Problem Choice	2023	Team Number
A	HiMCM	**14632**
	Summary Sheet	

SUMMARY

WE ARE HERE TO "SPREAD" the news about dandelions and their effects on our environment. Problem A consisted of two sub-tasks: (1) modeling the spread of a single dandelion across a one-hectare plot of land over the course of one year and (2) developing a method to measure the impact of a foreign species in a new environment.

To model the spread of dandelions over a plot of land, we developed a *Seed Agent Model*, which follows a dandelion's life cycle from seed to plant to puffball, including germination, growth, and death. We decided to use the Brownian Motion Process to model random dandelion movement and treat seed growth and death as a stochastic agent-based problem, meaning each seed was treated independently. To make our model adaptable to different regions and climates, we introduced several *hyperparameters* of temperature, light, wind, consumers, and soil nutrients, which are calibrated by location.

After we developed our model, we collected easily accessible data for our hyperparameters from three different climate regions: Clay, NY (temperate); Phoenix, AZ (arid); and Florida Keys, FL (tropical). Our results show that dandelion population growth tends to follow a similar cyclical trend year-over-year, where plant populations spike in their seed dispersal and blooming seasons and remain dormant otherwise, regardless of the region's climate; however, different regional climates resulted in different blooming seasons. Across all three regions, our model yielded a final population of around 500 to 900 dandelion plants after one year. Lastly, to test the validity of our model, we employed a Monte Carlo sampling simulation ($\pm 10\%$) and a sensitivity analysis of all hyperparameters, which revealed that light levels and seed lifespans significantly affect dandelion growth, whereas nutrients found in the soil were not as impactful.

For part two, we developed a "trade-off" model that considers both the benefits and disadvantages of introducing a foreign species. This was done by splitting the perceived impact into two components: harmful ecological and beneficial economic impact. Concerning the foreign species' impact on their new environment, we assessed their effect on native species and the speed at which they spread. To do this, we used a Lotka-Volterra system of differential equations and a logistic regression of population data, respectively. As for the economic potential of the foreign species, we conveyed the societal impact as a standardized monetary value. To do so, we used the bioeconomic Gordon-Schaefer model to represent the profit made from the plant and Sen's Welfare functions to calculate the plant's social benefits and human health impact.

From there, we applied the second model to three foreign species — dandelions, garlic mustard, and English ivy — and compared each plant's results. If the economic benefit was significantly higher than the ecological impact, we denoted the species as "non-invasive;" otherwise, it was labeled "invasive." By comparing each plant's results, we decided dandelions were invasive species due to their high ecological impact and below-par economic benefit. Similarly, garlic mustard plants were also considered as invasive, whereas English ivy was considered non-invasive. A benefit of this model is its generalizability to almost any type of plant due to our proposed *Data-Score-Divide* (DSD) framework. Using this framework, species can be automatically classified as invasive or non-invasive by a user-determined threshold to represent the balance between ecological and economic impact scores.

Overall, our agent-based and stochastic models thoroughly represent dandelions in the environment, how they spread, and their role as an invasive species. Easily adaptable, our models can be used on virtually any plot of land and for any type of invasive species. Although they seem irrelevant growing in your yard, the effects of dandelions on the environment can "blow" you away. #dandelionPRIDE

CONTENTS

1 Introduction

1.1 Background

Everyone knows the best way to make a wish: picking up a dandelion in its iconic "puffball" stage and blowing on it to release thousands of seeds into the environment. But how much do these seeds contribute to dandelion growth and spread? In this paper, we develop a mathematical model to predict the spread of dandelions across a plot of land and evaluate whether plants like dandelions are truly invasive.

Beginning as one tiny airborne seed, there are several phases that a dandelion progresses through in its lifetime. The first phase is when the seeds of a "puffball" spread, usually by wind. Once landed in soil, the seed will begin its plant phase, which has two main processes: (1) germination, where the seed breaks dormancy, and (2) growth as a plant, where the seed continues to adapt to the environment, eventually becoming the iconic bright yellow flower. Finally, in the third phase, the dandelion flower returns to a puffball stage to disperse even more seeds, thereby restarting the cycle. Although the cycle lasts about 50 days on average, it often resets randomly due to unpredictable variables, such as predator consumption, inclement weather, or a possibly unsuccessful growth stage [38].
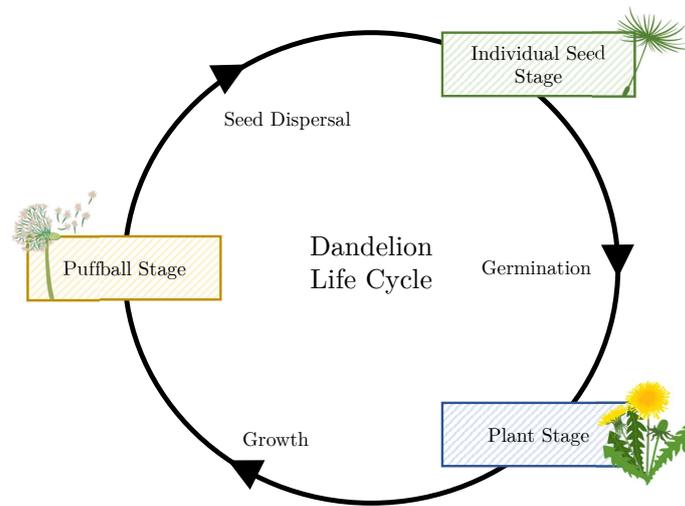
Figure 1: **Dandelion life-cycle.** A recurring process from seed to germination to growth, then the dispersal of seeds again.

1.1.1 Phase 1: Seed

Dandelion seeds begin their dispersal process from a puffball. However, only around 2-4% of the many seeds make it past the initial dispersal stage, the remainder of which is lost to consumers [7]. Upon landing, it will lay dormant until it meets certain environmental conditions to germinate and begin its plant growth process.

1.1.2 Phase 2: Plant

Next, the seed continues into the plant growth process. Here, certain environmental conditions, such as the right temperature, light, and nutrients in the soil can aid the growth of a dandelion. The dandelion is no longer considered growing when it has reached its puffball phase.

1.1.3 Phase 3: Puffball

The puffball stage occurs when the white seeds of the plant fully emerge, replacing the distinctive yellow flowers [4]. Once ready, the dandelion releases its seeds to the power of the wind which scatters the seeds into an everlasting journey.

1.1.4 Death

As a perennial plant, the taproot of the dandelion allows the plant to keep its roots for another year while the flowers and stem wither away in the winter. In the event of any damage to the portion of the plant above ground, the taproot remains rooted, and so, the dandelion continues to survive [35]. For a dandelion to die fully, either the taproot is entirely removed from the ground, or the plant reaches its life expectancy of 10-13 years [26]. In Figure 1, we summarize the life cycle of a dandelion (from seed to puffball), from which we draw inspiration for our dandelion spread model.

1.2   Problem Restatement

1. Develop a model to represent the spread of dandelions over the course of 1, 2, 3, 6 and 12 months, given the model begins with a single dandelion in its puffball stage adjacent to an open, one-hectare plot of land.

2. Formulate a mathematical model to determine an "impact factor" for invasive species.

   a. Compute an impact factor for dandelions to test the model.

   b. Determine the impact factor for two other non-native plant species of our choice that are often considered invasive.

2   PART 1: MODELING THE SPREAD OF DANDELIONS

2.1   Problem Analysis

Because a dandelion is a living organism, its life follows a cycle through the several phases outlined in the previous section and in Figure 1. In addition to the cyclic nature of dandelion growth, some unpredictability exists due to changing environmental factors such as the wind and consumers.

2.2   Assumptions

#	Assumption	Justification
1	Dandelion seeds that land outside of the one-hectare plot of land will not be considered.	The problem statement refers to a dandelion adjacent to a plot of land, so we interpreted that as modeling the growth of dandelions on that plot of land only.
2	Dandelion seeds behave mostly independently.	This allows us to develop an agent-based model to analyze each individual seed and easily adjust hyperparameters.
3	Although consumers are considered on the plot of land, human disturbances do not occur.	Because information is not provided on the whereabouts of the plot of land, we assume that no humans will cause a significant disturbance to the growth of the dandelions.
4	No natural disasters will occur during the 12-month period we are modeling.	Because the probability of a natural disaster occurring is low [5], our model does not consider it.
5	Nitrogen and potassium are the only nutrients affecting the growth of dandelions.	Higher potassium and nitrogen levels in the soil are the primary factors for increased seed germination and growth speed. [7].
6	Wind is the only factor that spreads the seeds from a puffball.	As stated previously, no human disturbances could cause seeds to spread; therefore, wind is the only other natural factor that can affect seed dispersal.
7	Our model begins on January 1st, and ends on December 31st.	To model an entire year and include climate-related factors in our model, we are assuming that the first dandelion of the model exists on January 1st and is modeled until December 31st.
8	The one-hectare plot of land is square (100 meters by 100 meters).	By specifying the dimensions of the plot of land to be a square, we can better model the movement of dandelion seeds.

2.3   Brief Overview

First, we model the spread of the seeds from an initial puffball across a one-hectare plot of land. Afterward, we minimize the scope of the problem, focusing on each individual seed to model the independent growth of a seed. Once each seed matures into a puffball, the process repeats.

To model the life cycles of a seed, we developed a *Seed Agent Model*, which considers environmental characteristics, such as wind, temperature, sunlight, and nutrients, to iterate each seed through a germination and plant growth life cycle. Finally, when new puffballs are fully developed, the Seed Agent Model is again applied to model the spread and growth of every new seed. Throughout our process, we incorporate environmental unpredictability by developing several stochastic processes.

Below is a table of variables used throughout our agent-based model.

Variable	Symbol	Description
Time Step	t	Each time step is a day where $t = 0$ is January 1 of a year.
Number of Moves	N	Number of moves a seed takes in its movement process.
Coordinate Shift	$\Delta x, \Delta y$	Change in x or y coordinate of a seed in every move.
Death Date	D	Time of death of a seed, sampled from an Exponential distribution.
Plant Growth	G	The stage of growth of a dandelion plant between 0 and 1.
Temperature Index	T_I	A standardized metric to describe optimal temperatures for dandelions.
Light Index	L_I	A standardized metric to describe optimal light levels for dandelions.
Nutrition Index	N_I	A standardized metric to describe optimal soil conditions for dandelions.

2.4 The Way Of The Wind: A Seed Dispersal Process

We first identified the primary contributing factor of dandelion seed movement to be the wind [45], following Assumption 6. Thus, we decided to use the Brownian Motion Process to model the spread of dandelion seeds. Brownian Motion stochastic processes are often utilized to model the stochastic, or random, movement of particles in a particular medium, such as seeds in open air.

To begin, we define the number of seeds of the initial puffball dandelion to be S_0. The puffball releases its S_0 seeds into a *stochastic Brownian motion process* in a 2D plane.

To mimic the random nature of wind in determining a seed's new position, each seed will complete two stages of an algorithm. First, a seed is assigned a random number of position translations, or "moves." Next, the distance the seed moves in that position is randomly determined for each position translation. This allows us to model the varying movements of the dandelion seeds in all directions.

2.4.1 Number of "Moves" a Seed Performs

First, we determine N, the number of "moves" a seed will perform on the 2D plane in time-step t. Next, we use a probabilistic model to account for varying winds in different climates and days. Therefore, we assign the random variable N with a non-homogeneous (changing over time) Poisson distribution and wind strength hyperparameter $m(t)$, denoting the mean number of steps at time step t. Poisson probability distributions are used when finding the number of events that occur in a given time frame. We can model this as

$$N \sim \text{Poisson}(m(t)) \tag{1}$$

$$\mathbb{P}(N = n) = e^{-m(t)} \frac{[m(t)]^2}{n!}$$

In other words, the number of moves a seed will take will be sampled from this Poisson distribution, which changes according to wind levels (Note: wind levels change based on the region and time). We use k to denote the k'th move out of n moves such that $k \in [0, n]$.

2.4.2 Distance Per "Move"

Next, we determine the change in position for each move k. The seed will move a specified distance Δx and Δy which are sampled from a Normal$(0, \sigma^2)$ such that the new coordinate for each seed after step k will be

$$(x_i, y_i)_{k+1} = (x_i, y_i)_k + (\Delta x, \Delta y) \tag{2}$$

$$\Delta x \text{ or } \Delta y \sim \text{Normal}(0, \sigma^2) \tag{3}$$
$$\mathbb{P}(\Delta x \text{ or } \Delta y = d) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{d}{\sigma}\right)^2}$$

Recall that from Assumption 1, dandelion seeds that travel outside our provided hectare of land are removed from the consideration of our model to simplify our model's scope.

2.5 Race the Clock: Beating a Death Date

After the seed is assigned a new location, the seed will go through two critical processes: (a) germination and (b) plant growth. However, the seed may die of natural causes before each phase. To account for this nuance, we introduce a *Death Date* for each seed that resets at the end of the germination and plant growth stage.

The first death date of a seed will be assigned after it lands on the ground and before it begins the germination process. The death date is determined by randomly sampling from a probabilistic survivorship curve (see Figure 2). If the seed successfully germinates before its assigned death date, it will be assigned a new death date for the plant growth cycle.
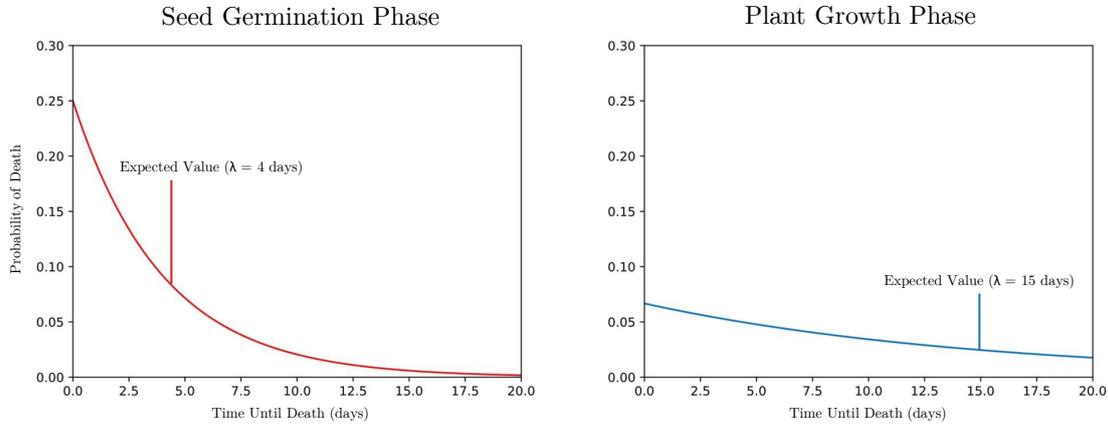
Figure 2: **The survival rate of a seed during two major phases.**

The survival rate of dandelions is often classified as a Type III survivorship curve [44], where shorter lifespans are more probable than longer lifespans. Thus, due to the low percentage of seeds that eventually germinate, it is more likely to randomly select a relatively earlier death date than a later death date [7].

Henceforth, we used a non-homogeneous (time-varying) Exponential distribution to model survival rates. However, per Assumption 2, the only time dandelions are not independent is when two seeds land in the same place, thereby eliminating one of the two.

The probability of the random variable of seed death date D at time t, $\mathbb{P}(D = t)$, can be determined by finding the *survivorship curve* of dandelions in their respective environments.

$$D \sim \text{Exponential}(\lambda(t))$$
$$\mathbb{P}(D = t) = \lambda(t) e^{-\lambda(t)t} \tag{4}$$

where $\frac{1}{\lambda(t)}$ is the average lifespan of a dandelion seed at time step t. For regional calibration, $\lambda(t)$ can be adjusted. For the sake of differentiating the death date of seeds and dandelion plants, we will introduce two hyperparameters $\lambda_s(t)$ and $\lambda_d(t)$ for the two stages, respectively [4]. Note that because no natural disasters nor human disturbances will occur on our hectare plot of land (Assumptions 3 and 4), neither of these lifespan hyperparameters will be arbitrarily low.

2.6 Stages of Dandelion Growth

After a dandelion has landed in the ground and a death date has been assigned, we begin to model the dandelion growth process. There are two phases: (1) seed germination and (2) the plant growth process.

Because the seed germination stage is relatively shorter than the much longer plant growth process, we use historical data to determine a germination length and a more extended process to model the plant development cycle.

2.6.1　Phase 1: Seed Germination

To obtain the time for a seed to germinate, we sample from a Normal $(\mu_{\text{germination}}(t), \sigma^2)$ distribution. Note that the mean duration of seed germination varies with regard to the climate and region [4]. Because seeds usually have different expected lifespans than dandelion plants, we assign a new death date from the seed survival distribution with a new hyperparameter $\lambda_d(t)$ after each seed germinates.

2.6.2　Phase 2: Plant Development

Once germinated, we define the growth function G based on three indices T_I, L_I, and N_I for temperature $\tau(t)$, light levels $l(t)$, and soil nutrient composition, respectively. When growth G is at 0, we say that a single dandelion just began growing (or germinating). When the growth G is at 1, we say that the dandelion is at the "puffball" stage and releases its seeds.

Temperature Index

For the *temperature index* $T_I : \tau(t) \longrightarrow [-1, 1]$, we define a piecewise function that assigns a negative score if the temperature (°F) $\tau(t)$ falls out of the ideal dandelion growth temperatures and a positive score if the temperatures are ideal. Positive scores are based on squared Euclidean distance from optimal temperatures $[50\,°\text{F}, 77\,°\text{F}]$ [4]. Non-ideal temperatures are normalized on a logistic curve, allowing us to scale scores between -1 and 1, while adding a non-linearity factor to weigh the optimal values more heavily. So, we can define T_I as

$$T_I = \begin{cases} \frac{-2}{1+e^{-0.1(\tau(t)-77)}} + 1, & \tau(t) \geq 77 \\ 1 - \frac{1}{63.5^2}(\tau(t) - 63.5), & 50 < \tau(t) < 77 \\ \frac{-2}{1+e^{-0.1(50-\tau(t))}} + 1, & \tau(t) \leq 50 \end{cases} \tag{5}$$

Light Index

For the *light index* $L_I : l(t) \longrightarrow [-1, 1]$, we again define a piecewise function which similarly assigns scores but for the optimal interval of light hours $[6\text{ hours}, 24\text{ hours}]$[38]. So, we define L_I as,

$$L_I = \begin{cases} \frac{-2}{1+e^{-0.5(l(t)-24)}} + 1, & l(t) \geq 24 \\ 1 - \frac{1}{15^2}(l(t) - 15), & 6 < l(t) < 24 \\ \frac{-2}{1+e^{-0.5(6-l(t))}} + 1, & l(t) \leq 24 \end{cases} \tag{6}$$

Nutrition Index

Lastly, we define the soil *nutrition index* $N_I : (k(t), n(t)) \longrightarrow [-1, 1]$ to be a function on potassium levels and nitrogen levels (see Assumption 5) in a given region where potassium and nitrogen levels are optimally between $[40, 80]$ and $[20, 40]$ ppm, respectively [2]. Based on this information, we formulate the following exponential transformation for the soil nutrient index. By defining a similar shape to the previous index functions and min-max scaling $k(t)$ and $n(t)$, we arrive at the following index function.

$$N_I = \exp\left(-\left(\frac{k(t) - 60}{40}\right)^2\right) + \exp\left(-\left(\frac{n(t) - 30}{30}\right)^2\right) - 1 \tag{7}$$

Using the three indices from above (Equations 5, 6, 7), we formulate the change in growth, $\frac{dG}{dt}(t)$ at time step t by scaling the sum of T_I, L_I, and N_I to match the time to fully grow a dandelion under optimal conditions (minimum 50 days) [19]. Furthermore, we scale the sum of the indices by $\frac{1}{3}$ to return to the -1 to

1 scale and another $\frac{1}{50}$ as optimal growing conditions (minimum 50 days) would result in $\frac{dG}{dt} = \frac{1}{50}$. Therefore, we define $\frac{dG}{dt}(t)$ as,

$$\frac{dG}{dt}(t) = \frac{1}{3 \cdot 50}\left(T_I + L_I + N_I\right) \tag{8}$$

$$\text{where } \max\left(\frac{dG}{dt}\right) = \frac{1}{50}$$
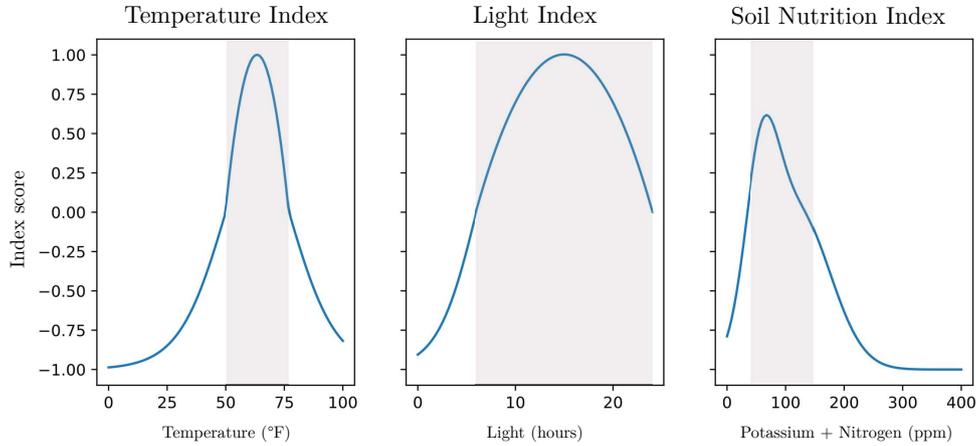
Figure 3: **From the left to right: T_I, L_I, N_I scores for plant growth.** Each shaded area represents the optimal interval for each unit of measure, which receives a positive score.

2.7 Seed Agent Model Overview

Our method of modeling the spread of dandelions from an agent-based perspective is summarized in Figure 4.
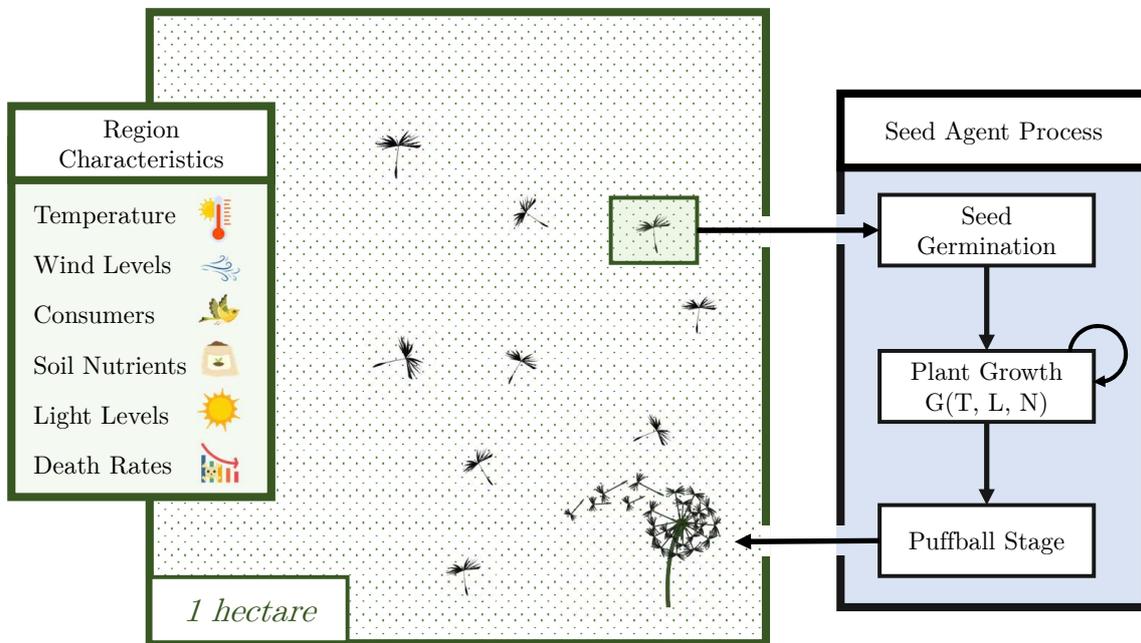
Figure 4: **Overview diagram of seed spread modeling.**

3   APPLICATION OF AGENT-BASED MODEL

To simulate our agent-based model in real-world scenarios, we chose three distinct plots of land from a temperate, arid, and tropical region with relatively low human interference. So, we randomly selected three "open" regions in the United States. The regions selected for the temperate, arid, and tropical regions are Clay, New York; Phoenix, Arizona; and Florida Keys, Florida, respectively.

3.1   Climate Calibration

To calibrate our model to each selected region, we define several of our *hyperparameters* from the previous section, which capture the difference in climate, consumers, and soil nutrition levels. We use the term hyperparameters to designate global values that are not directly passed into our model but still change the output of the model.

- Temperature — The local temperature (°F) over 2022.

- Light Levels — The local average hours of sunshine over 2022.

- Potassium and Nitrogen Levels — Local potassium and nitrogen levels (ppm) in the local soil. (Note that since Potassium levels are closely correlated with soil moisture, we did not choose to calibrate our model with rainfall, as potassium was indicative enough).

See Appendix 1 for more details on the implementation of our agent-based simulation.

3.2   Regional Results

3.2.1   Temperate Region: Clay, NY

We obtained data detailing the local temperature, light level, soil composition, and wind data of Clay, NY to calibrate our model's *regional hyperparameters* [1, 32]. Our results for this region are summarized in Figure 5 where t represents time and n represents the number of dandelions. Additionally, the graphs are scaled to the dimensions of the open plot of land we began with (100 meters by 100 meters). Recall that our first dandelion started at coordinate $(0, 0)$.
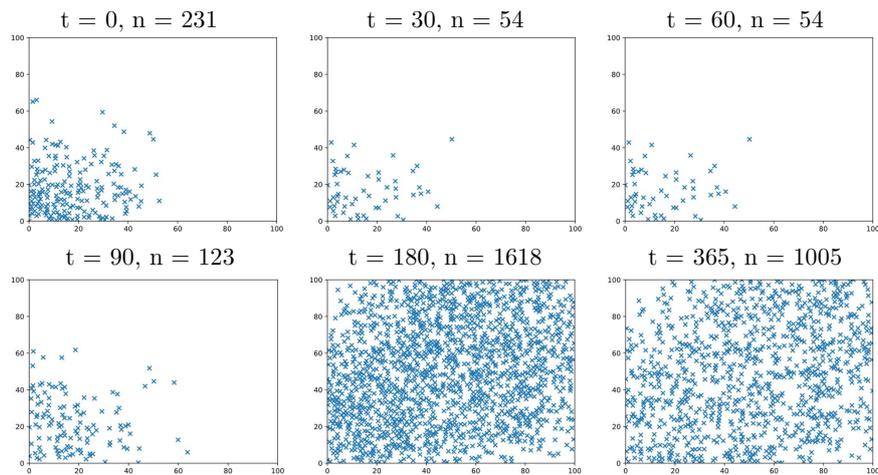
Figure 5: **Dandelion spread over one year in Clay, NY.** Each mark labels a spot where a dandelion (either in seed or plant phase) is located at that time stamp t.

One exciting observation is that the stochastic seed dispersal process seems to cover the plot of land over time uniformly. However, it is essential to note that *not all seeds survive* the dispersal process. After 30, 60, and 90 days of simulation from the beginning of the calendar year, the seeds began to travel even farther from the starting point. At the 180-day mark, we see an influx of seeds and dandelion plants due to the effects of optimal climate conditions for a dandelion's blooming phase. After the blooming phase, we see that through day 365, all surviving agents that are left are plants in their dormancy stage.

A visualization of the growth (Figure 6) of the dandelion population and seed population over time shows us two key results:

1. Dandelion plants go through two main blooming processes (spikes in population increase) throughout a year (spring and summer), which is consistent with other findings [7].

2. Dandelion plant populations remain dormant in winter and fall months, while dandelion seeds do not spread much, or at all, in the colder months, which is a unique property of perennial plants [7].
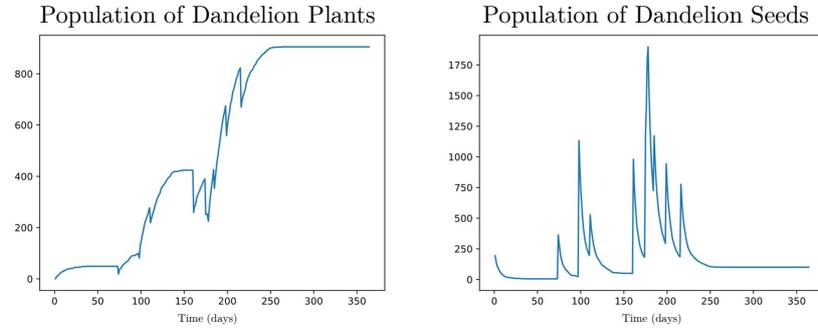
Figure 6: **Growth in population of dandelion plants and seeds over time in Clay, NY.** The simulation began with a single puffball at coordinate $(0,0)$.
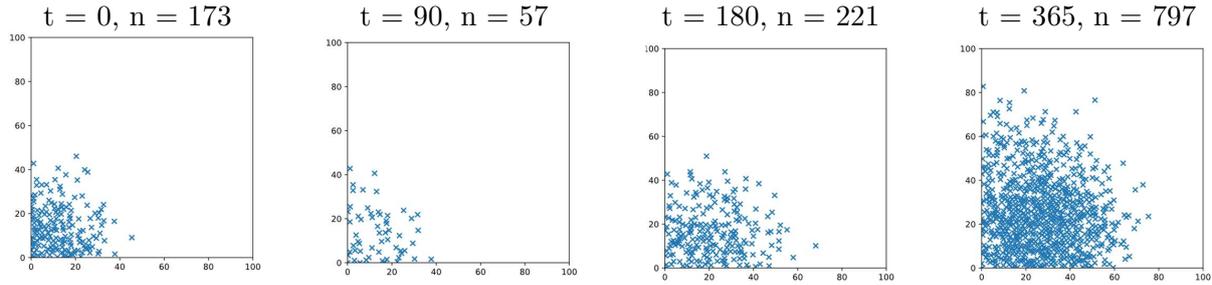
Figure 7: **Dandelion spread over one year in Phoenix, Arizona.**
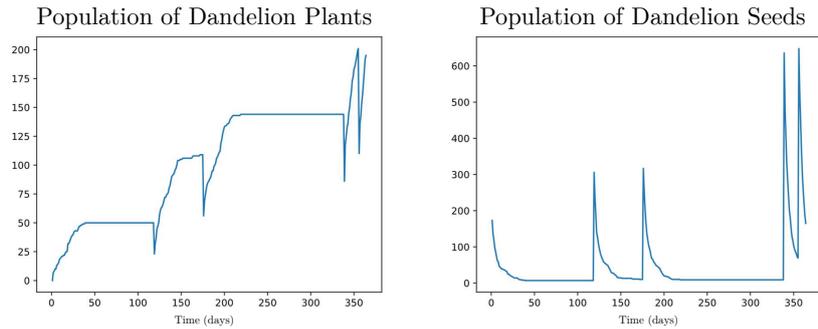
Figure 8: **Dandelion plant population growth in Phoenix, Arizona.**

3.2.2   Arid Region: Phoenix, AZ

In Phoenix, Arizona, we obtained new climate and soil nutrition data from several sources [2, 27, 11]. We then re-calibrated our model with these new hyperparameters and ran our simulated model.

One captivating observation of the spread of dandelions in an arid region is that their blooming seasons occur at different times of the year. This is most likely due to the availability of soil nutrients and optimal temperatures in the ending months of each year.

3.2.3   Tropical Region: Florida Keys, FL

By calibrating our model to Florida's tropical climate in the Florida Keys, we arrived at the following results for the dandelion growth and spread over one year. The calibrated climate data points were pulled from numerous sources as listed [22, 10, 28].

One riveting observation of tropical regions is that blooming seasons tend to occur during winters at $60 - 70°F$. Furthermore, we see that dandelion populations remain mostly stagnant over the summer months, as the summer heat may take away from dandelion growth. We can also conclude that dandelion populations follow a logistic growth pattern for the most part, which will be revisited in the following impact score model.

4   MODEL DISCUSSION

As a part of the creation of our agent-based model, where numerous climate and regional factors contribute to the spread of dandelion plants, we found it essential to conduct a sensitivity analysis to determine the most
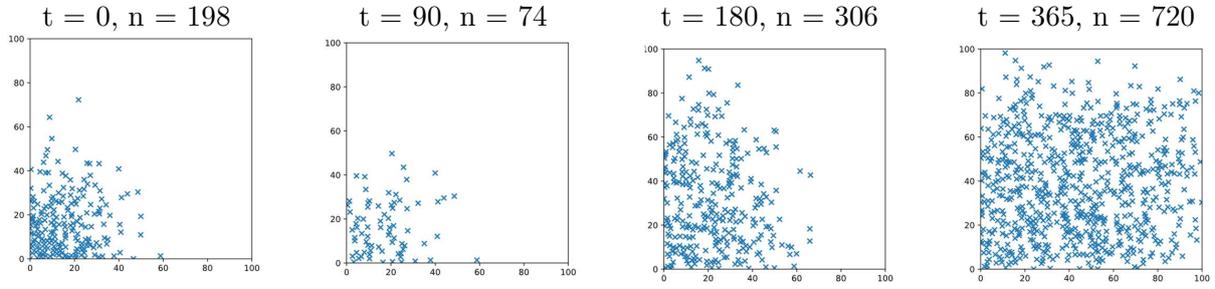
Figure 9: **Dandelion spread over one year in Florida Keys, Florida.**
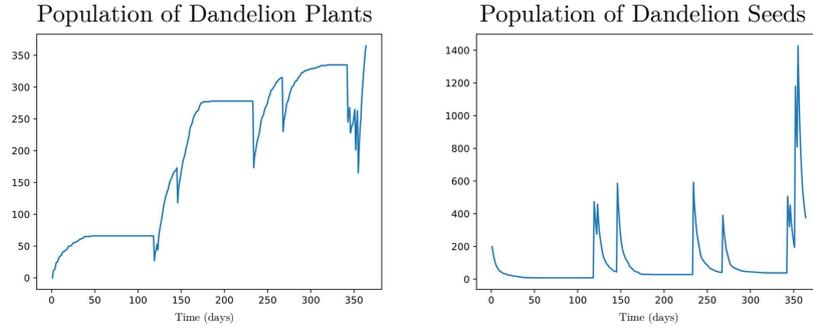
Figure 10: **Dandelion plant population growth in Florida Keys, Florida.**

impactful factors on dandelion growth. Beyond our sensitivity analysis, we also provide a brief discussion of our model's key strengths and weaknesses, as well as a scope of how our model can be readily generalizable to virtually any plot of land.

4.1   Model Hyperparameters

The following table consolidates all hyperparameters based on our model described above. Our sensitivity analysis therefore investigates the impacts of these hyperparameters on the spread of dandelion plants and seeds.

Hyperparameter	Symbol	Description	Default Value
Temperature	$\tau(t)$	Temperature (°F) over a year in a specific region. Dandelions grow optimally between $50-77°$ F [7].	Temperatures of Clay, NY (2022) [1].
Light Levels	$l(t)$	Hours of sunlight per day over a year. Dandelions grow best between 6-24 light hours per day [38].	Daily hours of light of Clay, NY (2022) data [1].
Wind Levels	$m(t)$	Average mph of wind per day over a year which affects movement processes.	Wind levels in Clay, NY (2022) [1].
K Levels	$k(t)$	Average amount of potassium in a regionally collected sample of soil in ppm over time. K levels account for soil moisture and richness [43].	K levels from a soil survey [12].
N Levels	$n(t)$	Average amount of Nitrogen (ppm) over time in a regionally collected sample of soil.	N levels from a soil survey [28].
Seed Lifespan	$\lambda_s(t)$	Average time until death for dandelion seeds.	4 days [40].
Dandelion Lifespan	$\lambda_d(t)$	Average time until death for dandelion plants.	15 days [35].
Initial Seeds	S_0	Number of seeds dispersed by each puffball.	2000 seeds [7].

4.2   Sensitivity Analysis

A standard methodology for sensitivity analysis is to randomly sample various values and analyze the model's output. Therefore, to randomly examine different sets of hyperparameters for our model, we employed

a Monte Carlo sampling simulation to randomly sample $\pm 10\%$ of default values. Based on the simulation, we modeled the effects of these changes on the growth of dandelion plant populations over time. These results are shown in Figure 11.
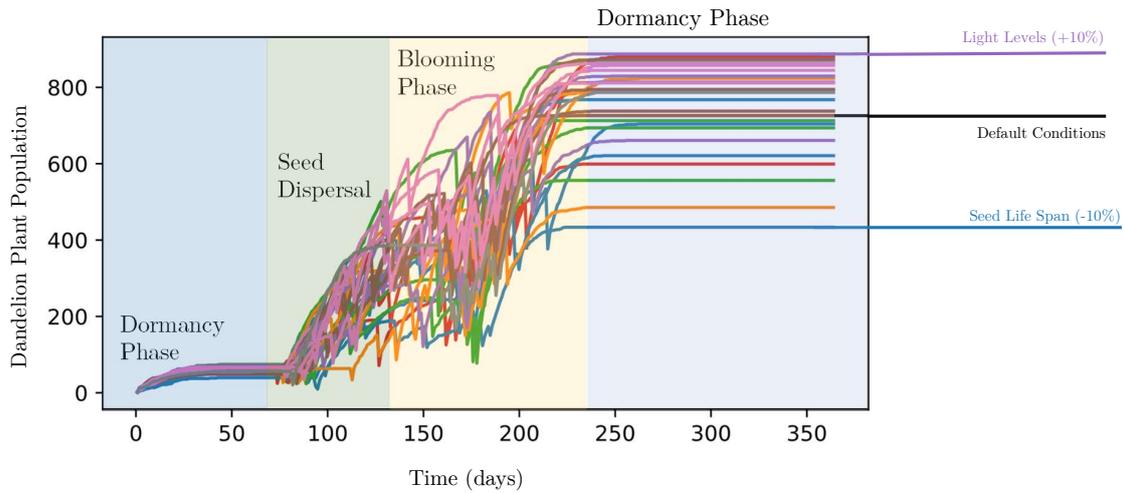
Figure 11: **Monte Carlo simulation sampling for hyperparameter sensitivity.**

In Figure 11, we demonstrate how, despite changes in our model's hyperparameters, dandelion populations tend to follow the same three-part process of dormancy, seed dispersal, and puffball blooming. At the end of the blooming phase, seeds are dispersed until the next blooming season reoccurs. Furthermore, dandelion plant populations generally approach an asymptote known as the carrying capacity of the plot of land, which in our case, is dependent on a hyperparameter specifying the minimum distance between two dandelions necessary to survive.

To numerically analyze our results from Figure 11, we analyzed the average Root-mean squared error (RMSE) between each factor's Monte Carlo simulation and our simulation with default conditions applied. The RMSE formula is commonly used in time-series analysis to determine the average deviation of a new time series from a ground-truth series [8].

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(x_{\text{new}_i} - x_{\text{ground truth}_i} \right)^2} \tag{9}$$

Using the formula for RMSE where N is the number of entries compared, we scaled each of these errors between 0 and 1 using min-max scaling and displayed our sensitivity analysis results in the radar chart below. Based on our analysis, light levels and seed lifespan have the most significant impacts on dandelion population growth (See Figure 12), which is consistent with findings from field observations [9].
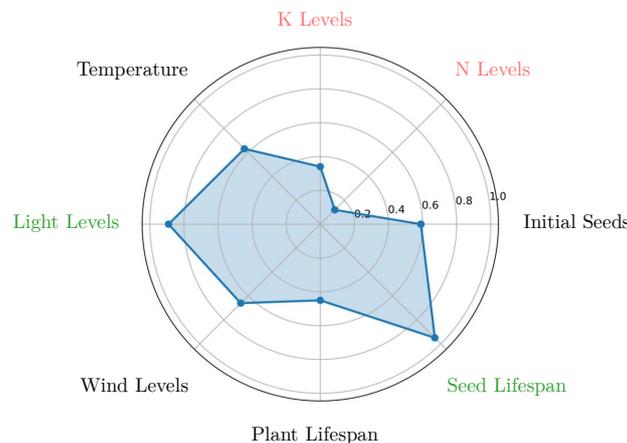
Figure 12: **Radar chart of model hyperparameters.** The two most impactful factors of our model were light levels and seed lifespan, but nutrient levels were least significant.

4.3 Strengths

1. Our model's main strength is its ability to **powerfully generalize**.

   For instance, by including numerous regional and climate hyperparameters into a single, robust framework, our model enables any investigator to thoroughly examine a plot of land for the growth of dandelions. Furthermore, all hyperparameters necessary can be easily obtained from open-access soil samples and climate analysis.

2. Our model employs both **deterministic and probabilistic** techniques.

   Our model reflects real-life behavior of a cyclic dandelion population growth pattern and allows for stochastic variation in dandelion movement and growth.

3. A thorough Monte Carlo simulation and sensitivity analysis performed **helps dandelion growers** and harvesting firms best grow dandelions through our analysis of impact factors.

   Our analysis concluded that light levels and seed lifespans impact dandelion growth the most; therefore, dandelion harvesters must prioritize these factors above all others to ensure that dandelions can grow most optimally.

4.4 Limitations

1. Some special data for dandelions could not be found, and so we assumed **that seeds act as independent agents** and follow growth cycles based on some key climate parameters.

   More abundant data resources can guarantee more robust accommodation for diverse climates in our models. For example, occurrences of natural disasters, significant human interference, and specific consumer-based behaviors would result in a more adept model.

2. Our model **derives climate and death rates from historical data**.

   Therefore, our model lacks predictive power for factors such as global warming and competition with other species. Despite this limitation, using stochastic processes ensures that our model tends toward realism while also maintaining model elegance.

5 PART 2: INVASIVE SPECIES TRADE-OFF MODEL

5.1 Problem Analysis

As with most events, the growth of foreign plants comes with benefits and drawbacks. Thus, to develop an invasive species impact score, we divide the impact into two primary categories: ecological impact and economic potential (see Figure 13). Through this concept, we can represent our "impact factor" as the trade-off between ecological and economic impacts.
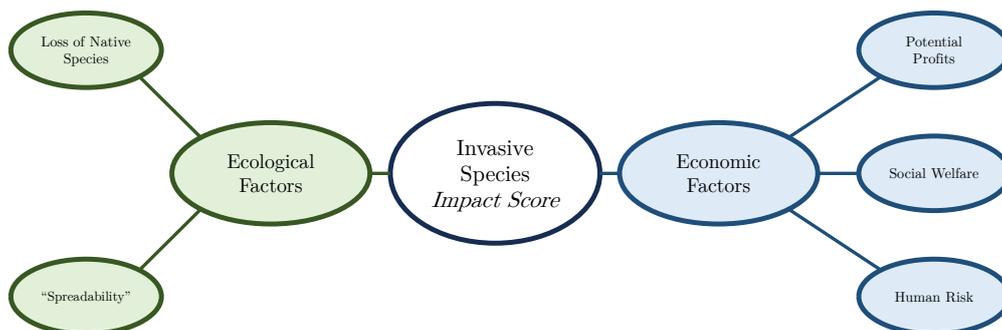
Figure 13: **Overview of trade-off model.**

5.2 Model Overview

Invasive species are non-native organisms that typically out-compete native species because of their ability to reproduce aggressively. Invasive species often harm their surrounding environments by disrupting food chains, which further causes ecological imbalance.

In the following section, we define *foreign species* as a non-native species introduced to a new region of interest. Furthermore, we define *spreadability* as the ease with which a foreign species can proliferate on a rural plot of land. We believe our proposed trade-off model will determine a species as invasive or simply foreign. Our model is useful for all perspectives because we model the "impact factor" as a comparison between the risks and rewards associated with foreign plants in a new environment.

5.3 Assumptions & Variables

In addition to our assumptions from Part 1, we provide more assumptions specific to Part 2 below. The following list of variables summarizes the key components of our trade-off model. We will derive each of the following functions (or scores) in this section.

#	Assumption	Justification
9	Native species are in competitive equilibrium before introducing a foreign species.	This key assumption allows us to use the Lotka-Volterra equations to compare before and after effects.
10	Foreign species populations grow logistically with a set carrying capacity.	This allows us to create a standardized metric based on logistic regression for measuring population growth.
11	A foreign species' social benefit and human health risk can be measured in dollars.	Measuring societal impacts in dollars allows us to create another standardized measure.

Variable	Symbol	Description
Native Species Population Changes	χ	Sum of changes in native species' population after introducing a foreign species (measured as a proportion).
Spreadability Rate	r	Rate of uninterrupted logistic growth of foreign species.
Economic Profit	π	Potential profit (USD) from a business centered around a foreign species.
Social Welfare Benefit	W	Implicit social benefit (measured in USD) from a foreign species.
Human Health Risk	R	Potential risk (measured in USD) of a foreign species to humans.

5.4 Ecological Impact Score

To begin, we determine two main measures of a foreign species' ecological impact: the overall change in the population of the native species and the rate at which a foreign species can spread (spreadability). Therefore, we formulate an ecological impact score as a combination of these factors, none of which are dispositive.

5.4.1 Effects on Native Species

To analyze the impacts of a foreign species in a new region, we need to measure the unintended consequences on native populations. To do so, we used the Lotka-Volterra model (a system of differential equations) to represent the changes in each native population. The Lotka-Volterra model is commonly used to analyze competing species' population dynamics, especially in our case between foreign and native species [30].

Assuming native species are in competitive equilibrium before the introduction of a new, foreign species (Assumption 9), we can use a generalized system of Lotka-Volterra equations for n native species, where we can treat population growth for each species as a single vector based on the populations of other species at each time step. Additionally, we use a matrix to model the weightings of how other native populations affect the growth of another species.

Mathematically, we let there be n native species where the population of a native species i where $1 \leq i \leq n$ is denoted by x_i. Therefore, the rate of change of a population *in relation to* other species can be represented as

$$\frac{dx_i}{dt} = r_i x_i \left(1 - \frac{\sum_{j=1}^{n} \alpha_{ij} x_j}{K_i} \right) \tag{10}$$

where r_i is the uninterrupted growth rate of species i, α_{ij} is the impact of species j on species i, and K_i is the carrying capacity of species i. The parameters α_{ij} can be combined into an n by n matrix of "interaction" parameters. Therefore, we denote this matrix of parameters the *interaction matrix*. For the first n native species, the Lotka-Volterra equations will yield a system of n differential equations with an n by n matrix α. When we introduce our new foreign species, the Lotka-Volterra model will yield a system of $n+1$ differential equations with an interaction matrix α of dimensions $n+1$ by $n+1$ where all values $\alpha_{ii} = 1$ for self-interaction of species. See Equation 11 for an example interaction matrix of three native species and one foreign species.

$$\alpha \text{ foreign species} = \begin{matrix} \text{Native Species 1} \\ \text{Native Species 2} \\ \text{Native Species 3} \\ \text{Foreign Species} \end{matrix} \underbrace{\begin{bmatrix} 1 & \alpha_{1,2} & \alpha_{1,3} & \alpha_{1,4} \\ \alpha_{2,1} & 1 & \cdots & \cdots \\ \alpha_{3,1} & \cdots & 1 & \cdots \\ \alpha_{4,1} & \cdots & \cdots & 1 \end{bmatrix}}_{\text{3 native species, 1 foreign species.}} \tag{11}$$

We used SciPy, a scientific software, to obtain numerical solutions (x_i) to the Lotka-Volterra system [31]. With population functions x_i, we can model the populations over time. For simplicity's sake, we assume that the changes in populations of 3 native species are representative enough of the impact of each foreign species. To create a standardized score, we developed a native population impact score, which measures the change of a native species' population as a ratio after one year since introducing a new foreign species. We call the sum of these ratios our native species impact score, represented by χ.

$$\chi = \sum_{i=1}^{n} \frac{x_i \text{ new}}{x_i \text{ old}} \tag{12}$$

5.4.2   Spreadability Index

To measure spreadability, we focus on the ability of a population to regenerate and grow without competitors and with optimal access to natural resources. Regarding population growth as a logistical growth problem (see Assumption 10), we evaluate the *spreadability* of a species. As most populations follow a logistic growth pattern with an ultimate environmental carrying capacity, we can generalize this process to most, if not all, common species [44]. We can perform logistic regression on foreign species population data, which is modeled by the following function for carrying capacity L, constant C, and proportionality constant of growth r (Equation 13).

$$N(t) = \frac{L}{1 + Ce^{-rt}} \tag{13}$$

Since r can easily describe the proportion of growth, we let our *spreadability index* be the value r scaled adequately by a factor of 100. Thus, our spreadability index can be represented as $100r$. Hereinafter, we can calculate our ecological impact score as the sum of χ and $100r$. Note that a lower ecological impact score is desired.

$$\text{Ecological Impact Score } = 100r + \chi \tag{14}$$

5.5   Economic Potential Score

Physical risk, human well-being, and cold, hard profits are all important considerations to determine the overall effect of a foreign species on humans. But, how can we measure these values, such as happiness? Money. Based on Assumption 11, we use monetary value as a standardized measure for analyzing the impact of foreign species on humans.

5.5.1 Net Profit

In bioeconomics, a standard model used to analyze the production value and potential maximum profit from harvesting a specific species is called the Gordon-Schaefer Model [47]. Gordon-Schaefer bioeconomic analysis methods can be applied whenever a species follows a logistic growth pattern, where a population levels off to some carrying capacity based on its respective environment. From Part 1, where we modeled the spread of dandelions, we can see that dandelions satisfy Assumption 10; therefore, we can apply the Gordon-Shaefer model.

To find the maximum sustainable profit of a foreign species, we construct an industry centered around harvesting this foreign species. Hence, we can apply the microeconomic *Profit Maximization Law*, where the optimal level of harvest, otherwise known as the maximum sustainable yield, can be found at the intersection of *marginal revenue* and *marginal cost* curves (MR=MC) [15].

In the context of a Gordon-Schaefer harvesting model, marginal revenue can be the revenue per harvest unit, and marginal cost can be the cost to harvest each species unit. As the marginal revenue decreases over time due to the Law of Diminishing Marginal Returns [20], we determine the optimal yield effort, which is the intersection of the MR and MC curves. Then, we find the corresponding point on a *Total Revenue* and *Total Cost* curve to obtain the maximum profit (π). See Figure 14 for a visual representation of the process. Note that each species will have a unique Gordon-Schaefer bioeconomic model calibrated based on industry data.
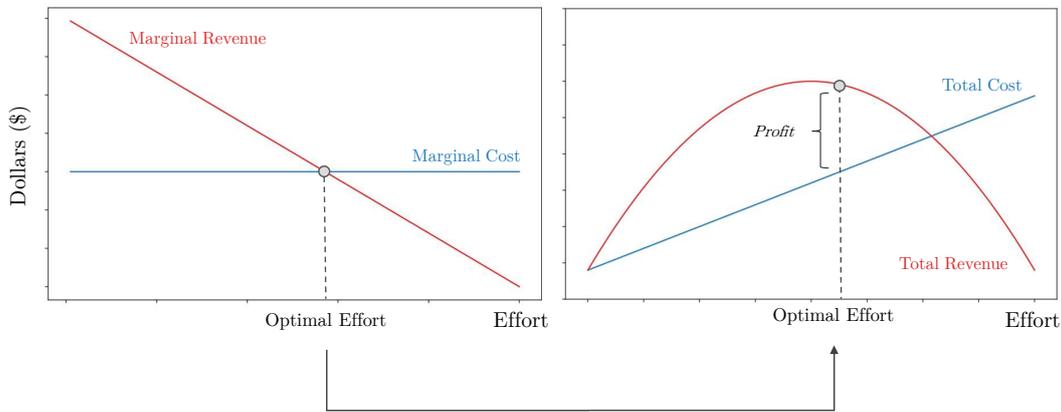
Figure 14: **Gordon-Schaefer model effort optimization for species harvesting.** Net Profit (π) can be obtained from the difference between Total Revenue and Total Cost.

5.5.2 Social Benefit

To calculate the social welfare benefit of a foreign species on society, we can calculate the *total social welfare* received by a population surrounding the foreign species and region in question. To determine the explicit dollar-value benefits of social welfare, we can utilize Sen's Welfare Function (which incorporates a Gini coefficient) to model the benefits of a particular species scaled to the region's income inequalities [48, 3].

Therefore, we can express the *total welfare benefit* (USD) W as a sum of individual benefits Y_i, Gini coefficient G, and regional population n,

$$W = (1 - G) \sum_{i=1}^{n} Y_i \tag{15}$$

5.5.3 Human Health Risk

To calculate human health risk as an economic factor, we analyze the average detriment a species has to humans as a sum of individual risks based on the potential amount of money lost. The sum can again be scaled similarly to Sen's Welfare Function using the Gini coefficient. Data for the detrimental costs of a species to human society can be found for virtually any species, allowing us to assume generalizability [23].

Therefore, we express the *total human risk* (USD) R as a sum of individual risks R_i, Gini coefficient G, and regional population n,

$$R = (1 - G) \sum_{i=1}^{n} R_i \tag{16}$$

Therefore, we can express the total economic benefit of a given species in a region as the sum of economic profit (π) and social benefit (W) minus the human health risk (R).

$$\text{Total Economic Benefit } = \pi + W - R \tag{17}$$

So, we combine the Ecological (Equation 14) and Economic (Equation 17) sub-scores as a coordinate pair to represent the trade-off between the two.

$$\textbf{Impact Score } = (\text{Ecological Score, Economic Score}) = (100r + \chi,\, \pi + W - R) \tag{18}$$

6   APPLICATION OF TRADE-OFF MODEL

By calculating our two scores, ecological and economic, we can apply our impact score calculations to virtually any foreign species. For the purposes of our discussion, we chose two additional foreign plant species from the United States region: *dandelions* in New York, *garlic mustard plants* in Illinois, and *English ivy* plants in Washington. These foreign plants have practical social benefits and negatively affect their new environments [39, 13, 25].

6.1   Dandelion Plants

6.1.1   Ecological Impact Score

As stated in the previous section, as part of determining the ecological impact score of dandelions, we need to examine their effect on native species. Thus, we chose the following ecologically related native species where dandelions reside in New York.

- **Bombus impatiens** (Bumblebees) — native pollinators common to eastern North America that visit dandelion species [16].

- **Malacosoma americanum** (Moths) — feeds on the leaves of the common dandelion in New York. A consumer-plant relationship with dandelions [6].

- **Cuscuta gronovii** (Scaldweed) — a climbing parasitic vine that extracts nutrients from attached-to plants [49].

Below, we determined an interaction matrix for the Lotka-Volterra model based on the interactions of each species in row i to the species in column j. For example, if species j was a natural predator of species i, the value of α_{ij} will be negative. If two species were mutualistic, we would expect that the element α_{ij} would be positive, thus improving both population's growth rates. Note that generally, element $\alpha_{ij} \neq \alpha_{ji}$ because the effect species i has on species j is generally opposite to the effect in the reverse order. Furthermore, the impact of the foreign species on native populations is found in the fourth row and fourth column of the matrix α.

Assuming a starting population of 100 plant units and a carrying capacity of 1000 plant units each, and a growth rate of 0.005 for bumblebees, as their populations mainly fluctuate by seasons, 0.06 for moths as their populations grow quickly, 0.01 for vines as their populations tend to stabilize, and 0.04 for dandelions as shown in the following logistic regression section [16, 6, 49]. We solve the systems of differential equations (Equation 10) with the following approximated interaction matrix.

$$\alpha_{\text{ dandelion}} = \begin{array}{c} \text{Bumblebees} \\ \text{Moths} \\ \text{Vines} \\ \text{Dandelions} \end{array} \underbrace{\begin{bmatrix} 1 & 0.25 & -0.5 & 2 \\ 0.25 & 1 & 0.6 & 0 \\ -0.5 & -0.5 & 1 & 2 \\ 2 & 0.25 & -1 & 1 \end{bmatrix}}_{\text{3 native species, 1 foreign species}} \tag{19}$$

From graphing the population solutions to the system, we can visualize the population changes in Figure 15. We find the change in populations over a single year to be $\frac{1200}{100}$ for moths, $\frac{250}{850}$ for vines, and $\frac{550}{100}$ for bumblebees. So, we can calculate the native species impact score to be
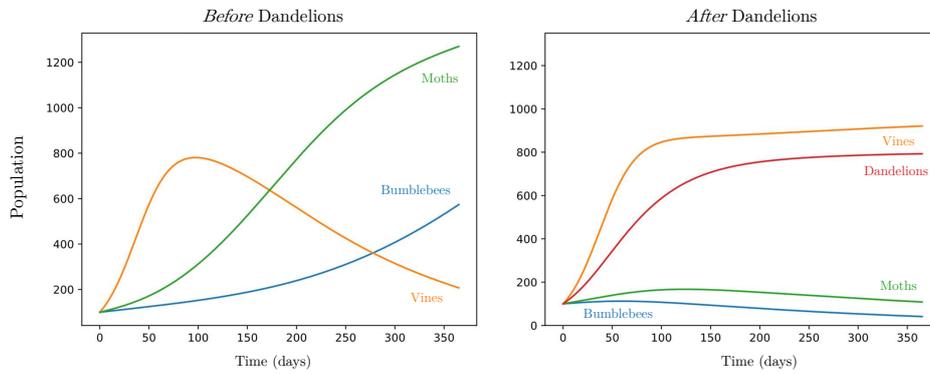
Figure 15: **Impact of dandelion plants on native species populations**.

$$\chi = \frac{1200}{100} + \frac{250}{850} + \frac{550}{100} = 17.794$$

Next, we calculate the spreadability index by performing logistic regression on dandelion population data from the previous model. We find that the carrying capacity of a region is approximately 900 dandelion plants (see Figure 6). Therefore, we calculate from logistic regression that the proportionality of growth is $r = 0.04$; therefore, our spreadability score is $100r = 4$.

Finally, we see that our total ecological impact score is $4 + \chi = 4 + 17.794 = 21.794$ (Equation 14).

6.1.2 Economic Potential

We constructed a small dandelion harvesting firm in New York to estimate economic profits. First, we can approximate the marginal cost of harvesting each dandelion to be around \$0.10 per pound of dandelions [35, 19]. Next, we can also approximate the marginal revenue curve. Marginal revenue decreases for every five pounds of dandelions sold with maximum marginal revenue at around \$4.50 per pound of dandelions [51]. So, we can model marginal revenue with a linear model with a slope of -0.2 and a y-intercept of 4.5. By integration, we can calculate the total revenue and cost curves based on effort levels, shown in Figure 16. Therefore, we calculate dandelion profits ($\pi_{\text{dandelions}}$) be around \$47.86.
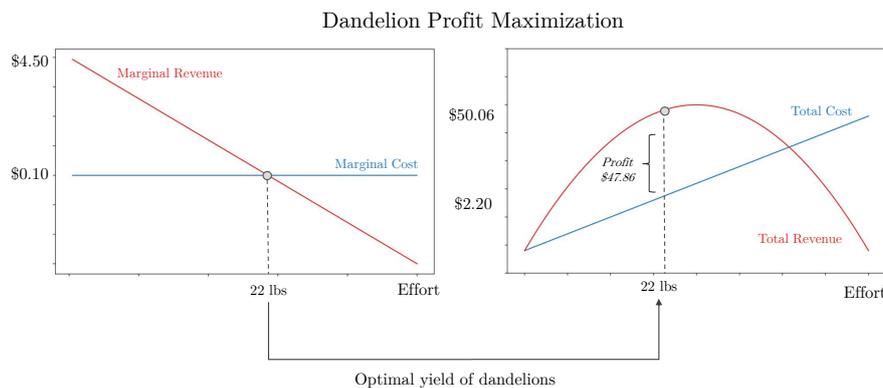
Figure 16: **Gordon-Schaefer model for dandelion profit yield.**

Additionally, we need to calculate the social welfare benefits and potential human risks. The typical population of an upstate New York town is approximately $n = 60,000$, and the net external benefit for each individual is approximately \$0.002 as dandelions are a rarely used medicinal plant and scarcely found soil loosener [29]. Lastly, New York's Gini coefficient is 0.52 [37]. Incorporating these aspects, the social welfare output is:

$$W = (1 - 0.52) \sum_{i=1}^{60000} 0.002 = \$57.60$$

.

Finally, we need to find the human health risk by calculating the amount of harm a dandelion has on its surrounding human population. Because dandelions do not generally seriously harm humans, we assign the potential human risk to be $R = 0$.

With the three components of the economic *potential* score, we can calculate the total economic *benefit* score to be

$$\text{Total Economic Benefit } = \pi + W - R = 47.86 + 57.60 - 0 = \$105.46$$

So, our final invasive species impact score is

$$\text{Impact Score} = (\text{Ecological Score, Economic Score}) = (21.794, \$105.46)$$

6.2 Garlic Mustard Plants

6.2.1 Ecological Impact Score

Next, to determine the ecological impact score of garlic mustard plants, we chose the following ecologically related native species where garlic mustard plants reside in Illinois.

- **White trillium** (Wildflower) — a native wildflower that grows in woodlands and is often threatened by garlic mustard invasion [33].

- **Plethodon cinerus** (Salamander) — an amphibian that lives under leaves and feeds on insects [46].

- **Hylocichla mustelina** (Wood thush) — a songbird that nests in forests where garlic mustard invasion reduces vegetation for nesting cover [24].

Below, we again determine an interaction matrix for the Lotka-Volterra model. Note that we again assume that the starting population of each species is 100 units with a carrying capacity of 1000 plant units each. We also assume a growth rate of 0.06 for wildflowers, as their populations mainly fluctuate by seasons, 0.03 for salamanders as their populations do not grow as quickly, 0.01 for songbirds as their populations tend to stabilize, and 0.1 for garlic mustard plants as their populations tend to proliferate. We finally solve the systems of differential equations shown in Equation 10 with the following interaction matrix (Equation 20).

$$\alpha_{\text{ garlic mustard}} = \begin{matrix} \text{Wildflower} \\ \text{Salamanders} \\ \text{Songbirds} \\ \text{Garlic mustard} \end{matrix} \underbrace{\begin{bmatrix} 1 & 0.75 & 1.5 & -1 \\ 0.35 & 1 & 0.65 & -1 \\ 0.5 & 0.5 & 1 & -1 \\ 0.25 & 0.25 & 0.15 & 1 \end{bmatrix}}_{\text{3 native species, 1 foreign species}} \tag{20}$$

In the fourth column of $\alpha_{\text{ garlic mustard}}$, we assigned a value of -1 to each element since garlic mustard plants harm all three native species populations.

By graphing the population solutions to the system, we can see the population changes in Figure 17. We find the ratio of change of populations over a single year to be $\frac{150}{100}$ for wildflowers, $\frac{800}{450}$ for salamanders, and $\frac{400}{200}$ for songbirds. So, we can calculate the native species impact score to be

$$\chi = \frac{150}{100} + \frac{800}{450} + \frac{400}{200} = 5.278$$

To calculate the spreadability index, we can start from an accepted fact that garlic mustard populations can double in a year in most growing conditions [23]. Therefore, by performing logistic regression with the x-axis being days and the y-axis as population, we observe that the r value is around 0.05. So, our spreadability index can be calculated as $100r = 5$. Thus, by Equation 14, our ecological impact score is $5 + \chi = 5 + 5.278 = 10.278$.

6.2.2 Economic Potential

As harvesting efforts for garlic mustard plants are minimal (we could not find any garlic mustard plant products via an online search), we conclude there is no possibility of a garlic mustard harvesting firm. This means that there is *no potential economic profit* ($\pi = 0$).

Next, we calculate a garlic mustard plants' social welfare benefits. The population of a typical town in Illinois is approximately $n = 50,000$, and the net external benefit for each individual person is similar to
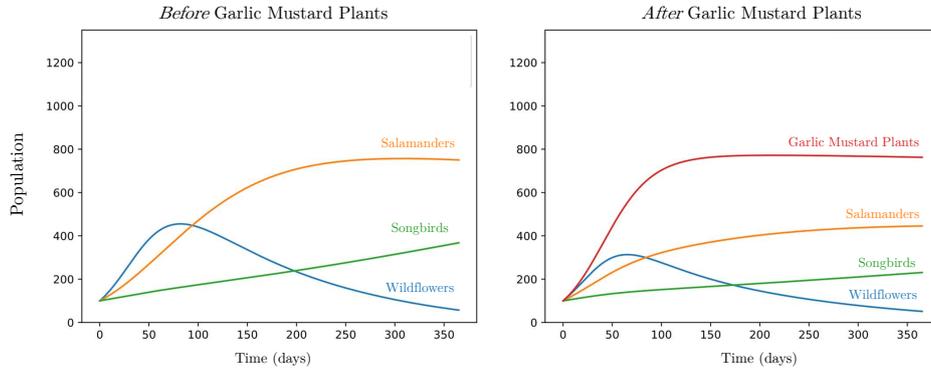
Figure 17: **Impact of garlic mustard plants on native species populations**.

dandelions at approximately \$0.001 as garlic mustard plants are a rarely used medicinal herb [23]. Furthermore, the Gini coefficient for Illinois is 0.48 [36], hence, the social welfare output is

$$W = (1 - 0.48) \sum_{i=1}^{50000} 0.001 = \$26.00$$

.

Lastly, we determine the human health risk by calculating the amount of harm garlic mustard plants have on its surrounding human population. Because dandelions do not generally seriously harm humans, we assign the potential human risk to be $R = 0$. [23].

With the three components of the economic potential score, we can calculate the score to be

$$\text{Total Economic Benefit } = \pi + W - R = 0 + 26.00 - 0 = \$26.00$$

So, our final invasive species impact score for garlic mustard plants is

$$\text{Impact Score} = (\text{Ecological Score}, \text{Economic Score}) = (5.278, \$26.00)$$

6.3 English Ivy Plants

6.3.1 Ecological Impact Score

To determine the ecological impact score of English ivy plants, we examine the top 3 most closely related native species where English ivy plants reside in Washington:

- **Kinnikinnick** (Shrub) — an evergreen shrub that grows in most conditions and attracts a variety of native pollinators and birds [50].

- **Asarum caudatum** (Wild ginger) — a low-lying groundcover plant with edible leaves [21].

- **Mahonia aquifolium** (Oregon grape) — another evergreen plant that produces berries and can grow in most temperate conditions in Washington [41].

For the last time, we determine an interaction matrix for the Lotka-Volterra model. We again assume that each species' starting population is 100 units with a carrying capacity of 1000 plant units each. We assume a growth rate of 0.02 for shrubs, as their populations mainly fluctuate by season, 0.04 for wild gingers as their populations grow rapidly across forest floors, 0.01 for Oregon grapes as their populations tend to stabilize, and 0.4 for English ivy plants as their populations tend to multiply. We solve the new Lotka-Volterra system with the following interaction matrix.

$$\alpha_{\text{english ivy}} = \begin{array}{c} \text{Shrubs} \\ \text{Wild ginger} \\ \text{Oregon grape} \\ \text{English ivy} \end{array} \underbrace{\begin{bmatrix} 1 & 0 & -0.5 & -0.25 \\ 0 & 1 & 0 & -1 \\ -0.5 & 0 & 1 & -1 \\ 0.25 & -1 & 0.15 & 1 \end{bmatrix}}_{\text{3 native species, 1 foreign species}} \tag{21}$$

In α english ivy, several elements are set to zero because they live in independent habitats, meaning they rarely compete for the same natural resources. However, English ivy plants compete vigorously with wild ginger, so the respective elements are -1 [21].
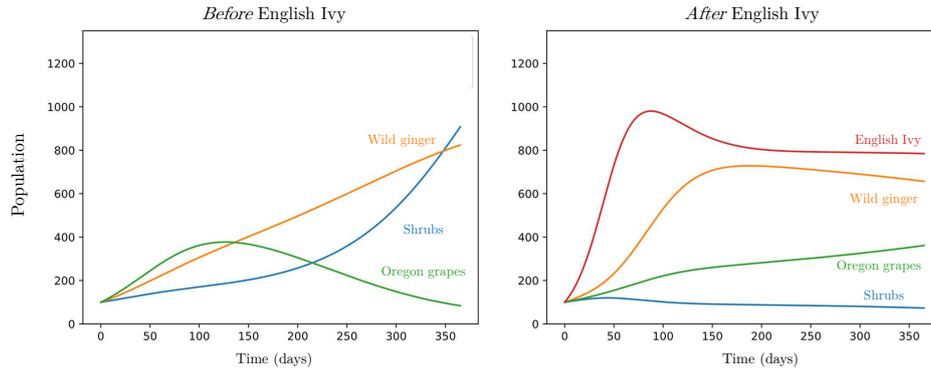
Figure 18: **English ivy impact on native species populations.**

By graphing the population solutions to the system, we can see the population changes in Figure 18. We find the change of populations over a single year to be $\frac{900}{90}$ for shrubs, $\frac{800}{600}$ for wild ginger, and $\frac{100}{300}$ for Oregon grapes. So, we calculate the native species impact score to be

$$\chi = \frac{900}{90} + \frac{800}{600} + \frac{100}{300} = 11.667$$

Next, we determine the spreadability index. We can start with an accepted fact that English ivy populations can double in around one and a half years in most growing conditions, even in the shade [34]! Therefore, by performing logistic regression with the x-axis being days and the y-axis as population, we find that the r value is around 0.035. So, our spreadability index is calculated as $100r = 3.5$. Therefore, by Equation 14, our ecological impact score is $3.5 + \chi = 3.5 + 11.667 = 15.167$.

6.3.2 Economic Potential

We assume that there exists some small landscaping firm that harvests and grows English ivy plants in a typical town in Washington, as many landscaping designers use English ivy to decorate their clients' backyards [42]. By analyzing the microeconomic model of this firm, we will be able to calculate our economic profits.

First, we estimate the expected marginal cost and revenue of harvesting and growing each pound of English ivy based on the Gordon-Schaefer model. We estimate that the constant marginal cost of harvesting each pound of English ivy is around \$15.00, as it takes around an hour of manual labor to safely harvest English ivy [18].

To model marginal revenue, we surveyed online shopping sources to estimate the revenue per pound of English ivy. Based on several pricing estimates [14], we conclude that the maximum pricing for a pound of English ivy would fall around \$100.00 and decrease with slope -0.5 per pound increased of English ivy. After applying the Profit Maximization Law, analyzing this firm with the Gordon-Schaefer model, and finding the intersection of the MR and MC curve, an optimal English ivy yield would be around 170 lbs for \$57.75 per pound.

To calculate economic profits (π), we subtract total costs (\$15.00 · 170) from total revenue (\$57.75 · 170), which leads us to conclude that $\pi = \$7225.00$.

As for measuring social welfare, English ivy has no benefits to humans other than as a landscaping plant, which is already an integral part of economic profits. Lastly, for the human risk factor, English ivy is inedible and contains toxic sap (yuck!). However, English ivy plants rarely harm humans. Therefore, we deduce that social welfare $W = 0$ and human risk $R = 0$. So, our total economic potential can be calculated as

$$\text{Total Economic Benefit } = \pi + W - R = \$7225.00$$

So, our final invasive species impact factor for English ivy plants in Washington is

$$\text{Impact Score} = (\text{Ecological Score}, \text{Economic Score}) = (15.167, \$7225.00)$$

6.4 Score Comparison

To analyze the differences in each score and the "classification" of each foreign species, we look at the trade-offs between ecological and economic scores in our definition of an impact factor.

Foreign Species	Ecological	Economic	Analysis
Dandelions	21.794	$105.46	As dandelions have a high ecological impact primarily due to the loss of native populations with a low return in economic potential, we conclude that dandelions are indeed *invasive species*.
Garlic mustard	5.278	$26.00	Since garlic mustard plants have a slight ecological impact and have little to no return on economic potential, we label garlic mustard plants as *invasive species*.
English ivy	15.167	$7225.00	As English ivy plants have a high impact on the environment yet also have a large return on economic potential, we conclude that English ivy plants *are not invasive*.

7 INVASIVE SPECIES TRADE-OFF MODEL DISCUSSION

7.1 Model Generalization

When applying our model to a larger sample of foreign plants to classify them as invasive or not invasive, it would be immensely beneficial to create an auto-classification system. Such a classification system would consist of a "line of division" in a plot where each axis would represent the ecological or economic scores. Since we do not have enough data points to create such a classification system, we design a *DSD* framework (Data-Score-Divide) below where an environmental scientist may use our trade-off model.

In our proposed *DSD* framework, environmental scientists would perform the following steps as described in Figure 19. First, a scientist must collect regional native population data, economic costs and revenues from similar firms, and determine dollar-value human risk costs in local populations. Next, data is to be evaluated via our proposed ecological and economic scores. Finally, scores for each foreign species should be plotted on an economic versus ecological score plot, where a line of division must be drawn by an evaluator to specify the split between invasive and simply foreign species. This division process is inspired by Support Vector Machines, which are commonly used in machine learning classification tasks [17].

We did not conduct a sensitivity analysis on our trade-off model because, in a sense, the environmental scientists who *divide* the scores are the primary influences of the final decision.

The *DSD* Process: Data, Score, Divide
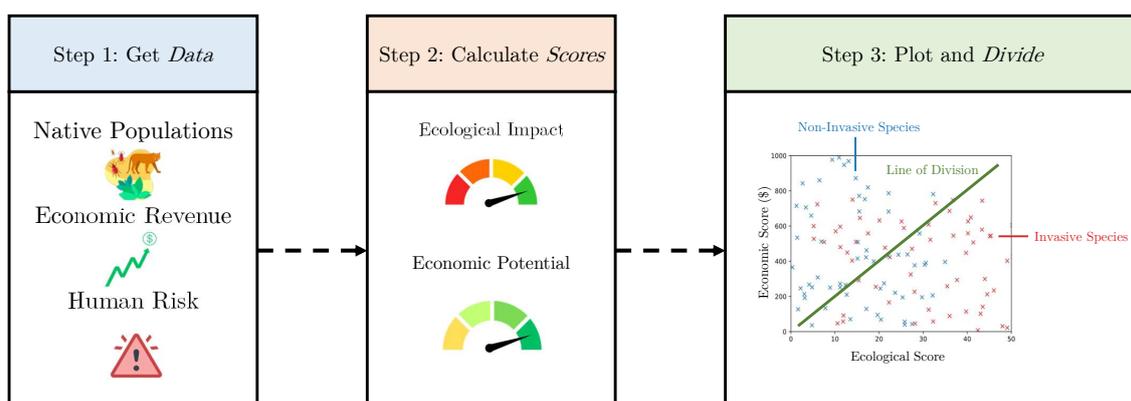
Figure 19: **The *Data-Score-Divide* (DSD) framework for evaluating foreign species.**

7.2 Strengths

1. Our model offers unique **insights into the trade-offs** of foreign species, in particular, invasive species.

   Unlike impact models that only examine the ecological benefits and downfalls of a foreign species, our model examines both ecological and economic trade-offs.

2. Our model is adaptable to virtually **any region and foreign plant** combination.

   This means that our model can be applied to different scenarios and contexts, such as invasive species management, biodiversity conservation, and ecological restoration. Our model can adapt based on different regional native plants and economic conditions. We also provide a general-use framework that can be quickly applied to many species at a time.

7.3 Limitations

1. Intra-population dynamics between foreign and native species that are unknown **can be difficult to predict**.

   In the previous three contexts of analyzing the trade-offs between dandelion, garlic mustard, and English ivy plants, the most distinct aspect of these species is that they are widely classified as invasive or a "weed." This means that for unknown foreign species relationships, scientists must input some observational data into the model.

2. New economic benefits and pitfalls **may not follow historical trends**.

   In our analysis of dandelion and English ivy plants, several businesses centered around dandelion and English ivy harvesting already exist. This allows us to build upon previous work to calculate new economic profits and yield. However, when we take a look at the case of garlic mustard plants, few garlic mustard firms exist, which hinders our ability to predict economic outcomes.

8   Conclusion

In this paper, we developed a mathematical model to predict the spread of dandelions and evaluate whether they are genuinely invasive. Our seed agent model considers the different phases of a dandelion's life cycle, including seed dispersal, germination, plant development, and the puffball stage. The model also considers environmental factors such as temperature, light, and soil nutrients. The paper discusses the application of the model to three different regions and analyzes the ecological impact and economic potential of other invasive species.

To evaluate foreign species' trade-offs, we formulate an impact score based on ecological and economic factors, which evaluated dandelions and garlic mustard plants as invasive and English ivy plants as simply foreign. Furthermore, our model is adaptable as we provide a unique framework for different regions and foreign plant combinations.

We hope that our model can be a step forward in understanding how dandelions and other foreign plants can spread in a new environment and aid the understanding of the most impactful environmental factors on dandelion growth. In the future, we aim to share our models with environmental scientists worldwide and fine-tune them to even more region-plant combinations. We also envision that our model can be showcased with an easy-to-use graphical interface, adjustable hyperparameters, and a place for data upload—all of which can significantly benefit foreign species analysis.

9 References

[1] Aladin. Wolcott, NY - Climate & Monthly weather forecast.

[2] Arizona Agricultural Division. Guide to Arizona Agriculture.

[3] W. Bank. World Bank Open Data — data.worldbank.org. `https://data.worldbank.org/indicator/SI.POV.GINI`.

[4] Board of Pesticides Control: Maine DACF. "Dandelion-Taraxacum Officinale." Dandelion: Got Pests? *Board of Pesticides Control: Maine DACF*, 2022.

[5] W. J. W. Botzen, O. Deschenes, and M. Sanders. The Economic Impacts of Natural Disasters: A Review of Models and Empirical Studies. *Review of Environmental Economics and Policy*, 13(2):167–188, July 2019. Publisher: The University of Chicago Press.

[6] Butterflies and M. of North America. Eastern Tent Caterpillar Moth Malacosoma americanum (Fabricius, 1793).

[7] C. CALS. Dandelion | CALS.

[8] O. Calzone. MAE, MSE, RMSE, and F1 score in Time Series Forecasting, Apr. 2022.

[9] W. Chepil. Germination of Weed Seeds I. Longevity, Periodicity of Germination, and Vitality of Seeds in Cultivated Soil. *Scientific Agriculture*, 26(7):307–346, July 1946. Publisher: NRC Research Press.

[10] W. . Climate. Average monthly hours of sunshine in Key West (Florida), the United States of America.

[11] W. . Climate. Average monthly hours of sunshine in Phoenix (Arizona), the United States of America.

[12] G. Communications. Interpreting Phosphorus and Potassium Levels, Dec. 2018.

[13] T. N. Conservancy. Garlic Mustard: Invasive, Destructive, Edible — nature.org.

[14] G. G. Direct. English ivy. `https://gardengoodsdirect.com/products/english-ivy`.

[15] I. Economist. The Profit Maximization Rule | Intelligent Economist — intelligenteconomist.com.

[16] N. W. Federation. Common Eastern Bumble Bee. `https://www.nwf.org/Educational-Resources/Wildlife-Guide/Invertebrates/Common-Eastern-Bumblebee`.

[17] M. A. Hearst. Support vector machines — ieeexplore.ieee.org. `https://ieeexplore.ieee.org/document/708428`.

[18] B. Homes and Gardens. Caring for Easy-to-Grow English Ivy as a Houseplant. `https://www.bhg.com/gardening/plant-dictionary/houseplant/english-ivy/`.

[19] G. K. How. Dandelion Harvest Guide – How To Harvest Dandelion Plants.

[20] Investopedia. Law of Diminishing Marginal Returns: Definition, Example, Use in Economics — investopedia.com. `https://www.investopedia.com/terms/l/lawofdiminishingmarginalreturn.asp`.

[21] P. Nursery. Asarum caudatum: Wild Ginge. `https://www.portlandnursery.com/natives/asarum`.

[22] T. A. Obreza. Importance of Potassium in a Florida Citrus Nutrition Program, 2003.

[23] F. of Mississippi River. Invasive species: Garlic mustard. `https://fmr.org/invasive-species-garlic-mustard`.

[24] C. L. of Ornithology. Wood Thrush Overview, All About Birds. `https://www.allaboutbirds.org/guide/Wood_Thrush/overview`.

[25] I. Org. English Ivy (Hedera helix) — invasive.org. `https://www.invasive.org/alien/pubs/midatlantic/hehe.htm`.

[26] G. Organic. Garden Organic | Dandelion.

[27] M. Ottman. Arizona Grain Research and Promotion Council September, 2008, 2008.

[28] W. L. Pritchett, C. F. Eno, and M. N. Malik. The Nitrogen Status of the Mineral Soils of Florida. *Soil Science Society of America Journal*, 23(2):127–130, 1959. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj1959.03615995002300020015x.

[29] A. H. Rachael Ajmera. Dandelion: Health Benefits and Side Effects — healthline.com. `https://www.healthline.com/nutrition/dandelion-benefits`.

[30] ScienceDirect. Lotka-Volterra Model - an overview | ScienceDirect Topics.

[31] SciPy . Scipy.

[32] N. W. Service. NOWData - NOAA Online Weather Data. Publisher: NOAA's National Weather Service.

[33] U. F. Service. Great White Trillium. `https://www.fs.usda.gov/wildflowers/plant-of-the-week/trillium_grandiflorum.shtml`.

[34] U. F. Service. Hedera helix. `https://www.fs.usda.gov/database/feis/plants/vine/hedhel/all.html`.

[35] F. Show. Growing Dandelions For Dollars. `https://www.farmshow.com/a_article.php?aid=20546`.

[36] . W. St. How Income Inequality in Illinois Compares to Other States. `https://247wallst.com/state/how-income-inequality-in-illinois-compares-to-other-states/`.

[37] Statista. Gap between rich and poor by state U.S. 2022.

[38] S. M. Stewart-Wade, S. Neumann, L. L. Collins, and G. J. Boland. The biology of Canadian weeds. 117. *Taraxacum officinale* G. H. Weber ex Wiggers. *Canadian Journal of Plant Science*, 82(4):825–853, Oct. 2002.

[39] M. Szydlowski. Dandelions: Fight them or love them? — columbiatribune.com.

[40] M. S. University. Dandelion.

[41] O. S. University. Mahonia aquifolium | Landscape Plants. `https://landscapeplants.oregonstate.edu/plants/mahonia-aquifolium`.

[42] P. University. English Ivy in the Landscape. `https://extension.psu.edu/english-ivy-in-the-landscape`.

[43] E. University of Minnesota. Potassium for crop production.

[44] L. A. Urray. Campbell biology, 2016.

[45] L. Wang, S. Li, and L. Zhang. Separating mechanism of dandelion (Taraxacum officinale) seed: Results from model analysis and separation force measurement. *Wear*, 477:203846, July 2021.

[46] A. Web. AmphibiaWeb - Plethodon cinereus. `https://amphibiaweb.org/species/4126`.

[47] Wikipedia. Gordon-Schaefer model - Wikipedia — en.wikipedia.org. `https://en.wikipedia.org/wiki/Gordon-Schaefer_model`.

[48] Wikipedia. Social welfare function - Wikipedia — en.wikipedia.org. `https://en.wikipedia.org/wiki/Social_welfare_function#:~:text=In%20welfare%20economics%2C%20a%20social,possible%20pair%20of%20social%20states`.

[49] M. Wildflowers. Cuscuta gronovii (Swamp Dodder). `https://www.minnesotawildflowers.info/flower/swamp-dodder`.

[50] WNPS. Plant Profile: Kinnikinnick (Arctostaphylos uva-ursi). `https://www.wnps.org/blog/plant-profile-kinnikinnick`.

[51] M. Young. How to Make Money Selling Dandelions — farmfitliving.com. `https://farmfitliving.com/make-money-selling-dandelions/`.

10   Appendices

10.1   Appendix 1: Seed Agent Modeling Code

./seedagentmodel.py

```python
1   import numpy as np
2
3   ### Hyperparameters ###
4   S0 = 2000 # seeds dispersed per puff ball
5   ld_death_seeds = 4 # average life expectancy of seeds in days
6   ld_death_dandelion = 15 # average time to get consumed
7   germ_mean = 12 # average germination time in days
8   optimal_time_to_grow = 50 # optimal time to grow in days
9   sim_length = 365 # simulation length in days
10  min_dist = 1 # minimum distance between seeds in meters
11  side_length = 100 # side length of square plot in meters
12
13  ### DEFAULT (TEMPERATE) REGIONAL DATA ###
14  wind_mean = [int(np.random.normal(loc=100, scale=100)) for i in range(365)] #
    average wind speed in mph
15  temperatures = [int(20*np.sin(2*np.pi* x/(360))+55) for x in range(365)]
16  light_hours = [int(6*np.sin(2*np.pi* x/(360))+10) for x in range(365)]
17  k = [int(np.random.randint(10, 100)) for x in range(365)]
18  n = [int(1.05*np.random.randint(0, 50)) for x in range(365)]
19
20  class Seed:
21      ### CONSTRUCTOR ###
22      def __init__(self, t, x, y):
23          self.t = t
24          self.growth_state = 0
25          self.coordinate = [x, y]
26          self.death_date = self.get_death_day(t, ld_death_seeds)
27          self.germination_date = self.get_germination_date(t+1)
28          self.germination_state = 0
29
30      ### STAGE 1: BROWNIAN MOTION ###
31      def brownian_motion(self):
32          ### POISSON DISTRIBUTION FOR NUM STEPS ###
33          def num_steps(t):
34              m = [abs(wind_mean[t % 365]) for x in range(365)]
35              num = int(np.random.poisson(m[t % 365], 1))
36              return num
37
38          ### BROWNIAN RANDOM WALK ###
39          steps = num_steps(0)
40          for i in range(steps):
41              self.coordinate[0] += np.random.normal(loc=0, scale=2)
42              self.coordinate[1] += np.random.normal(loc=0, scale=2)
43
44      ### STAGE 2: CALCULATE DEATH DATE ###
45      def get_death_day(self, t, mean):
46          ld = [int(mean*np.sin(2*np.pi* x/(360))+(mean/2)) for x in range(365)]
47          #death_time = t+int(np.random.exponential(ld[t % 365]))
48          death_time = t+int(np.random.exponential(mean))
49          return death_time
50
51      ### STAGE 3: SEED GERMINATION ###
52      def get_germination_date(self, t):
53          return t+int(np.random.normal(germ_mean, germ_mean))
54
55      def dGdt(self, t):
56          def T_I(temp, max=77, min=50):
57              middle = (max+min) / 2
58              if temp > max:
59                  return -2 / (1 + np.exp(-0.1*(temp - max))) + 1
60              elif temp < min:
61                  return -2 / (1 + np.exp(-0.1*(min-temp))) + 1
62              else:
63                  return 1 - (1/(middle-min)**2) * (temp-middle)**2
64
65          def L_I(light_hrs, max=24, min=6):
66              middle = (max+min) / 2
67              if light_hrs > max:
68                  return -2 / (1 + np.exp(-0.5*(light_hrs - max))) + 1
69              elif light_hrs < min:
70                  return -2 / (1 + np.exp(-0.5*(min-light_hrs))) + 1
71              else:
72                  return 1 - (1/(middle-min)**2) * (light_hrs-middle)**2
73
74          def N_I(k, n, k_max=80, k_min=40, n_max=40, n_min=20):
75              k_avg = (k_max+k_min)/2
76              n_avg = (n_max+n_min)/2
77              k_score = np.exp(-((k-k_avg)/(k_max-k_min))**2)
78              n_score = np.exp(-((n-n_avg)/(n_max-n_min))**2)
79              return k_score + n_score - 1
80
81          return (1/(3 * optimal_time_to_grow)) * (T_I(temperatures[t % 365]) +
                L_I(light_hours[t % 365]) + N_I(k[t % 365], n[t % 365]))
82
83  ### INITIAL STATE ###
84  seeds = [Seed(0, 0, 0) for i in range(S0)]
85  for seed in seeds:
86      seed.brownian_motion()
87      if (seed.coordinate[0] < 0 or seed.coordinate[1] < 0 or seed.coordinate[0] >
        side_length or seed.coordinate[1] > side_length):
88          seeds.remove(seed)
89
90  ### FILTERING COORDINATES
91  def euclidean_distance(coord1, coord2):
92      x1, y1 = coord1
93      x2, y2 = coord2
94      return np.sqrt((x2 - x1) ** 2 + (y2 - y1) ** 2)
95
96  def remove_close_seeds(seeds, threshold=min_dist):
97      filtered_coordinates = []
98      n = len(seeds)
99      for i in range(n):
100         keep = True
101         for j in range(i + 1, n):
102             if euclidean_distance(seeds[i].coordinate, seeds[j].coordinate) <
                threshold:
103                 keep = False
104                 break
105         if keep:
106             filtered_coordinates.append(seeds[i])
107     return filtered_coordinates
108
109 seeds = remove_close_seeds(seeds)
110 population_seeds = []
111 population_plants = []
112 ### BEGIN SIMULATION ###
113 for t in range(1, sim_length):
114     x = sum(1 for seed in seeds if seed.germination_state == 0)
115     y = sum(1 for seed in seeds if seed.germination_state == 1)
116     population_seeds.append(x)
117     population_plants.append(y)
118
119     for i in range(len(seeds)-1, -1, -1):
120         seed = seeds[i]
121
122         ### CHECK DANDELION DEATH ###
123         if seed.death_date == t and seed.germination_state == 0: # before growth
            stage
124             seeds.pop(i)
125         elif seed.death_date == t and seed.germination_state == 1: # in growth
            stage
126             seed.growth_state = 0
127             seed.death_date = seed.get_death_day(t, ld_death_dandelion)
128             seed.t = t
129
130         ### CHECK GERMINATION ###
131         if seed.germination_date == t:
132             seed.germination_state = 1
133             seed.death_date = seed.get_death_day(t+1, ld_death_dandelion)
134
135         ### CHECK PLANT GROWTH ###
136         if seed.germination_state == 1:
137             seed.growth_state += seed.dGdt(t)
138
139             ### REACHES PUFF BALL STAGE
140             if (seed.growth_state < 0):
141                 seed.growth_state = 0
142
143             if (seed.growth_state >= 1):
144                 seed.growth_state = 0
145                 seed.death_date = seed.get_death_day(t+1, ld_death_dandelion)
146
147                 ### BEGIN BROWNIAN MOTION AGAIN
148                 new_seeds = [Seed(t, seed.coordinate[0], seed.coordinate[1]) for
                    i in range (S0)]
149                 for new_seed in new_seeds:
150                     new_seed.brownian_motion()
151                     if (new_seed.coordinate[0] < 0 or new_seed.coordinate[1]< 0
                        or new_seed.coordinate[0] > side_length or
                        new_seed.coordinate[1] > side_length):
152                         new_seeds.remove(new_seed)
153                 seeds += new_seeds
154                 seeds = remove_close_seeds(seeds)
```