events

We model precipitation events as a collection of spatially and time increment-dependent Poisson distributions with magnitudes independently and identically distributed (iid) according to another spatially

dependent Gamma distribution, mostly following [DNO18]. Formally, for each location (i, j) in Ω, let $N_{i,j}(\Delta t)$ denote the number of precipitation events that occur in a time increment of Δt and assume that each event that occurs has a magnitude distributed according to a Gamma distribution. The probability mass function of $N_{i,j}(\Delta t)$ is by definition

$$\mathbb{P}(N_{i,j}(\Delta t) = n) = e^{-\lambda_{i,j}(\Delta t)} \frac{\lambda_{i,j}(\Delta t)^n}{n!}, \tag{6}$$

and the total precipitation at the (i, j) location in a timestep of size Δt $P_{i,j}(\Delta t)$ is simply the sum

$$P_{i,j}(\Delta t) = \sum_{k=1}^{N_{i,j}(\Delta t)} Y_i, Y_i \sim \text{Gamma}(k_{i,j}, \theta_{i,j}). \tag{7}$$

One of the key properties of this model is that it captures days with no rainfall quite naturally - indeed, plugging in $n = 0$ to the above Poisson distribution gives

$$\mathbb{P}(N_{i,j}(\Delta t) = n) = e^{-\lambda_{i,j}(\Delta t)} \frac{\lambda_{i,j}(\Delta t)^0}{0!} = e^{-\lambda_{i,j}(\Delta t)} > 0. \tag{8}$$

Further, it follows that $\lambda_{i,j}(\Delta t)$ has an easy physical interpretation, as $\mathbb{E}[N_{i,j}(t)] = \lambda_{i,j}(\Delta t)$ - it is simply the number of precipitation events we *expect* to occur on a timestep of size Δt at location (i, j). To stay simple, $\lambda_{i,j}(\Delta t)$ scales linearly with respect to Δt. A hyperparameter representing the frequency $f_{i,j}$ of precipitation events is provided and then we simply compute $\lambda_{i,j}(\Delta t) = f_{i,j} \cdot \Delta t$.

This Poisson-Gamma approach provides an incredible amount of flexibility - the frequency matrix can be initialized in any manner and can even be a function of time, just like the parameters of the Gamma distribution. In particular, we can specify wetter and drier regions of Ω freely as the effect is internalized within this subroutine. One approach to simulating drought or abundance of precipitation events would just involve tweaking $\lambda_{i,j}$ for all (i, j) in some subdomain of Ω. We do not settle on this approach of modeling for various reasons (mostly numerical stability, see subsubsection 3.4.4), but it provides a strong physical justification and rooting for this submodule.

3.2.2 Precipitation Diffusion

Following precipitation, our model should account for the remaining components of the water cycle; in particular, the additional water resultant from precipitation should diffuse through the domain and, in some sense, by retained by the ground until eventually it is reduced through evapotranspiration. The 2d heat equation provides a natural avenue for capturing this behavior:

$$u_t = \alpha \Delta u = \alpha \left(u_{xx} + u_{yy} \right), \tag{9}$$

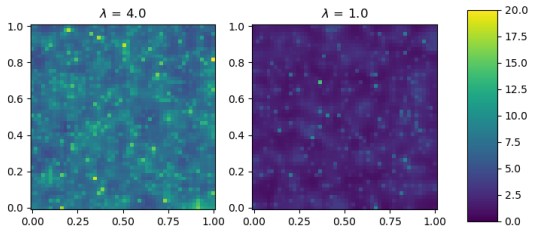
where in this case $u : \Omega \times \mathbb{R} \to \mathbb{R}$ represents the amount of rainwater at each location and Δ denotes the Laplacian operator. We numerically approximate a solution on Ω using a standard finite difference method approach detailed briefly in section 4.

Two example runs of combining the subroutines in subsubsection 3.2.1 and subsubsection 3.2.2 for different values of $\lambda_{i,j}$ are displayed in Figure 2a. To give visual intuition of a continuous time progression of the realization, one can imagine the grid “twinkling" or “sparkling."

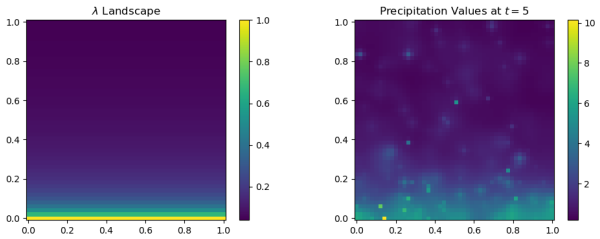
3.2.3 Bodies of Water

To model bodies of water, we introduce “point sources" into the landscape that constantly supply water. These point sources behave as literally just an additive factor, *except* in the presence of drought events. To represent an approximation “diffusion" of this constant source, we model the impact area of the body of water as a 2D Gaussian. Mathematically, we discretize the function

$$z(x, y) = Ae^{[(x-x_0)^2+(y-y_0)^2]/\sigma^2} + \epsilon, \tag{10}$$



(a) Precipitation example runs for 100 timesteps, $\Delta t = 0.01$, and two different $\lambda_{i,j}$ values (as an example constant for all i, j).



(b) Example of an interpolated nontrivial landscape and its results.

Figure 2: Water Simulation Examples

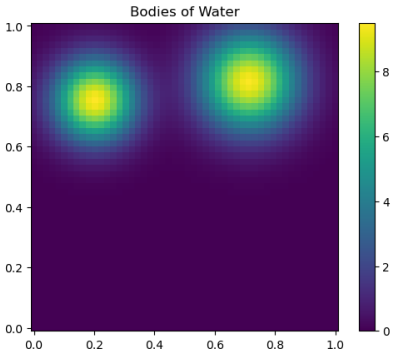


Figure 3: Example generation of bodies of water.

where A, σ, and (x_0, y_0) are randomly generated and represent the scaling magnitude, effective radius, and center location respectively. ϵ is a small constant so that the discretization appears a bit nicer.

Visually, an example manifestation of bodies of water is displayed in Figure 3.

3.3 Simulating Nutrients

We model the grid of available nutrients in a much simpler manner because of the assumption that the movement of nutrients is negligible. Our physical motivation is that we want the amount of nutrients to grow faster in some locations compared to others and otherwise asymptotically approach the maximum amount of nutrients that *could* be held at each location.

Thus, quite naturally, we attach a logistic growth model $Y_{i,j} : \mathbb{R} \to \mathbb{R}$ representing to units of nutrients available to each index (i, j):

$$Y_{i,j}(x) = \frac{M_{N,i,j}}{1 + e^{-\sigma_{i,j}x}}, \tag{11}$$

where $M_{N,i,j}$ is the maximum amount of nutrients that can exist in this location of soil and $\sigma_{i,j} > 0$ is a scale parameter representing how fast the logistic growth is. We restrict to positive $\sigma_{i,j}$ so that $Y_{i,j}$ is monotonically increasing.

x is some "location" that lives along an abstract real line that we will traverse over during iteration - we stress that there is little physical intuition to be had about values of x. Implementation wise, we need to support two operations: moving forward some time step Δt and growing nutrients, and updating the current location given some changes in the values due to other factors in the system such as consumption.

For the former, we keep it simple and simply move from $x \mapsto x + \Delta t$. We emphasize that the reader should not take this to mean x can be interpreted as time, as this interpretation falls apart when implementing the latter operation.

The latter operation can be formulated as the problem: given some new state value y', what value of x

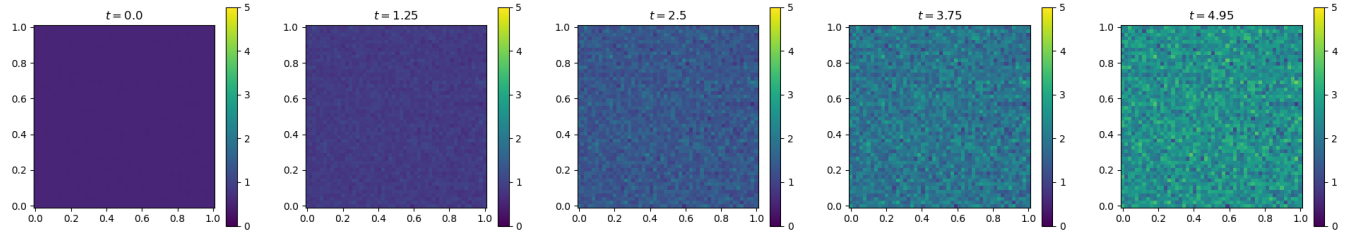


Figure 4: Example of nutrient growth, initialized with all values at 0.5.

are we at? We answer this by inverting $Y_{i,j}(x)$, which can be easily computed with some algebra as:

$$Y_{i,j}^{-1}(y) = \frac{1}{\sigma_{i,j}} \log \frac{y}{M_{N,i,j} - y}. \tag{12}$$

So given some new state value y', we update our $x \mapsto Y_{i,j}^{-1}(y')$. Clearly, this can and almost always will be a shift to the left (since the changes come from consumption), so interpreting x as time is not correct. An example is displayed in Figure 4.

3.4 End to End Simulation

We aggregate all of these previous computations and subroutines into formulating a notion of localized carrying capacity.

3.4.1 Carrying Capacity - Water + Nutrients

We consider two types of carrying capacities: Soil Water Carrying Capacity for Vegetation (SWCCV) and Soil Nutrient Carrying Capacity for Vegetation (SNCCV). For our model, we wanted two key properties of carrying capacity: (1) it must be species specific and (2) it must depend both on space and time. We investigate SWCCV first and define SNCCV in the same manner.

One of the most popular models used to determine SWCCV is the **water balance equation**. This is a simple compartment model that represents the inflow and outflow into a particular location on Ω. Assuming everything in the following equation is i, j indexed on the square mesh of Ω, we can model $W(t)$, the **surplus of water units** at (i, j) at time t, as

$$W(t + \Delta t) = W(t) + P(\Delta t) - E(\Delta t, k_t) - R(\Delta t), \tag{13}$$

where P is the precipitation as computed in subsubsection 3.2.1, E encapsulates all forms of evapotranspiration which is a function of the population k_t, and R is runoff. R depends on the Plant Available Water Holding Capacity (PAWHC) denoting the maximum amount of units of water this location of soil can hold onto. Operating under our assumptions (namely assumption 1), accounting for runoff is as simple as elementwise minimizing $W(t)$ with a matrix denoting the PAWHC at each point in the domain i.e. all runoff disappears.

This definition is slightly different than most literature, as we drop an additional soil surplus term and instead view W as the *surplus of water* rather than raw volume. So, a value of $W(t) = 0$ implies that the system is operating with no surplus while $W(t) > 0$ and $W(t) < 0$ imply the system is operating on a surplus or a deficit, respectively.

As a consequence, we can define SWCCV as the critical $K_W^*(t)$ such that

$$K_W^*(t) = \sup\{k_t : W(t + \Delta t) \geq 0\} = \sup\{k_t : W(t) + P(\Delta t) \geq E(\Delta t, k_t) + R(\Delta t)\}, \tag{14}$$

which we note is a simple supply-demand argument[2] and captures the same idea as the traditional definition, which defines $K_W^*(t)$ as the maximum population density at which water consumption is equal to the water supply i.e. $W(t + \Delta t) - W(t) = 0$, if W was instead the volume of water.

[2]Interestingly, this is not the only time economic-looking ideas appear in the carrying capacity computation - perhaps there is more to be fleshed out about a link between the two.

We impose the assumption that E is surjective onto $\mathbb{R}^+$, otherwise Equation 14 could evaluate to ∞. This is not an outrageous assumption at all and can usually be engineered to always be true in practice - in our case, we define E as the product of Δt and k_t with some rate d_s at which the species s depletes the resource (units are resource units per unit biomass per unit time).

We can define SNCCV similarly by analyzing a compartment model for nutrition (in this case much simpler):

$$N(t + \Delta t) = Y(x_t) - E(\Delta t, k_t) - R(\Delta t), \tag{15}$$

where N is the density of nutrients, Y is the nutrient function as defined in Equation 11, and E and R are defined analogously as in 13. x_t denotes the abstract location on the real line that we are at the time t.

Thus, SNCCV is $K_N^*(t)$ such that

$$K_N^*(t) = \sup\{k_t : N(t + \Delta t) \geq 0\} = \sup\{k_t : Y(x_t) + R(\Delta t) \geq E(\Delta t, k_t)\}, \tag{16}$$

again making similar assumptions about E.

3.4.2 Carrying Capacity - Allocation

Now, we must convert both $K_W^*(t)$ and $K_N^*(t)$, which are the amounts of "biomass" at which the surplus of resources is fully consumed, into specific values for the species $K_{W,s}^*(t)$ and $K_{N,s}^*(t)$ which will be aggregated and plugged into the species specific Equation 1. We pose the following scheme based on the idea of "constructing an optimal allocation," inspired remotely by auctions and uses principles from information theory. We will use $K_W^*(t)$ as the example but the exact same procedure is applied to $K_N^*(t)$ in practice.

We firstly notice that instead of thinking about the problem as allocating units of biomass, we transform it into a more concrete problem of allocating units of the resource, in this example water. Namely, we instead consider allocating the **available preconsumption resources** to the species, or the quantity

$$C_W(t) = W(t) + P(\Delta t) - R(\Delta t). \tag{17}$$

If we follow the mentioned surjective assumption, this is also the same as $E(\Delta t, K_W^*(t))$, as the $\sup$ in Equation 14 is always attained.

Primarily, we should be allocating these resources in a manner that captures the *ability to expand* of each species... or the growth rate! A bigger (relatively) growth rate means the species expands more aggressively, so we allocate more of preconsumption resource to it, while a more passively growing species should get less of the surplus.

The trick is to convert the vector of growth rates $\mathbf{r}$ into a probability mass function (pmf) through a transformation. We considered two approaches - firstly, the most obvious transformation is to normalize by the sum of all growth rates i.e. compute

$$p_i = \frac{r_i}{\sum_j r_j}. \tag{18}$$

This does normalization properly but problematically results in polarized values that cause instabilities in our numerical solvers for the PDEs.

Instead, we take inspiration from maximum entropy distributions in information theory and choose to take a softmax function (equivalently the Gibbs/Boltzmann distributions in statistical mechanics):

$$p_i = \frac{e^{r_i}}{\sum_j e^{r_j}}. \tag{19}$$

To see how creating the pmf this way is better, consider a case where $\mathbf{r} = (0.01, 0.9)$. If we allocate according to rule 1, we get that species 1 would receive roughly 1% of the resources. Alternatively, if we

compute as a softmax, species 1 would instead receive around 29% of the allocation. Simply, using the softmax allocation instead of the summation based normalization reduces the extremity of the allocations so that extremely lopsided allocations do not occur as often. This helps smooth out the changes in the carrying capacity over time,especially when the number of species gets larger, which implicitly smooths out and helps avoid stiff gradients.

After obtaining the pdf, the rest is straightforward: we allocate $p_s C_W(t)$ of the resources to species s, and then divide by their consumption rate to get the carrying capacity in terms of units of biomass. In symbols:

$$K^*_{W,s}(t) = \frac{p_s C_W(t)}{d_{W,s}}, \tag{20}$$

where $d_{W,s}$ denotes the rate at which species s consumes W.

Clearly, you can repeat this routine to obtain $K^*_{N,s}(t)$ for all s as well.

3.4.3 Carrying Capacity - Aggregation

We wish to now aggregate $K^*_{W,s}$ and $K^*_{N,s}$. Two methods were proposed, both with reasonable physical interpretations, so we ended up choosing the one that made the numerics tractable.

The first aggregation method is to simply take

$$K^*_s = \min(K^*_{W,s}, K^*_{N,s}). \tag{21}$$

Borrowing from stoichiometry, the intuition is that one of the resources becomes a sort of “limiting reagent" and the carrying capacity cannot exceed that value. Unfortunately, $\min$ can cause extremely ragged and noisy carrying capacities, resulting in stiff gradients that cannot be properly computed and modeled by numerical algorithms. Instead, we adopt an **average aggregation method**, or

$$K^*_s = \frac{K^*_{W,s} + K^*_{N,s}}{2}. \tag{22}$$

A strong argument for why average aggregation is actually better than the limiting aggregation method is if we considered more than just 2 resources. If we wanted to extend the model further, it becomes very clear that minimum is probably no longer the true carrying capacity - instead, average is a better approximation as it combines all the values rather than just throwing them out. Species can also evolve to substitute certain nutrients for others in a sort of "multicollinearity" effect.

Overall, the big takeaway is that doing the aggregation through averaging not only makes the numerics work but has physical merit and generalizes/scales to larger simulations better.

An example carrying capacity at an arbitrary timestep is shown in Figure 5. See the caption and 3.4.5 for a more in depth explanation of some of the features, especially the dark black box and blob of high K^*_Ws.

3.4.4 Extreme Precipitation Events

There are a lot of methods ranging from simple to complex on how to simulate an Extreme Precipitation Event (EPE). As aforementioned, we definitely could have gone into each component of the model and tweaked individual parameters, such as tweaking the $\lambda_{i,j}$ of each Poisson distribution. However, we avoided this approach because it would involve tuning even more parameters. In particular, it makes simulating contrived environments difficult, as we notice later on that our ecosystem was sensitive to input parameters - anything that was not in a relatively balanced and competitive initial state usually results in 1 species dominating.

So, for an EPE, we instead accumulate all minute effects of the shift in weather into one “weather constant" γ_{EPE} that is constant but unique to this EPE and scale the production of water accordingly based

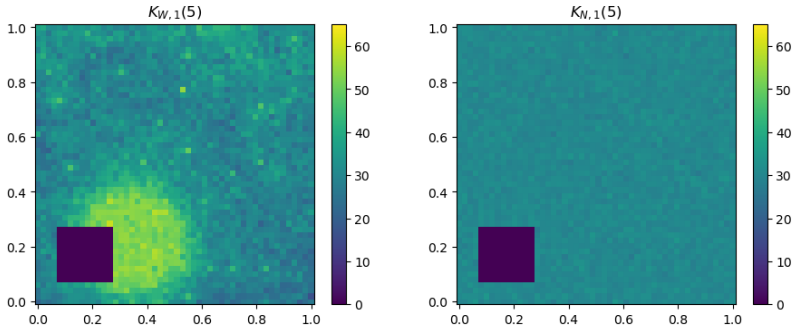


Figure 5: Example precipitation (left) and nutrient (right) carrying capacities. The blob of high water carrying capacities is a body of water. The dark block in both is an example of a pollutant mask - in this case a very simple one that zeros out the resources in that area, so the carrying capacity is likewise 0. This example is explored further in many of the later sections, especially section 6.

on how long has passed since the start of the EPE. We want the growth to be proportional to the existing amount, so we choose to multiply the water surplus by a term that is γ_{EPE} to a time dependent exponent. We formalize this by beginning with a few definitions.

Definition 3.1 (Interpolation Methods)**.** *Let $I = [\ell, r], 0 \leq \ell \leq r$ be the current EPE interval. An **interpolation method** is any mapping $t_{diff} : I \to [0, \infty)$.*

***V-interpolation** is defined as*

$$t_{diff}(t) = \begin{cases} t - \ell & t \leq \frac{\ell+r}{2}, \\ r - t & t > \frac{\ell+r}{2}. \end{cases} \tag{23}$$

***Half Valley Interpolation** is defined as*

$$t_{diff}(t) = \begin{cases} t - \ell & t < \frac{3\ell}{4} + \frac{r}{4}, \\ \frac{3\ell}{4} + \frac{r}{4} - \ell & \frac{3\ell}{4} + \frac{r}{4} \leq t \leq \frac{\ell}{4} + \frac{3r}{4}, \\ r - t & t > \frac{\ell}{4} + \frac{3r}{4}. \end{cases} \tag{24}$$

The shapes of these two methods are displayed in Figure 6. Now, assuming we working on a localized scope in both the species sense and the location sense, we define the **modified water input** $C'_W(t)$ as

$$C'_W(t) = C_W(t) \cdot \gamma_{EPE}^{t_{\text{diff}}(t)}, \tag{25}$$

where t_{diff} is any interpolation method - we restricted ourselves to using v-interpolation and half valley interpolation, but any methods work. Because this entire scaling is localized, we note the flexibility of our model to simulate extreme precipitation events in a subset of the domain - we simply restrict the effect of Equation 25 to this subset. After computing this modified input, we compute a **modified carrying capacity** accordingly using the allocation and aggregation in subsubsection 3.4.2 and subsubsection 3.4.3 but applied to $C'_W(t)$.

3.4.5 Pollution

The way our structure is set up (Figure 1) permits extremely easy access to introducing pollution and other environmental degradation. We simply can model them as an additional “mask" applied elementwise to the water and nutrient values. So, we can compute water as

$$C_W(t) = B \odot W(t) + A \odot P(\Delta t), \tag{26}$$

where $\odot$ denotes the element-wise Hadamard product of matrices.

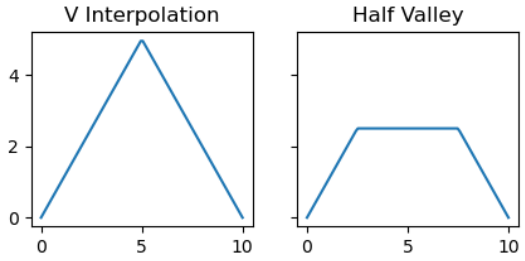


Figure 6: Shapes for different drought exponent interpolation methods, see Definition 3.1.

So, an example we use consistently is imagining the construction of a building or parking lot on a subset of Ω, such that there are no longer resources there. We call this the **parking lot pollutant** from now on. The parking lot pollutant's mask matrix would look like a block matrix of 0s in the subdomain and 1s everywhere else. For example, on a small grid of 5, it would look like

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{27}$$

In terms of implementation details, for big pollution events i.e. ones that make drastic changes in certain regions of the domain, we do need to have a bit of a warmup period in order to smooth out the gradients a bit and make the problem computable. But, we do note that other than this minor caveat of potentially creating a gradient that is too large for the PDE solvers, this framework of introducing pollutants allows for *any* type of pollutant one could want to add simply by specifying the behavior of a certain mask matrix.

3.5 Implementation

The surprising fact of this section is that everything can be very trivally implemented - the vast majority of operations are simple matrix operations and/or sampling from distributions. We were able to implement the entire simulation in pure `NumPy` and Python.

Letting $|S|$ be the number of species and G be the grid size, the complexity of the simulation to run T timesteps is $\mathcal{O}(|S|TG^2)$, which is not great but is still computable for reasonable parameters. Speedups (namely parallelization) to mitigate this nearly quartic complexity are discussed in section 7.

4 Numerical Schemes

The above discussion constitutes a system of N (which is how much?) coupled time-dependent ordinary and partial differential equations which we evaluate numerically. As all equations are time dependent, and we are primarily interested in modelling population dynamics, a global time stepper for integrating the competitive Lotka-Volterra equations and individual time evaluations within a single time step comprise the full solver. We employ the explicit Runge-Kutta method RK45 to evaluate time steps while embedding updates from other models within each time step. The choice of RK45 is based on its theoretic stability and convergence properties. Overall, the numerical scheme may be summarized with the following diagram:

4.1 Integration and Time Stepping through Competitive Lotka-Volterra Equations

The outer loop of our model consists of the system of competitive Lotka-Volterra equations, which are integrated through time by RK45 to determine new population counts. We treat the time stepping through

these equations as the global time while other system effects are incorporated within a time step to change species-specific parameters.

We briefly review the scheme for a system of N unknowns and N ODEs given by the linear system $\mathbf{v}'(t) = A\mathbf{v}(t)$. Let k be the times step size and t_n be the adaptively determined step size. The approximation $\mathbf{w}^n \approx \mathbf{v}(t_n)$ is obtained by solving

$$\mathbf{w}^{n+1} = \mathbf{w}^n + \frac{1}{6}(m_1 + 2m_2 + 2m_3 + m_4) \tag{28}$$

where

$$m_1 = kA\mathbf{w}^n \tag{29}$$

$$m_2 = kA\left(\mathbf{I} + \frac{k}{2}A\right)\mathbf{w}^n \tag{30}$$

$$m_3 = kA\left(\mathbf{I} + \frac{k}{2}\left(\mathbf{I} + \frac{k}{2}A\right)\right)\mathbf{w}^n \tag{31}$$

$$m_4 = kA\left(\mathbf{I} + kA\left(\mathbf{I} + \frac{k}{2}A\left(I + \frac{k}{2}A\right)\right)\right)\mathbf{w}^n. \tag{32}$$

4.2 Finite Difference Discretization of Water Diffusion

As mentioned in section subsubsection 3.2.2, the chosen model of water precipitation diffusion is the heat equation with diffusivity parameter α. We numerically evaluate this equation using a standard finite difference scheme, which we describe in brief here. For $\Delta x, \Delta y$ representing node spacings in the x and y directions, the differential operators are discretized using centered differences:

$$\partial_t P - \alpha\Delta P = \frac{P_{i,j}^{k+1} - P_{i,j}^k}{\Delta t} - \alpha\left[\left(\frac{P_{i-1,j}^k - 2P_{i,j}^k + P_{i+1,j}^k}{\Delta x^2}\right) + \left(\frac{P_{i,j-1}^k - 2P_{i,j}^k + P_{i,j+1}^k}{\Delta x^2}\right)\right] = 0 \tag{33}$$

where $P_{i,j}^n$ denotes the approximation of $P(x,t)$ at time step n and gridpoint $x_{i,j}$. Choosing a uniform mesh such that $\Delta x = \Delta y = h$, the above scheme is equivalent to

$$P_{i,j}^{k+1} \approx \left(1 - \frac{4\Delta t\alpha}{h^2}\right)P_{i,j}^k + \Delta t\alpha\left(\frac{P_{i,j-1}^k + P_{i-1,j}^k + P_{i+1,j}^k + P_{i,j+1}^k}{h^2}\right) \tag{34}$$

which yields the Courant-Friedrichs-Levy condition $\Delta t \leq \frac{h^2}{4\alpha}$.

We can easily do compute this at each time step on the interior of the domain. To determine the boundary conditions, we note that would like to simulate this ecosystem as a small piece in a larger ecosystem, which can be done by imposing zero Neumann conditions on $\partial\Omega$. To discretize this, suppose (i,j) was a location on $\partial\Omega$ and $i = 0$ (everything analogously holds for $i, j \in \{0, n-1\}^2$). We can approximate $(\partial_x P)$ on the boundary by

$$0 = (\partial_x P)_{0,j}^t \approx \frac{P_{-1,j}^t - P_{1,j}^t}{2\Delta x} \implies P_{-1,j}^t = P_{1,j}^t. \tag{35}$$

Thus, we can pad the discretized approximation with a layer of *ghost cells* whose values exactly match the values 1 layer in from the boundary. To move a timestep forward, we can simply update all locations including $\partial\Omega$ using the update rule in Equation 34.

4.3 Method of Lines Approximation of Linearized Fisher-KPP

As the time stepping of the system is adaptively managed by RK45, it remains to spatially approximate the solution to the non-dimensionalized linearized Fisher-KPP PDE of the form $u_t = \Delta u + u$ with

homogeneous Dirichlet 0 boundary conditions. Recall from subsubsection 3.1.2 that this PDE is solved N times for each species population density while subject to the continuity condition Equation 3. The spatial derivatives are replaced with the finite difference

$$\Delta u + u \approx \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} + u_{i,j} \tag{36}$$

on a uniform grid of width h consisting of M^2 grid points. The PDE is thus approximated by a system of M ODEs in the independent variable t which is integrated in time at each grid point. The boundary values are determined by the no-flux conditions of Fisher-KPP, specifically that the boundary terms in the domain are set to 0 while the initial condition is determined by the population density in the domain at the preceding time step. Explicitly, the linear system of ODEs is posed in $M + 1$ unknowns and $M + 1$ equations as $\mathbf{u}'(t) = A\mathbf{u}(t)$ where

$$\mathbf{u}(t) = \begin{bmatrix} u_0(t) \\ u_1(t) \\ \vdots \\ u_M(t) \end{bmatrix}, \quad A = \begin{bmatrix} 4 - h^2 & -1 & -1 & \cdots & -1 \\ -2 & 4 - h^2 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & -2 & 4 - h^2 \end{bmatrix}. \tag{37}$$

The above scheme is evaluated on each component of the population densities during a single time step of the outer integration loop of the Lotka-Volterra equations. To perform the time integration, we employ a second RK45 time stepper to integrate Fisher-KPP for each species' population densities over $[t_n, t_{n+1}]$ where t_n and t_{n+1} are the current and next times determined by the Lotka-Volterra integrator, respectively. The primary motivation for the second integrator is the adaptivity of time stepping, as this system develops stiff gradients and may lead to non-convergence of the method overall. At the expense of increased computational overhead, this combination yields a stabler numerical scheme.

To evaluate the validity of the discretization for the Fisher equation, we provide Figure 7, which confirms the results of Tang and Weber [TW91]. Specifically, due to a time-dependent balance of the diffusion and linear term, a local perturbation grows to a steady-state solution that replaces the quasi-steady state gradually. Concretely, this implies that the state $u = 0$ in Ω is replaced by the state $u = K$ when considering dimensionalized values. In the following, we perform 100 time steps in the range $[0, 10]$ of the competitive Lotka-Volterra model for two species coupled with the Fisher-KPP equation for both population densities. The initial conditions are specified by

$$u_1(x, 0) = 80 \exp^{-((x-0.5)^2+(y-0.5)^2)}, u_2(x, 0) = 20 \exp^{-((x-0.5)^2+(y-0.5)^2)} \tag{38}$$

which are two normally distributed Gaussians centered at the centroid of the square domain Ω. The growth rates are set to $(r_1, r_2) = 0.72, 1.27$, carrying capacities to $(K_1, K_2) = (60, 100)$, and $\{\alpha\}_{0,1} = 0.1, \{\alpha\}_{1,0)} = 1$ where $\{\alpha\}_{ii} = 1$ by convention. We observe the rapid diffusion of population 2, which is initialized far below its carrying capacity and thus experiences significant growth from the competitive LV equation. For comparison, we provide a reference solution of only the competitive LV equations for these initializations to illustrate the differences in solution. In particular, while competitive LV uniformly distributes the initial populations, our model more realistically captures the transfer dynamics of the populations. Although the solved PDE is linearized, it is clear that the structure is preserved when non-dimensionalizing, meaning for this initialization, the linearization is valid.

5 Simulating Multi-Species Interactions with Environmental Stress

Using the model described above, we perform a variety of simulations to analyze the long-term behavior of multi-species habitats under a variety of environmental effects to explore the question of biodiversity increase due to species symbiosis. A theoretical limitation on the fidelity of the simulation comes from Smale's 1976 [Sma76] theorem which states that the competitive Lotka-Volterra system is compatible with any dynamical behavior in general for large numbers of N. In other words, even under the requirement on α,

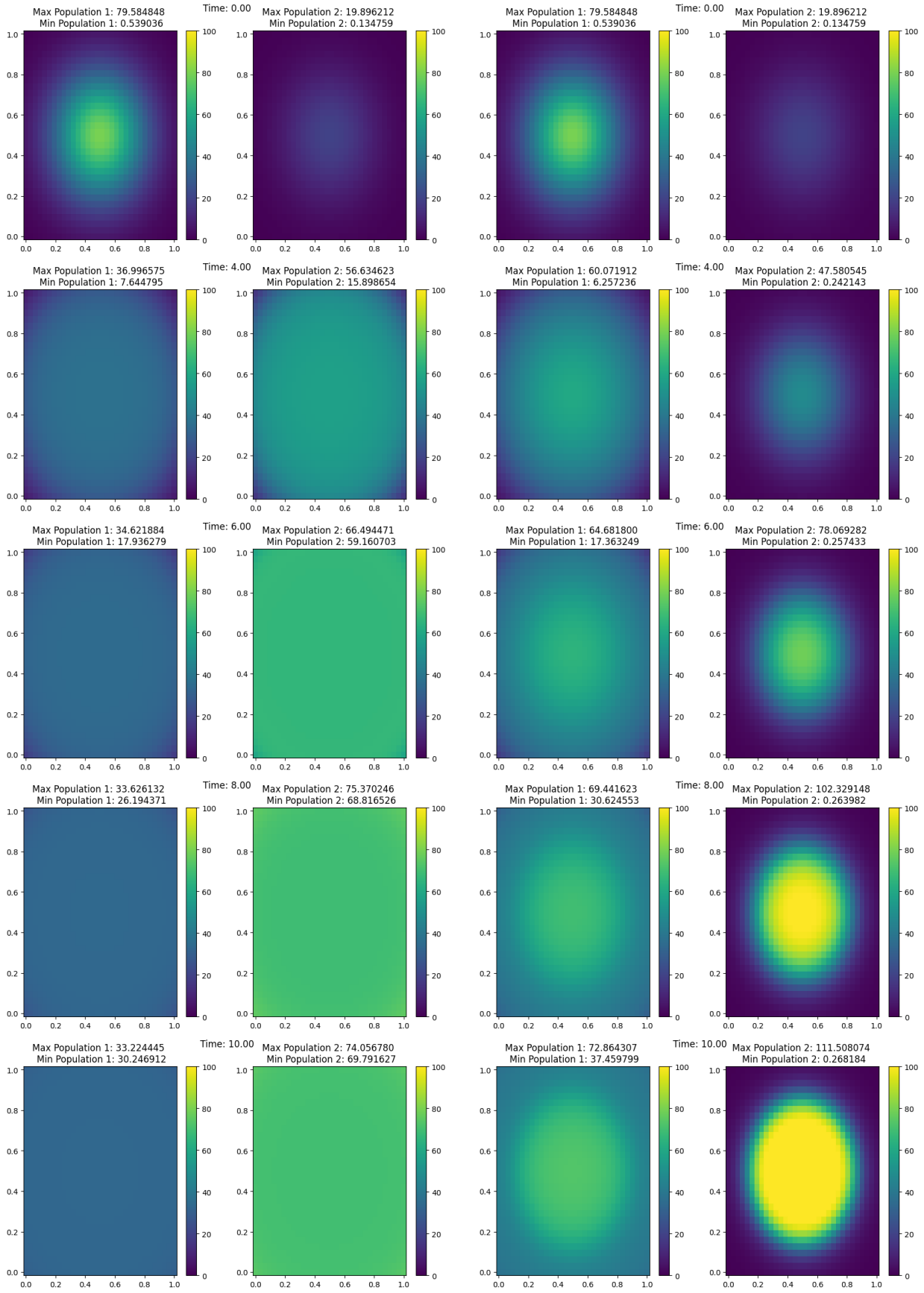


Figure 7: Diffusion of population and interface propagation of two populations initialized as univariate Gaussians. Left: classical competitive LV equation with the same initialization and parameters, Right: Coupled competitive LV and Fisher-KPP diffusion illustrating physically reasonable population diffusion throughout the unit square.

5.1 Quantifying Biodiversity

Firstly, we must attribute some actual meaning to biodiversity. As intuition might serve, we would like to say that a fairly even population distribution is more biodiverse while particularly uneven distributions, perhaps those having only 1 or 2 species in abundance out of several more in total, would be less biodiverse; we will propose using the *true diversity* metric or *Hill numbers* of order $q = 2$. [CCJ16] For order q, the formula is given as

$$ {}^{q}D := \frac{1}{\sqrt[q-1]{\sum_{i=1}^{N} p_i \left(p_i^{q-1}\right)}} \tag{39} $$

where N is the total number of types in the dataset and $p_i := \frac{P_i}{\sum P_j}$ the *proportional abundance* of the i'th type (in our case, a type $\sim$ a species), which we have taken to mean the odds of, over a uniform distribution assuming replacement, selecting species i. The interpretation is somewhat straightforward, though dependent on q: we read it as the 'effective number' of types s.t. more abundant types are given more weight in the generalized average as q is increased. In the $q = 1$ case, the measure ${}^{1}D$ can be read as the effective number of common/ typical species; in the $q = 2$ case, the measure can be read as the effective number of dominant/ very abundant species in the ecosystem; beyond $q \geq 3$, there is little additional meaning associated to particular values and the variations with increases in q become increasingly marginal.[CCJ16] While it may seem sensible at face value to work towards a plot of ${}^{q}D$ against q, consider the discontinuity at $q = 1$ and that the marginal information gain of larger q becomes almost meaningless very quickly; it is for these reasons that we will choose to work with ${}^{2}D$, allowing us to penalize particularly sparse and particularly abundant species while retaining intuition on the interpretation of the numbers. Additionally, for $q = 2$, we regain the inverse Simpson index.

Lastly, consider the following optimization problem

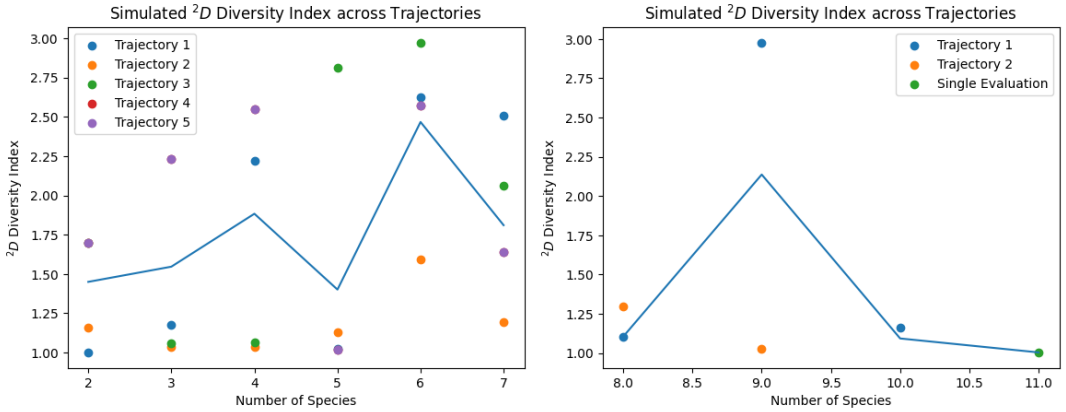
$$ \arg\max_{N \in \mathbb{N}} \frac{1}{k} \sum_{i=1}^{k} {}^{2}D\left[\mathbf{S}_{k;\Omega_k;W_k;N}\right] \tag{40} $$

for a sequence of simulations S_k for N-species dynamics given initial conditions Ω_k and subject to weather and pollution patterns W_k with k sufficiently large. Though computing this directly may itself not be especially meaningful given the high dimensionality, susceptibility to chaotic dynamics, and stochasticity, the process of doing so will certainly paint a picture of the relation we are attempting to capture.

5.2 Symbiotic Effects in Dynamic Environments

We first analyze symbiotic effects emerging in habitats with varying numbers of species subject to the same antagonistic effects from the environment. To do so, we simulate environments in which there are irregular weather cycles including periods of drought reducing water availability in regions of the domain, reduction of habitat rendering regions of the domain uninhabitable, and pollution effects which deplete the environment's available resources. Keeping these stressors the same against studies, we vary the number of species present in the habitat, randomly determine the positive definite matrix of inter-species interaction coefficients α_{ij} and growth rates.

Figure 8a is a sampling of 5 random trajectories of evolution for the system along which time-averaged biodiversity scores ${}^{2}D$ are computed and a plot of the average ${}^{2}D$ scores is plotted. Figure 8b represents a smaller trajectory of the same process with a larger number of species. As antagonistic environmental events are spatially and temporally random and species parameters are randomized, this sampling is representative of the overall dynamics of the system as the number of plant species increases. The figures depict an oscillatory behavior of the biodiversity metric with respect to the number of species present



(a) Trajectories for $N = 2, \ldots, 7$ (b) Trajectories for $N = 8, \ldots, 11$

Figure 8: Simulated ^2D across various trajectories. We observe a spike at 6 species.

in the sample, though an apparent trend is that a considerable increase in biodiversity is observed when $N = 6$, which is true across all trajectories. A similar effect can be observed for $N = 4$, though it is less significant both in terms of size of increase and frequency of occurrence.

In line with the problem statement, the number of required species for a symbiotic inter-species effect to occur is at least $N = 4$, where a steep increase in the biodiversity index is observed on average. A clearer trend is that at $N = 6$, the simulated plant community undergoes a greater symbiotic effect and approaches a population equilibrium which is stable under external stresses. On the other hand, the steep reduction in ^2D for $N = 5$, which occurs between two steep increases at $N = 4, 6$ suggests that there is not a clear correlation between the number of species and the resulting biodiversity index. Indeed, Figure (b) confirms a similar oscillatory behavior where the biodiversity index increases dramatically at $N = 9$, though at a lesser extent than at $N = 6$ and diminishes significantly to the theoretical minimum at $N = 11$. This oscillatory biodiversity behavior suggests that a raw count of the number of species in an environment is not an adequate proxy for the stability of the ecosystem, though for certain values of N (at $4, 6$, and 9) there are weak indicators of a positive correlation between N and ^2D.

5.3 Species Antagonism and Interaction Structure

The specification of the interaction matrix and intrinsic growth rates for each species constitute two factors by which inter-species effects can be studied across a range of scenarios representing different species. As positive definiteness of the α interactions matrix is a necessary condition for stability of the ecosystem, we only consider scenarios in which this requirement is upheld.

5.3.1 Theoretical Equilibrium Point Convergence

Previous theoretical analyses of the competitive Lotka-Volterra system have yielded theoretical equilibrium points for the system wherein different species may co-exist under a normalization of populations to $0 \leq x_i \leq 1$ [SWA05]. Specifically, if each species is identical in its interactions with exactly two other species and has no interaction with other species in such a way that each row of the inter-species interaction coefficients matrix α is an index permutation around a circle of $\langle 1, \alpha_1, 0, 0, \alpha_2 \rangle$, then it is known that the system attains a coexisting equilibrium point given by

$$\mathbf{x} = \frac{1}{\alpha_1 + \alpha_2 + 1} \tag{41}$$

meaning each species attains the same population count. Using the full model with randomly initialized population data in the range $(0, 1)$, interaction matrix

$$\alpha = \begin{bmatrix} 1 & 0.25 & 0 & 0 & 0.5 \\ 0.5 & 1 & 0.25 & 0 & 0 \\ 0 & 0.5 & 1 & 0.25 & 0 \\ 0 & 0 & 0.5 & 1 & 0.25 \\ 0.25 & 0 & 0 & 0.5 & 1 \end{bmatrix} \tag{42}$$

with uniform growth rates of 0.1, two drought events of order $0.4, 0.5$ at two disjoint periods of time $[0.1, 0.4] \cup [0.7, 0.9]$ at distinct subregions, the average populations for each species reaches an equilibrium of ≈ 0.122 with exact values given by $[0.1193, 0.1216, 0.1221, 0.1224, 0.1261]$ indexed by species number. Although the theoretical coexisting equilibrium point for these interaction coefficients is not attainable due to the external stresses experienced by our simulated population, we nevertheless obtain results of a stable co-existence equilibrium for the system wherein there is substantial spatial overlap between populations and they all attain approximately the same equilibrium population size. The theoretical equilibrium population is approximately 0.571, but the introduction of irregular drought cycles across the domain leads to considerable long-term system convergence population dampening. We perform a longitudinal study of the equilibrium convergence an analogous series of simulations for $N = 5, 6$ with variable column sums to quantify different levels of inter-species activity.

Table 1: Difference between Simulated and Theoretical Equilibrium Populations

Species Count	$\alpha - \mathbf{I}$ Column Sum				
	0.1	0.25	0.50	0.75	0.95
5	-0.788	-0.691	-0.522	-0.449	-0.365
6	-0.511	-0.364	-0.155	-0.121	8.179e-3

In Table 1, we provide the difference between the simulated and theoretical equilibrium equilibrium populations attained by the system with identical external stressors. In each evaluation, α is a positive definite matrix with a unit diagonal with the additional property that $\alpha - \mathbf{I}$ is a specified value ($\mathbf{I}$ is the $N \times N$ identity matrix). These scenarios are evaluated along a single trajectory of the system with identical stressors. The apparent trend is that the addition of another species enables the simulated system to converge closer to the theoretical equilibrium. Although in general, the stressed equilibrium populations are substantially smaller than in the idealized case, we can nevertheless consider this as numerical evidence that a symbiotic relationship allows a system to converge to a state closer to the theoretical equilibrium.

We perform a second study with the same set-up to quantify the differences in simulated biodiversity. Notice that in the theoretical equilibrium with N species, each has a proportion abundance of $1/N$ and maximizes ${}^2D$ to the value of N.

Table 2: ${}^2D$ Evaluations for Simulated Equilibria

Species Count	$\alpha - \mathbf{I}$ Column Sum				
	0.1	0.25	0.50	0.75	0.95
5	4.469	3.916	4.265	4.506	4.321
6	5.208	5.312	5.195	4.624	3.605

In the above, we have collocated the computed ${}^2D$ values for each simulated system where higher values are better. In each case, N is the theoretical equilibrium maximum and the above quantifies the difference between the simulated and theoretical equilibrium biodiversities. A key observation is that the equilibrium biodiversity differences for a larger N are also substantially larger for every initialization of α. Thus, while increasing the number of species improves the simulated population equilibrium, the biodiversity of the resulting system is inferior to the case with fewer species in some situations, especially when one pair of inter-species interactions is very antagonistic (specifically at $\alpha - \mathbf{I} = 0.95$).

However, our model indicates that the biodiversity index decreases monotonically as two species compete more strongly (as $\alpha - \mathbf{I} \to 1$) and increases considerably as the competitive effects decrease. Additionally, while the biodiversity index relative to the theoretical maximum may not necessarily improve with the addition of a new species, Table 2 strongly indicates that decreasing competitive effects in the system *ceteris paribus* leads to an improved biodiversity index.

Comparing the two tables suggests a trend with respect to the addition of a new highly-competitive species. In particular, the number of species may reduce the difference between the simulated and theoretical equilibrium populations at the expense of biodiversity in the environment. This is reminiscent of Chris D. Thomas' theory of introducing invasive species to unstable ecosystems with the aim of regenerating other species by creating a suitable habitat or food source for existing populations.

5.3.2 Interesting Species Types

We now employ our simulation to investigate the interactions of species that have a distinctive type. In this small case study, we consider a system with 3 species parameterized as follows. We consider a specific α matrix:

$$\alpha = \begin{pmatrix} 1 & 0.1 & 0.2 \\ 0.4 & 1 & 0.3 \\ 0 & 0 & 1 \end{pmatrix}. \tag{43}$$

Namely, species 3 is a unique type: we can consider it as a very "resilient" agent as the effects of the other species on it are 0. We use growth rates as $\mathbf{r} = (0.5, 0.7, 0.5)$. Finally we also specify the demand vectors as

$$\mathbf{d}_W = (0.07, 0.07, 0.08), \mathbf{d}_N = (0.06, 0.06, 0.045). \tag{44}$$

So, not only is 3 very resilient, it is more reliant on water than the other agents but much less reliant on nutrients. We will observe that this setup dampens the effect of droughts on species 3. The results are displayed at various times in Figure 9.

Even though the growth rate is the same as species 1 and the water demand higher, it ends up outperforming because of its resilience and lower nutrient demand - it serves as verification that our model can capture these behaviors that are observed in real environments.

5.4 Varying Drought Frequencies

The previous section's analysis runs the problem with some minor droughts occuring throughout. In this section, we run a similar simulation, keeping the grids and seeds the same, but with 4 drought events occuring sparsely throughout the domain and on various subsections of the domain; their types, interval, γ_{EPE}, and location specified by upper left (UL) and lower right (LR) indices (on the mesh) are displayed in Table 3.

We run the simulation 5 times for the same seeds as before with $n \in \{2, 3, 4, 5, 6\}$ to observe the effect of drought on ${}^2D$. Unfortunately, the simulation is slow without parallelization beyond $n = 7$ - this is discussed in the shortcomings. The results are plotted in Figure 10 along with a mean line.

Certainly, it is still apparent that having more species up to a point seems to contribute to greater diversity scores, and based on our random seeding the optimal value seems to between 5 and 6 species.

Type	Interval	γ_{EPE}	Location
Half Valley	$[0.2, 0.4]$	0.2	UL: $(10, 10)$, LR: $(15, 15)$
Half Valley	$[0.0, 0.3]$	0.4	UL: $(20, 20)$, LR: $(30, 30)$
Half Valley	$[0.5, 0.7]$	0.3	UL: $(10, 20)$, LR: $(15, 25)$
Half Valley	$[0.7, 1.0]$	0.3	UL: $(0, 0)$, LR: $(15, 15)$

Table 3: Drought configurations

However, only having 5 trajectories is certainly not enough to invoke the strong law of large numbers, yet we can clearly observe the desired relationship - with more species, biodiversity is more robust to droughts and extreme precipitation. Furthermore, the results

5.5 Long-term Species Stabilization

In the previous sections, we have explored issues of population dynamics by considering a variety of weather, pollution, and habitat reduction effects while changing the number of co-existing species and their intrinsic properties. The preceding analyses have suggested an ecosystem management strategy that may rescue declining plant populations from extinction [Zay07]. Specifically, the introduction of invasive new species into the ecosystem that interact weakly with all existing plant populations which in the equilibrium case may stabilize the population dynamics to regenerate. Although this oftentimes comes with a negative effect on the biodiversity of the domain, there are several examples in the wildlife management sciences that utilize this exact technique for ensuring long-term viability of endangered ecosystems [SSO11].

6 Ablation Study: No Diffusion vs Diffusion

We analyze what happens when the model is run completely without the Fisher KPP diffusion. In particular, it reveals a missing desirable behavior that only occurs when diffusion is included. See Figure 11.

We notice that the parking lot pollutant indeed makes its impact as expected, but upon further inspection we actually notice some of the population seeping into the parking lot void when diffusion is used. The minimum value of population at $t = 5$ with no diffusion is essentially 0, but the value when diffusion is encountered is actually > 1.

We claim that the result with diffusion makes drastically more physical sense. There will never be locations with purely 0 biomass unless the species has essentially gone extinct, and especially this will not occur in real environments in a region where there is a healthy community of plants. This parking lot pollutant is placed inside where most of the population resides (on top of a large water source), so it makes sense that there is a nonneglible diffusion of the outer population into the inner void.

7 Strengths and Weaknesses

After that barrage of case studies, we now turn to analyzing what we believe are the strenghts and shortcomings of our model, and how the shortcomings could potentially be addressed. For strengths, we hope that it is clear why we believe our framework is strong.

1. **The model is incredibly general.** There is so much customization that can occur because our model accounts, captures, and simulates so many different dynamics.
2. **Each component itself is simple and allows for extensions.** On the other side of the same coin, although our model is general, each component is not complex or esoteric at all - we were able to implement everything just using Python packages `NumPy` and `py-pde`.
3. **Solution Fidelity and Scalability** It is difficult to match the fidelity and precision of a solution via numerically solving differential equations in this fashion. Due to well-understood convergence and error bound results for the methods we used, decreasing the time-step and spatial-step in a fashion according to the CFL condition allows for configurable precision.

As for shortcomings, there are a few that unfortunately prevented deeper exploration using the general framework.

1. **Computationally expensive, especially as the number of species grows.** Assuming the grid is of size $n \times n$, this is because we are essentially introducing a collection of n^2 new PDEs for the

computer to solve.

There are a variety of speedups we could but did not have time to implement, e.g. employing parallelization and vectorization strategies for the iterative procedures. Our finite difference method for solving the heat equation is currently computed using two `for` loops in Python, but the operation over the domain is embarassingly parallel. This would offer a considerable speedup and give more freedom to run demanding simulations, but we were not able to successfully implement it.

2. **Curse of Dimensionality** Our model is deeply intricate and complex, meaning that in order to get meaningful results such as biodiverse systems at an equilibrium, the parameters must be carefully chosen so that no species completely dominates another. The quantities being dimensionless does not help much either with this point, as it is hard to inject real world intuition. We ended up having to work with a small set of parameters that we knew worked rather than having the time to explore and experiment with many new environments. The solution here is probably to reintroduce dimensionality and go through the tedious trouble of computing dimensionality transforms.

References

[Sma76] S. Smale. "On the differential equations of species in competition". In: *Journal of Mathematical Biology* 3.1 (1976), pp. 5–7. DOI: 10.1007/bf00307854.

[TW91] S. Tang and R. O. Weber. "Numerical Study of Fisher's equation by a Petrov-Galerkin finite element method". In: *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics* 33.1 (1991), pp. 27–38. DOI: 10.1017/s0334270000008602.

[AG04] G. J. Ackland and I. D. Gallagher. "Stabilization of Large Generalized Lotka-Volterra Foodwebs By Evolutionary Feedback". In: *Phys. Rev. Lett.* 93 (15 Oct. 2004), p. 158701. DOI: 10.1103/PhysRevLett.93.158701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.158701.

[SWA05] J.C. Sprott, J.C. Wildenberg, and Yousef Azizi. "A simple spatiotemporal chaotic Lotka–Volterra model". In: *Chaos, Solitons amp; Fractals* 26.4 (2005), pp. 1035–1043. DOI: 10.1016/j.chaos.2005.02.015.

[Zay07] Packer Laurence Zayed Amro Constantin Serban. *Successful Biological Invasion despite a Severe Genetic Load — PLoS One. 2007; 2(9): e868.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1964518/. [Accessed 21-Feb-2023]. 2007.

[SSO11] Martin A. Schlaepfer, Dov F. Sax, and Julian D. Olden. "The Potential Conservation Value of Non-Native Species". In: *Conservation Biology* 25 (2011).

[Ehr15] Rachel Ehrenberg. "Global Forest Survey finds trillions of trees". In: *Nature* (2015). DOI: 10.1038/nature.2015.18287.

[CCJ16] Anne Chao, Chun-Huo Chiu, and Lou Jost. *Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers*. Springer, 2016. URL: https://doi.org/10.1007/978-3-319-22461-9_8.

[DNO18] Nelson Christopher Dzupire, Philip Ngare, and Leo Odongo. "A poisson-gamma model for zero inflated rainfall data". In: *Journal of Probability and Statistics* 2018 (2018), pp. 1–12. DOI: 10.1155/2018/1012647.

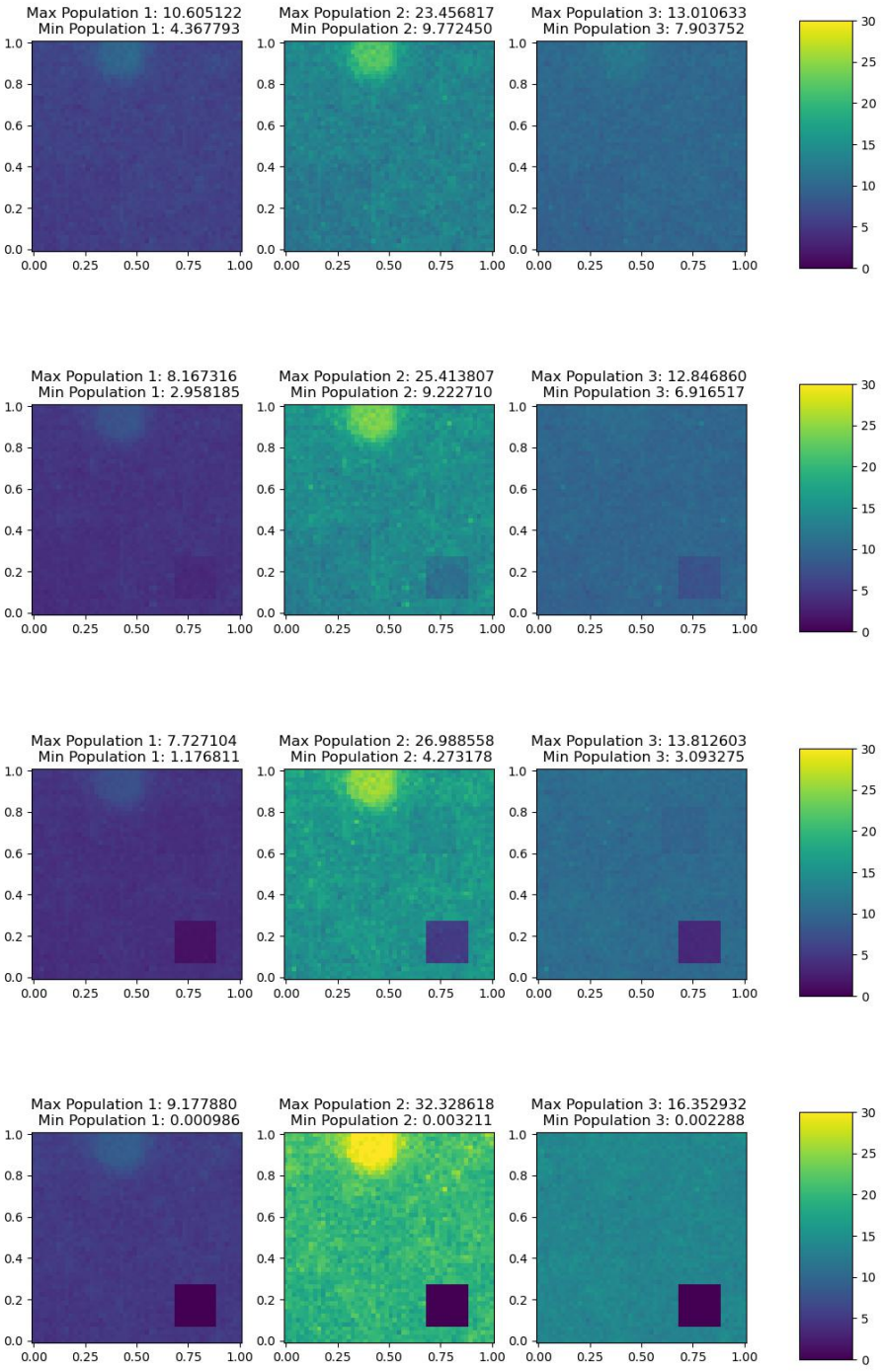


Figure 9: Realization of simulation for 3 species from top to bottom at $t = 1, 2, 3, 5$. Species 3 is a resilient species. We can observe that it outperforms the first species completely, and is also quicker to recover from a drought than the other two due to its low nutrient demand; at $t = 2$ it has already essentially recovered from the drought that occurred on interval $[0, 1.5]$ centered approximately at $(0.2, 0.2)$, while a dark spot is still present for the other species.

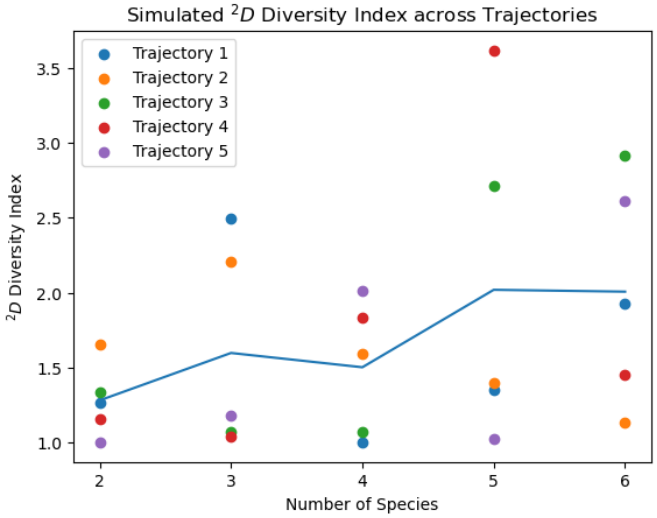
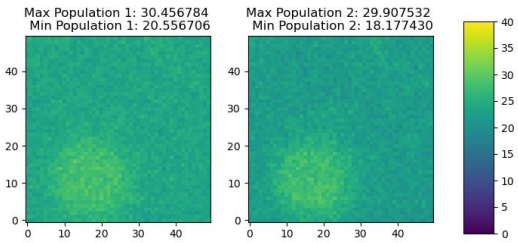
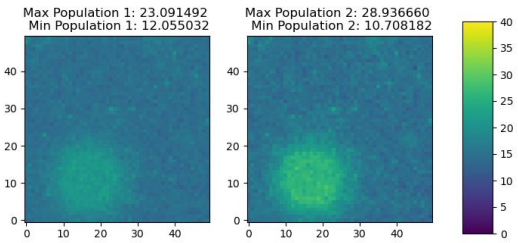


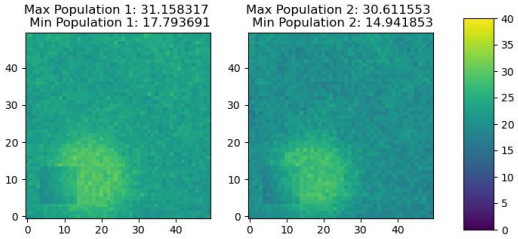
Figure 10: ^2D Values for 5 realizations of the simulation with additional and stronger droughts for each $n \in \{2, 3, 4, 5, 6\}$.



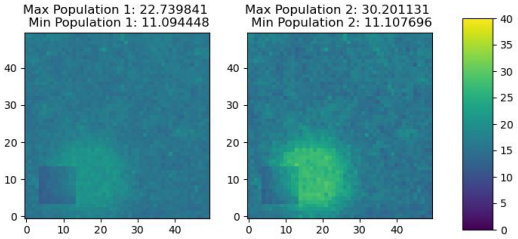
$t = 1$, No Diffusion



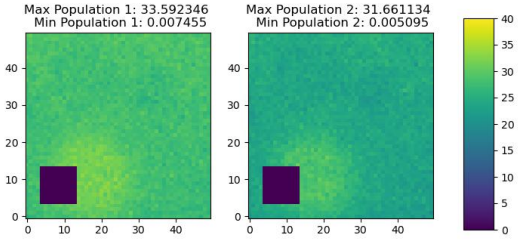
$t = 1$, Diffusion



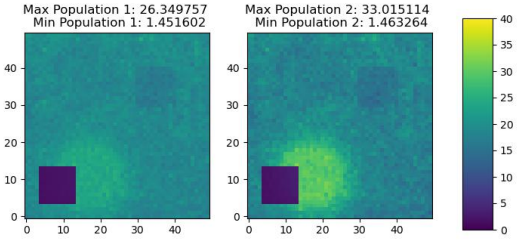
$t = 2.5$, No Diffusion



$t = 2.5$, Diffusion



$t = 5$, No Diffusion



$t = 5$, Diffusion

Figure 11: Ablation study on the Fisher KPP diffusion module. We run with a set seed (notably the same seed as Figure 5) and compare values when no diffusion and diffusion are used.