Team Control Number

13719

Problem Chosen

A

2023
HiMCM/MidMCM
Summary Sheet

Every Single One Matters

Invasive plants are always severe problems for human beings. They always have a negative influence on the local ecosystem, economy, and even people's health. The dandelion is one of the annoying invasive plants for gardeners. It is known for its ability to spread across a large area and adapt to different conditions. Many factors can influence the growth of dandelions, such as temperature, rainfall, and sunlight. In our work, we build a model to simulate, visualize, describe, and discuss the impact of different factors.

The first problem requires a model to predict the dispersion of dandelions, considering climate factors. To this end, we extract various climate, geography, and biodiversity data from the **Google Earth Engine**. Our model incorporates temperature in a proposed **germination model**; temperature, humidity, solar radiation, and pH in the **growth model**; and wind speed, direction, and properties of plants in the **seed spreading model**. We utilize the **Cellular Automata** concept and **simulate the dispersion of three invasive plants in 14 locations**. Results reveal significant differences in the coverage rate among different locations. The sensitivity analysis shows clear dependency and correlation of climate effects based on our model.

An invasive impact factor should solve the second problem to describe the level of impact of an invasive plant in a region. We choose **various objective and subjective indicators**, including climate, human activities, vegetation, etc. The simulation results from the first problem are also extracted as indicators. We use the **Analytic Network Process** and Technique for Order Preference by Similarity to Ideal Solution (or **TOPSIS**) to estimate accessing factors. The impact of three kinds of invasive plants, including dandelions, **Centaurea Solstitialis** and **Solidago Canadensis**, in 14 worldwide locations is quantified by the proposed Impact Factor which incorporates TOPSIS. Fitting on data of dandelions and evaluating the fitted model on the other two plants, results show that the Impact Factor could accurately identify locations that have been invaded.

In conclusion, by building a mathematical model, we can have a deep understanding of the spread of these invasive plants and quantify their impact factor. In this case, we can control these plants better and make good use of them. If we know them quite well and make good use of them, instead of destroying the ecosystem and local economy, these invasive plants can positively influence citizens' health.

AI Use Report: We did not use AI to support our modeling and writing.

Keywords: Invasive plants, Cellular Automata, Analytic Network Process, TOPSIS

Contents

1 Introduction

1.1 Background

Invasive species are serious problems in many places. An invasive species, also known as an alien species, is a non-native organism that becomes popular in a specific area. It is extremely harmful to the local ecosystem, economy, and human health. These species reproduce quickly and push aside native plants because of their lack of natural predators and controls in new habitats. It is very harmful to local biodiversity and can even cause the extinction of native species. Hence, human beings need to control invasive species and increase the biodiversity.

In this case, we need to have a deep understanding of the growth and the spread of invasive plants. The growth of invasive plants is closely related to temperature, rainfall, and sunlight. Only when we know how these factors contribute to the growth of invasive plants, can we find a perfect solution to analyze the impact of the invasion and improve the local ecosystem.

Among these invasive species, dandelion is a headache for many gardeners. Dandelion is known for its notable bright flowers and distinctive "puffball" seed head. With a parachute-like structure, dandelion can spread its seeds everywhere through wind. However, even though dandelion is native to Eurasia, this plant now can be found worldwide. It invades many different areas, which causes great harm to the local ecosystem. The dandelion has an amazing ability to reproduce and adapt. The dandelion can quickly cover the whole area if the condition is available. It will plunder other plants' living space and nutrition, leading to the disappearance of other native plants. Some research also finds out that the average distance dandelion seeds can travel is as far as 97 kilometers. As long as dandelions take root in the soil, it is very hard for people to get rid of them, for their roots can grow to 6 to 18 inches long.

1.2 Problem Restatement

For Problem 1, assuming a mature dandelion exists adjacently to open land, a model is required to predict the dispersion of dandelions in 1, 2, 3, 6, and 12 months, considering the impact of climates on the growth of dandelions. A simulation approach that can integrate various environmental conditions should be proposed to accurately predict the dispersion results.

For Problem 2, we should try to discuss the complicated relationship between dandelions, other flora, and human beings. It is often regarded as an invasive species because of its ability to thrive in different areas. However, every part of the dandelion is edible and can be used as medicine. This also applies to other invasive plants. Thus, we need to propose an impact factor, which should be tested on dandelions, using mathematical models to evaluate the impact of an invasive plant.

1.3 Our work

The workflow of our work is illustrated in Fig. 1.1. Variables required for modeling are first extracted from Google Earth Engine and are applied to the modeling of both problems. For the first problem, numerical formulations are constructed to model the behaviors of germination, growth, and wind dispersion of seeds related to climate, physics, and geographic variables, and are utilized in a proposed Cellular Automata algorithm to simulate the spreading of invasive plants, especially dandelions. For the second problem, an Analytic Network Process is built to model the relationship among various variables, which considers the effects of climate, physics, geography, biodiversity, and human activity, after which a model of impact factor is constructed using TOPSIS. For both problems, we analyze the results and discuss the sensitivity, advantages, and disadvantages of our models. Besides dandelions, the other two invasive plants selected for the second problem are **Centaurea solstitialis** and **Solidago canadensis**.

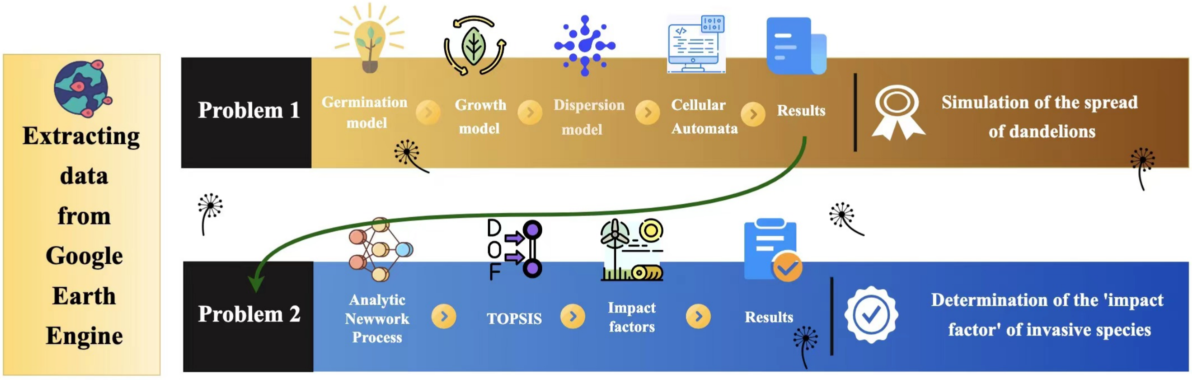

Figure 1.1: The workflow of our work.

2 Assumptions and Justifications

Assumption & Justification 1. The investigated open land is assumed to be perfectly flat. The simulation of complex and randomized geographical characteristics is beyond our scope.

Assumption & Justification 2. The investigated one-year period starts from 1st, January 2019 to 31st, December 2019 for locations in the northern hemisphere, and from 1st, July 2018 to 30th, June 2019 for locations in the southern hemisphere. The starting time of the requested one-year period for the problem may influence the results because of seasonality, thus should be restricted.

Assumption & Justification 3. Once plants become mature, their seeds finish spreading in one day. The flying time of seeds is relatively short compared to the life cycle of the entire plant, which can be ignored.

Assumption & Justification 4. When the precipitation is high enough (higher than 0.2 m/day), seed dispersion will be suspended and seeds will be destroyed. This assumption is important for modeling plants like dandelions that expose their seeds to the outside.

Assumption & Justification 5. Effects of influential variables on other plants considered in this context are the same as those on dandelions if they are not found in the literature. To the best of our knowledge and time, we have already found and applied influences of the temperature, pH, and heights of the two additionally considered plants. If more time and resources are available, the specific properties can be obtained from experiments.

3 Variables

Symbol	Description	Symbol	Description
μ	Mean	z_0	Roughness length
σ	Standard deviation	h	Displacement height
T	Temperature	P	The annual productional value of agriculture, forestry, and fishing of the country
t	The specific date	r_{prop}	The propagation rate
T_{best}	The best temperature for seeds to grow	S_{area}	Resolution of the environment data
R	The accumulated growth of a plant	S_{country}	The area of the country
r	The growth ratio value in the history	I_+	The set of all positive indicators
h	Specific humidity	I_-	The set of all negative indicators
s	Solar radiation	S^+	Positive ideal solution
D	Dispersal distance	S^-	Negative ideal solution
K	Kármán constant	IF	Impact factors
F	Terminal velocity	D^-	The distance of a data record to the negative ideal solution
H	Height of seed release	D^+	The distance of a data record to the positive ideal solution
W	Vertical wind speed		

4 Data acquisition

In this context, we will consider several influential variables related to climate, physics, vegetation, geography, and humans. Before we explain their effects on our modeling in the following sections, the data acquisition approach is first described in this section.

We obtain influential variables in the form of time-series data or scalar data through the Google Earth Engine Code Editor, a cloud computing platform for satellite imagery and geospatial datasets. The source and information of each variable can be seen in Table 4.1. Variables are from different public datasets and have different pixel resolutions. They are reprojected to a consistent 55.7 km resolution by aggregating and averaging the original pixels. For simplification and efficiency, we will simulate the daily behavior of plants, so the hourly variable zero plane displacement height is first averaged to daily data. We extract time-series data of these variables from 1st, January 2019 to 31st, December 2019 for locations in the northern hemisphere, and from 1st, July 2018 to 30th, June 2019 for locations in the southern hemisphere to get consistent climate periods. If some values are absent in time-series data, they are imputed by the average of presence values. If a scalar variable is invalid for a location, the value is imputed using the average value of other locations.

All 14 locations are considered in our results. Three selected locations are reported to have been invaded by dandelions. Similarly, for the additionally investigated plants, Centaurea solstitialis and Solidago canadensis, four and three locations are selected, respectively. These locations can be seen in Table 4.3. To investigate the influence of different climates, we select four more locations with four major climate types, i.e. tropical monsoon climate, temperate continental climate, temperate monsoon climate, and Mediterranean climate, respectively, as shown in Table 4.3.

Examples of the collected time-series data for three locations used in this section are shown in Fig. 4.2.

Table 4.1: Data sources and information of spatial-temporal influential variables obtained from Google Earth Engine.

Category	Variable	Dataset	Resolution (m)	Reduced Resolution (m)	Time scale	Unit
Climate	Soil surface temperature	ERA5-Land	11132	55659	Daily	Celsius
Climate	Air temperature (2 m above surface)	ERA5-Land	11132	55659	Daily	Celsius
Climate	Total precipitation	ERA5-Land	11132	55659	Daily	m
Climate	Eastward wind speed (10 m above surface)	ERA5-Land	11132	55659	Daily	m/s
Climate	Northward wind speed (10 m above surface)	ERA5-Land	11132	55659	Daily	m/s
Climate	Eastward wind speed (10 m above surface)	ERA5-Land	11132	55659	Hourly	m/s
Climate	Northward wind speed (10 m above surface)	ERA5-Land	11132	55659	Hourly	m/s
Climate	Total net solar radiation	ERA5-Land	11132	55659	Daily	J/m^2
Climate	Specific humidity	GLDAS-2.1	27830	55659	4-hourly	Mass fraction
Physics	Zero plane displacement height	MERRA-2	69375	55659	Hourly	m
Geography	Soil surface pH	OpenLandMap	250	55659	/	10x
Geography	Multi-Scale Topographic Position Index	ERGo	270	55659	/	/
Vegetation	Enhanced Vegetation Index	MODIS	463	55659	Daily	/
Vegetation	Normalized Difference Vegetation Index	MODIS	463	55659	Daily	/
Vegetation	Topographic diversity	ERGo	270	55659	/	/
Human	Percent of grass cover	CGLS	100	55659	Yearly	%
Human	Percent of shrub cover	CGLS	100	55659	Yearly	%
Human	Percent of cropland cover	CGLS	100	55659	Yearly	%
Human	Percent of urban cover	CGLS	100	55659	Yearly	%

5 Simulation of wind dispersion of plant seeds

In Section 5.1, we introduce influential variables considered in our model. In Section 5.2, we describe the source of the data used in our model. In Section 5.3, we build a model to consider the life cycle of plants whose seeds are dispersed by wind, which is divided into three separate stages: germination, growth, and seed dispersion considering wind. In Section 5.4, we perform Cellular Automata simulations under various conditions, analyze results, and draw quantified conclusions.

Table 4.2: Locations considered for three plants. Each of the locations has been invaded by one of these plants.

Country	City/State/Region	Longitude	Latitude	The reported invasive plant	Ref.
Chile	The Andes at central Chile	-69.71	-26.77	Taraxacum officinale	[4]
Japan	North Hokkaido	142.27	44.82	Taraxacum officinale	[8, 9]
United States	State of Alaska	-151.17	64.93	Taraxacum officinale	[10]
China	Shiyan	110.80	32.63	Solidago canadensis	[14]
China	Shaoxing	120.58	29.99	Solidago canadensis	[16]
Belarus	Homyel City	30.99	52.43	Solidago canadensis	[15]
China	Yili	81.31	43.90	Centaurea solstitialis	[11]
United States	State of Oregon	-122.68	45.51	Centaurea solstitialis	[12]
United States	State of California	-118.16	34.36	Centaurea solstitialis	[12]
United States	State of Ohio	-83.01	39.96	Centaurea solstitialis	[13]

Table 4.3: Locations considered to investigate different climates.

Country	City/State/Region	Longitude	Latitude	Climate
Saudi Arabia	Riyadh	46.66	24.70	Tropical monsoon
Germany	Berlin	13.15	52.61	Temperate continental
China	Beijing	116.80	40.13	Temperate monsoon
United States	San Francisco	-122.10	37.90	Mediterranean

5.1 Influential variables

To determine the model of the spread pattern of the dandelion, we discover the influential variables of its growing process and spreading process. In the growing process, temperature, soil pH, solar radiation, and humidity have a significant impact on the speed of this period. Also, the wind has an impact on the spread of the dandelion seeds.[1]

Temperature has an important impact on dandelion growth. Too high or too low temperatures weaken their photosynthesis, resulting in slower growth. In addition, high temperatures can also cause water evaporation and weaken enzyme activity, inhibiting the growth of dandelions. Therefore, there is an optimal temperature for the growth of dandelions, and dandelions grow fastest at this temperature. Above or below its optimum temperature will slow down its growth rate.[2]

Similar to many other plants, humidity and solar radiation influence the growth of the dandelion in a positive way, which gives it a favorable situation to grow. The higher the humidity and solar radiation, the faster the dandelion grows.

Soil pH is another factor influencing the growth speed of the dandelion. Soil with inappropriate pH will reduce the effective components of the soil. For example, in highly alkaline soil, the solubility and effectiveness of calcium and magnesium elements will decrease. Soil that is too acidic or too alkaline will inhibit the activity of soil microorganisms. Similar to the temperature, there is a most optimal pH, where the dandelion grows the fastest.[3]

Wind is the factor influencing the spread of the dandelion seeds. The wind speed and the direction of the wind determine the position the seed finally arrives, it consequently forms the final distribution of the dandelion.

Extreme weather is also considered. We assume that when the precipitation is high enough (higher than 0.2 m/day), seed dispersion will be suspended and seeds will be destroyed.

5.2 The model of growth

In this section, we formulate the germination, growth, and seed dispersion to perform cellular automata simulations.

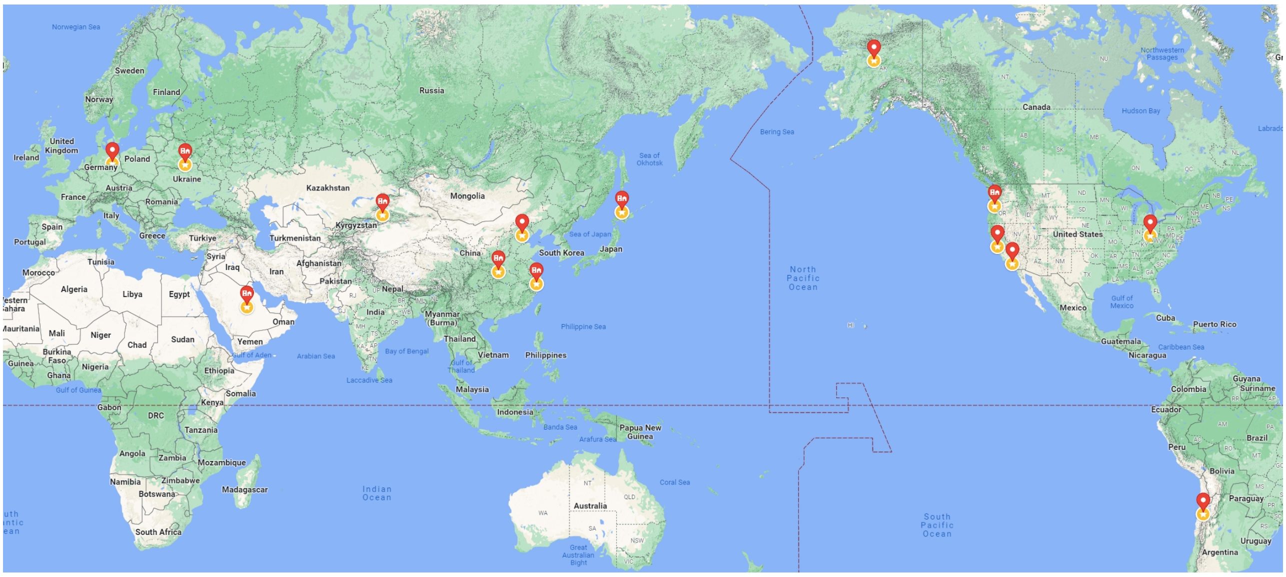

Figure 4.1: A map of all 14 locations considered, containing ten locations shown in Table 4.2 that one of the investigated plants has invaded and four locations shown in Table 4.3 occupied by four major climate types.

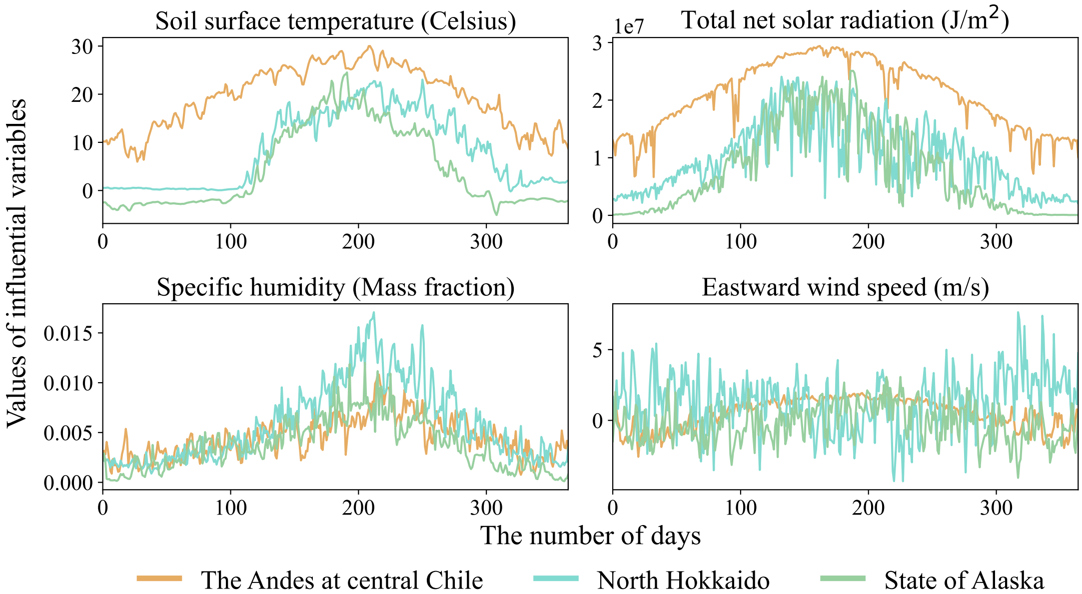

Figure 4.2: Examples of the collected time-series data for three locations.

5.2.1 Germination probability

Seeds may not grow up to a plant. Hoya et al. [17] examined the cumulative germination probability of tetraploid dandelions concerning time under different temperatures, as shown in Fig. 5.1. It can be concluded that less than 70% of seeds can germinate.

An empirical formula can be concluded according to results from Hoya et. al [17]. Inspired by the cumulative distribution function (CDF) of the normal distribution, and assuming that the mean μ and standard deviation σ of the normal distribution are functions of the temperature, the cumulative germination probability can be defined with the observed 70% limitation as

$$\Phi_g(t,T) = \frac{0.7}{\sigma(T)\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left\{-\frac{[u-\mu(T)]^2}{2\sigma^2(T)}\right\} \mathrm{d}u. \tag{5.1}$$

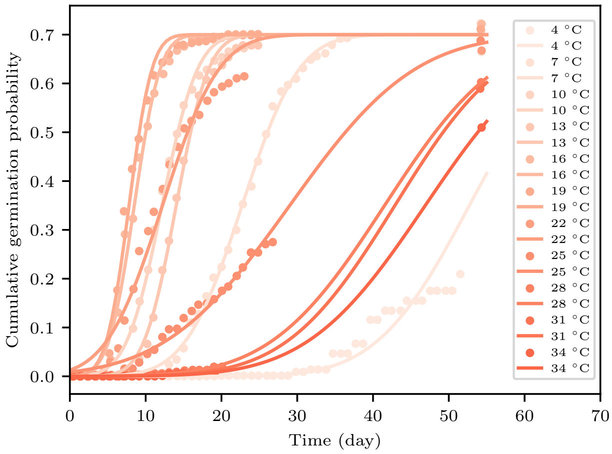

Figure 5.1: The cumulative germination probability data of tetraploid dandelions concerning time under different temperatures, and curves fitted using Eq. (5.1). Data points are obtained from Hoya et al. [17].

The formula is numerically evaluated by the `norm.cdf` function in the `scipy` Python package and can be fitted using curves of different temperatures through the `curve_fit` function in `scipy`, obtaining mean values and standard deviation values for each temperature; therefore, $\mu(T)$ and $\sigma(T)$ can be fitted by assuming that they are quadratic functions, resulting in

$$\left.\begin{aligned} \mu(T) &= 0.1698T^2 - 5.928T + 62.45 \\ \sigma(T) &= 0.02813T^2 - 0.8233T + 10.38 \end{aligned}\right\}, \tag{5.2}$$

as shown in Fig. 5.2.

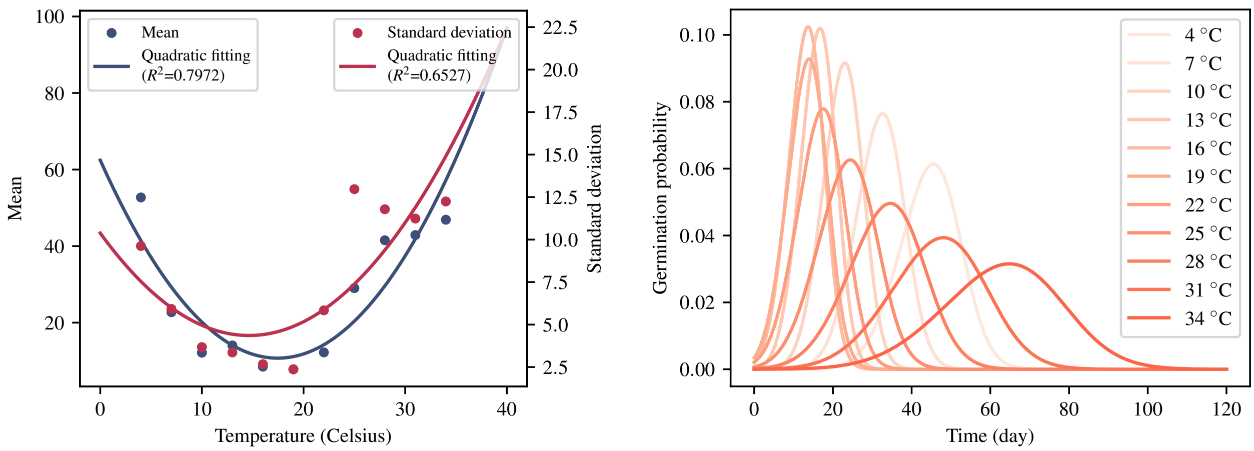

Figure 5.2: **Left**: Fitted mean μ and standard deviation σ values from Eq. (5.1) against temperature values T, and further quadratic fitting of $\mu(T)$ and $\sigma(T)$. **Right**: Germination probability concerning the index of the day, given by the fitted $\mu(T)$ and $\sigma(T)$ shown on the left and Eq. (5.1).

The cumulative probability Eq. (5.1) can also be expressed as

$$\Phi_g(t, T) = \Phi_g(t-1, T) + [1 - \Phi_g(t-1, T)]p_g(t, T), \tag{5.3}$$

where $p_g(t, T)$ is the probability that the seed, which has not germinated before the day t, germinates at the day t. The expression leads to

$$p_g(t, T) = \frac{\Phi_g(t, T) - \Phi_g(t-1, T)}{1 - \Phi_g(t-1, T)}, \tag{5.4}$$

which means that the germination probability of a seed can be directly given by the index of the day t and the temperature of the day by substituting Eq. (5.1) and Eq. (5.2) into the above equation, as shown in Fig. 5.2.

Making use of the above formulations, some useful conclusions can be drawn for our model. The fitted $\mu(T)$ is the day that the cumulative probability increases most rapidly under a certain temperature ($\partial\Phi_g(t,T)/\partial t$ reaches its maxima at $t = \mu(T)$); therefore, its minimum point is regarded as the most suitable temperature T_{best} for germination. According to Eq. (5.2), the most suitable temperature for germination is 17.45 °C. Under the most suitable temperature, using a Monte Carlo Simulation for 10,000 seeds, the average number of days needed for germination is 11.33.

A reminder is that, under the most suitable germination temperature, the maximum point of the probability for a seed to germinate $p_g(t, T = T_{\text{best}})$ does not coincide with (and is higher than) the maximum point of the increasing rate of the cumulative probability $\mathrm{d}\Phi_g(t, T = T_{\text{best}})/\mathrm{d}t$ because the number of seeds that have not germinated is decreasing day by day. We also stress that although Eq. (5.1) is modified from the form of the CDF of the normal distribution mainly because of its trend and its asymptotic line, it can be any other form such as a polynomial and $p_g(t,T)$ is not conceptually related to the probability distribution function of the normal distribution.

5.2.2 Growth

We start by defining the accumulated growth

$$R = \sum_{i=1}^{t} r(t), \tag{5.5}$$

which is the summation of growth ratio $r(t)$ values in the history. Starting from 0 on the germination day, the accumulated growth increases according to the growth ratio each day, which should represent the influence of influential variables. When R exceeds 1, the plant is regarded as mature and can spread seeds.

The relationships between influential variables and the growth ratio are modeled. Four main factors influencing the growth of the dandelions are considered: temperature T, specific humidity h, solar radiation s, and soil pH. They are first normalized by the following rule:

$$\overline{x} = \begin{cases} 0, x < x_{\min}, \\ \dfrac{x - x_{\min}}{x_{\max} - x_{\min}}, x_{\min} \leq x \leq x_{\max}, \\ 1, x > x_{\max}, \end{cases} \tag{5.6}$$

where x might be T, h, s, or pH. For $x_{\min}$ and $x_{\max}$, temperature T and pH may have certain ranges in which a certain plant can grow; however, we assume that plants can grow under all humidity and solar radiation conditions and only the growth rate varies, so $x_{\min}$ and $x_{\max}$ are defined by global minimum and maximum values for h and s, obtained from their data sources. Next, four growth parameters $\alpha_T(\overline{T})$, $\alpha_h(\overline{h})$, $\alpha_s(\overline{s})$, and $\alpha_{\text{pH}}(\overline{\text{pH}})$ are introduced for these variables respectively, all of which are between 0 and 1. According to the discussion of Section 5.1, the growth rate increases with the increase of humidity and solar radiation. The tanh function is chosen to model the effect for $\alpha_h(\overline{h})$ and $\alpha_s(\overline{s})$:

$$\alpha_{\text{monotonic}}(\overline{x}) = \tanh(3\overline{x}) = \frac{e^{3\overline{x}} - e^{-3\overline{x}}}{e^{3\overline{x}} + e^{-3\overline{x}}}, \quad \overline{x} \in [0,1], \tag{5.7}$$

where $\alpha_{\text{monotonic}}(\overline{x})$ might be replaced by $\alpha_h(\overline{h})$ or $\alpha_s(\overline{s})$. For temperature and pH, the growth rate increases with the increase of temperature and pH until the most optional situation (say $\overline{T}_{\text{best}}$ and $\overline{\text{pH}}_{\text{best}}$, respectively) and then decreases with the factors increasing. The tanh function is again used, but a platform-shaped function is constructed:

$$\alpha_{\text{platform}}(\overline{x}) = \begin{cases} \tanh\left(3\dfrac{\overline{x}}{\overline{x}_{\text{best}}}\right), & \overline{x} \in [0, \overline{x}_{\text{best}}) \\ -\tanh\left(3\dfrac{\overline{x}-1}{1-\overline{x}_{\text{best}}}\right), & \overline{x} \in [\overline{x}_{\text{best}}, 1] \end{cases} \tag{5.8}$$

where $\alpha_{\text{platform}}(\overline{x})$ might be replaced by $\alpha_T(\overline{T})$ or $\alpha_{\text{pH}}(\overline{\text{pH}})$. The function images of $\alpha_{\text{monotonic}}(\overline{x})$ and $\alpha_{\text{platform}}(\overline{x})$ can be shown in Fig. 5.3.

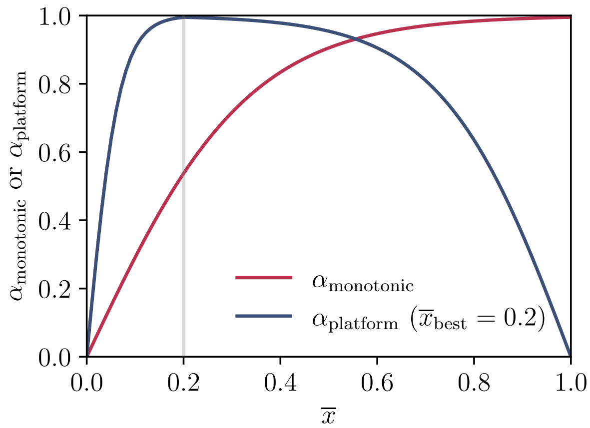

Figure 5.3: The functions of growth parameters given by Eq. (5.7) and Eq. (5.8). $\overline{x}$ is $\overline{h}$ or $\overline{s}$ for $\alpha_{\text{monotonic}}(\overline{x})$, and is $\overline{T}$ or $\overline{\text{pH}}$ for $\alpha_{\text{platform}}(\overline{x})$.

With the expressions given by Eq. (5.7)-(5.8), the four introduced parameters are multiplied together to get the growth index $\mathcal{A}(t)$ of a specific day t under the specific environment:

$$\mathcal{A}(t) = \alpha_T(\overline{T}(t)) \cdot \alpha_h(\overline{h}(t)) \cdot \alpha_s(\overline{s}(t)) \cdot \alpha_p(\overline{\text{pH}}(t)). \tag{5.9}$$

Finally, the growth ratio is modeled by

$$r(t) = \frac{1}{d_{\text{max}}}\left(\frac{d_{\text{max}}}{d_{\text{min}}}\right)^{\mathcal{A}(t)} + \mathcal{N}\left(0, \frac{1}{4}\left(\frac{1}{d_{\text{min}}} - \frac{1}{d_{\text{max}}}\right)\right), \tag{5.10}$$

where $d_{\text{min}} = 65$ is the lowest number of days needed for dandelions to be mature with the fastest growth rate while $d_{\text{max}} = 118$ [6, 7] is the highest one, and $\mathcal{N}$ represents a random variable drawn from the normal distribution, which is added to introduce randomness on growth. As shown in Fig. 5.4, the growth rate under the most suitable environment, which is $\alpha_h(\overline{h}) = 1$, $\alpha_s(\overline{s}) = 1$, $\alpha_T(\overline{T}) = \overline{T}_{\text{best}}$, and $\alpha_{\text{pH}}(\overline{\text{pH}}) = \overline{\text{pH}}_{\text{best}}$, approximately reaches 1, and the number of days required for the accumulated growth R to exceed 1 (i.e., mature) under the ideal environment is approximately d_{min} if no randomness is considered. Under a worse (or not such ideal) environment with a series of lower growth index $\mathcal{A}(t)$, the number of days needed will be higher than d_{min}, until $\mathcal{A}(t)$ approximates 0 so that the number of days needed approximates d_{max}.

A remark should be given here. Germination of seeds is considered only related to the temperature, different from growth. Although other influential factors are not considered in germination, they will still prevent growth if the environment is not suitable for the plant. Consequently, they will not further spread seeds. **Therefore, the calculated coverage rate only considers plants grown to a specific level, i.e.** $R > 0.1$**.**

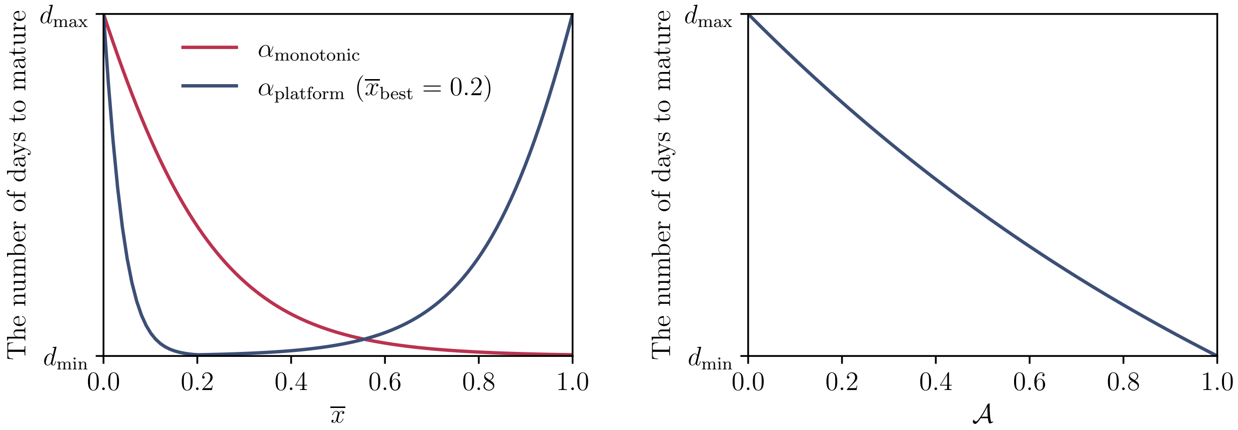

Figure 5.4: **Left**: The number of days needed to mature from germination without randomness, assuming that one influential variable is given, and others are fixed at their most suitable values for growth. $\overline{x}$ is $\overline{h}$ or $\overline{s}$ for $\alpha_{\text{monotonic}}(\overline{x})$, and is $\overline{T}$ or $\overline{\text{pH}}$ for $\alpha_{\text{platform}}(\overline{x})$. **Right**: The number of days needed to mature without randomness given a fixed growth index.

5.2.3 Wind dispersion

Wind effects on the dispersion of seeds are modeled in this section. When a plant grows up (the accumulated growth R from Eq. (5.5) exceeds 1), the plant finishes generating seeds that can be dispersed. If the type of plant, such as dandelions, spreads its seeds by wind, the wind direction, wind speed, and properties of seeds may affect the landing positions.

The wind speed and properties of seeds are first used to model the landing distance of a seed from its origin plant. A model for seed dispersal by wind is proposed by Nathan et al. [5] is employed in our modeling:

$$D = \frac{u_*}{K(F-W)}\left((H-h)\ln\left(\frac{H-h}{ez_0}\right)+z_0\right), \tag{5.11}$$

where D is the dispersion distance, u_* is the friction velocity estimated by

$$u_* = \frac{K \cdot U_{10}}{\ln \frac{10-h}{z_0}}, \tag{5.12}$$

in which U_{10} is the horizontal wind speed measured 10 m above the ground, $K \approx 0.40$ is the von Kármán constant, F is the terminal velocity, H is the height of seed release, W is the vertical wind speed (positive if upward), z_0 is the roughness length (the height at which the wind speed is theoretically equal to zero because of shear forces), and h is the displacement height, which scales the vertical distribution of those shear forces in the surface canopy.

The wind direction is considered to model the landing position. The daily wind direction average is defined as a radian angle obtained by analyzing the arctangent value of the ratio of two perpendicular velocity components. Additionally, since we obtain daily averaged values of influential variables, the uncertainty of wind direction in a day should also be counted. Therefore, we obtain the hourly wind direction data in a day in the same way and calculate the standard deviation of the hourly data for each day. Eventually, with the average and standard deviation calculated in advance, we obtain a normally distributed random daily wind direction for each spread seed.

Meanwhile, the height from which each seed starts spreading may differ, even if the seeds start from the same plant. This randomness is realized by selecting a random height between the general maximum and minimum of the height of a plant species, such as a dandelion.

5.2.4 Cellular Automata simulation

With the germination and growth properties of plants, and wind dispersion characteristics at hand, the behavior of the life cycle of a plant can be explicitly simulated and analyzed using Cellular Automata, a modeling technique that discretizes both time and space. Cells in regular grids follow the same rule to update their status according to themselves and their surroundings at each time step, building up an automatic and dynamic system that does not require global mathematical or physical formulations.

To simulate the growth and spreading of plants, a Cellular Automata algorithm is modeled. A schematic flow diagram is shown in Fig. 5.5 to address the process of the proposed algorithm. The investigated space is discretized as a square grid, where each cell in the grid contains multiple plants. The time step is one day, which is consistent with the timescale of our time-series datasets.

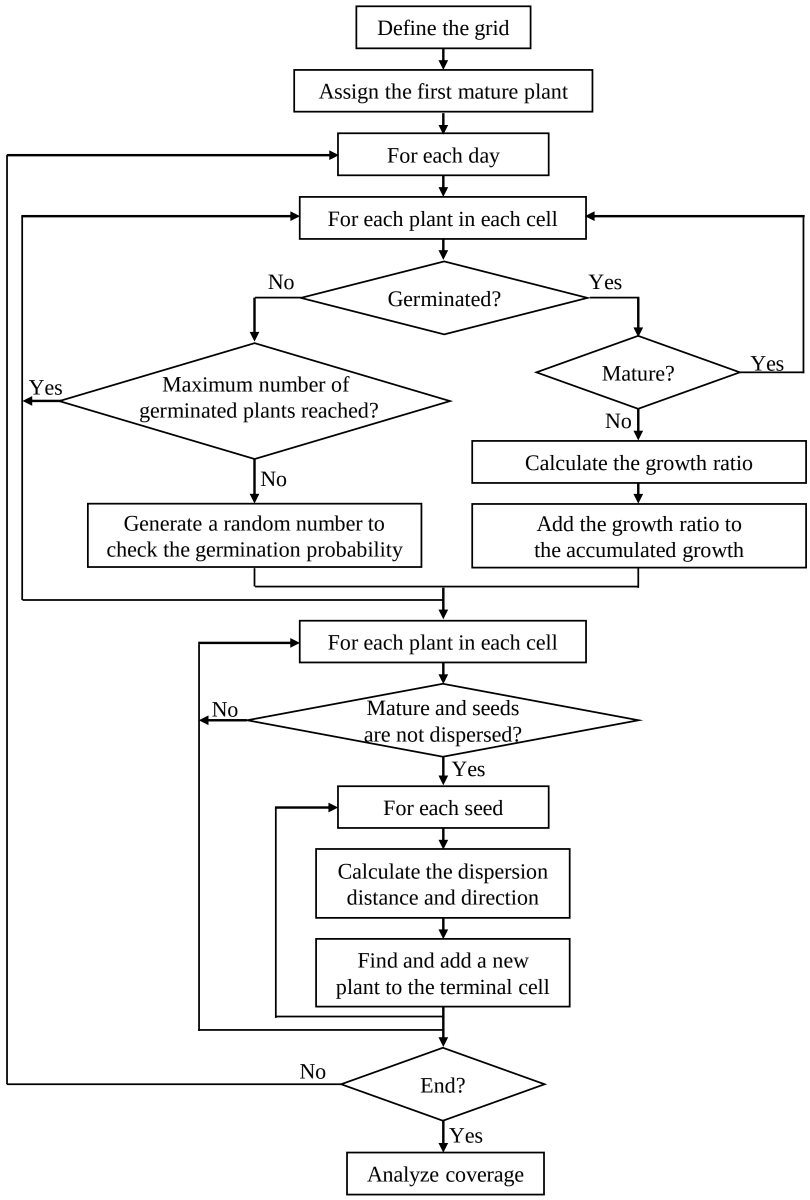

Figure 5.5: The flow diagram of the Cellular Automata simulation.

Step 1: At each time step, the germination of each seed in each cell is first simulated. A random number is drawn from the standard uniform distribution and is compared with the probability that the seed germinates on the current number of days from the day of planting (see Eq. (5.4)). We stress that the minimum planting distance is considered. Resources, such as nutrients, solar radiation, and space, are limited; therefore, given the area of the cell (which is set to 1 m^2), the maximum number of germinated plants is consequently limited to 100 plants per cell.

Step 2: For each germinated plant in each cell, the growth ratio of the plant on the day $r(t)$ is calculated by Eq. (5.10) and is added to its accumulated growth R if R has not exceeded 1 yet (i.e., the plant is not mature).

Step 3: After simulating germination and growth, the spreading of seeds is then simulated. For each seed of each mature plant ($R > 1$) in each cell, the wind speed and other properties are first used to estimate the spreading distance using Eq. (5.11). Note that several randomized variables are considered for plants or seeds. Then the randomized wind direction is sampled to determine the termination location of the seed. Finally, the plant will not generate seeds [18] or be counted in the growth calculation.

Step 4: After termination locations of all seeds in all cells are calculated and gathered, the number of new seeds in each cell is calculated, and these new seeds are planted and are waited for the germination check in Step 1 in the next time step.

At the end of each time step, the coverage rate is calculated by summing up all mature plants which is then divided by the maximum possible number of plants in the area, limited by the above-mentioned minimum planting distance. Two different definitions are used to calculate the coverage rate:
Area Related: The ratio of covered area that contains plants whose accumulated growth $R > 0.1$.
Number Related: The ratio of plants whose accumulated growth $R > 0.1$, to the total capacity of the entire area.

The simulation of the wind dispersion involves considerable randomness. The following Table 5.1 shows the variables that involve randomness and their distribution. In the table, the standard deviation of the terminal falling velocity is estimated via a special method: According to the previous research by Sun and Guo [19], the terminal falling velocity of the dandelion seed is proportional to its gravity to the power of 0.75. Therefore, given the weight range of regular dandelion seeds (which is 0.8-2.0 g per thousand seeds), the standard deviation of the terminal falling velocity can be estimated.

Table 5.1: Randomness involved in wind dispersion simulation.

Variable Name	Type of Randomness	Parameters
Number of spread seed	Uniform in a limited range	Number of seeds on each flower head (151-200),
	Uniform in a limited range	Number of flower heads (1-10)
Wind direction	Normal distribution	Average (daily wind direction),
	Normal distribution	Standard deviation (obtained from hourly wind direction)
Terminal falling velocity	Normal distribution	Average (0.39 [20]),
	Normal distribution	Standard deviation (*estimated via special method)

During the simulation of the growth part, the properties of different plants are described by the parameters listed in Table 5.2, which are later used to calculate the growth ratio increase of plants. The rest of the variables considered in the simulation are decided by the environment and do not involve randomness. They are included in the following Table 5.3.

Due to the randomness in our model, all simulations are repeatedly performed five times, and the average values are obtained for analysis.

5.3 Results and discussion

5.3.1 Simulation

The simulation domain is a 200 m× 200 m square, but note that **the calculation of coverage rates is performed in a semi-circle with the area 1 hectare, and the initial plant is on the circle center to meet the requirement of the problem.** Since the wind direction matters in our simulations, the effect of the angle of the semi-circle will be discussed below.

Table 5.2: Properties of plants used in the simulations.

Parameter	Taraxacum officinale	Solidago canadensis	Centaurea solstitialis
Minimum Temperature ($°C$)	0.0	-20.0	10.0
Maximum Temperature ($°C$)	30.0	40.0	30.0
Best Temperature ($°C$)	18.0	20.0	22.0
Minimum pH	4.0	4.8	5.5
Maximum pH	7.1	7.5	7.0
Best pH	8.0	6.2	6.5
Seed Spread Height Range (m)	0.05-0.45	0.30-0.90	0.10-1.00

Table 5.3: Nonrandom environmental factors

Variable Name	Related Stage	Detailed Influence
Precipitation (m per day)	Wind dispersion	All dispersion suspend when reaches 0.2
Zero-plane-displacement height (m)	Wind dispersion	See h in Eq. (5.11)
Roughness Length	Wind dispersion	Set to 0.03 [21]. See z_0 in Eq. (5.11)

We perform the simulation on all 14 locations selected for the convenience of sensitivity analysis, and the final results of Alaska, Chile, and Hokkaido are shown in Fig. 5.6. **The coverage rates during the year calculated in the 1-hectare semi-circle** under two different definitions mentioned above are illustrated in Fig. 5.7. Among the three selected locations, **Chile suffers from dandelions most, with more than 40% (and sometimes over 60%) of the land of the 1-hectare domain covered by dandelions.**

The number of dandelions increased dramatically during the last 3 months in Chile because the dispersion might follow an exponential rule. However, the first impact on the coverage rate is later than 100 days. We stress that the growth period (excluding germination time) of dandelions is approximately 65 to 118 days. Considering that the average number of days required is around 11.3, as shown in Section 5.2.1, and since the coverage rate is calculated for plants that have grown to $R > 0.1$ (Section 5.2.2), the first impact on the coverage rate will be later than 100 days.

The increase in coverage rates shows clear seasonality in all cases. The number of days required for dandelions to be mature might cluster around a certain value that depends on the environmental factors (i.e. the growth ratio $\mathcal{A}$) and are not the same because of the randomness in Eq. (5.10). Therefore, the increase of the coverage is generally staircase-shaped but smooth. We can also observe that, in Chile, the period between each dispersion break out is shorter than that in the other two locations, suggesting a more suitable germination environment and a higher growth rate for dandelions, which might be the main reason for the invasion report in Chile. Hopefully, this seasonality may give us some inspiration in preventing the deterioration of dandelion invasions, such as taking measures before a coming increase trend.

Moreover, in Fig. 5.7, we can observe a clear inclination of the dispersion. This may probably be attributed to the prevailing wind direction and other reasons. Similarly, such property of wind dispersion can also provide us with inspiration related to invasion prevention.

5.3.2 Sensitivity analysis

The relative location of the investigated 1-hectare land that the initial plant is adjacent to should be discussed. The spreading of seeds heavily depends on the wind direction of the day and is partially determined by the wind direction uncertainty. Starting from the angle shown in Fig. 5.6, the semi-circle rotates anti-clock-wise and the coverage rates are calculated. As shown in Fig. 5.8 where we select one of the repeated simulations to study the relative location, the difference in coverage is significant for the evaluation of seed dispersion and plant invasion.

To validate the robustness and accuracy of our model, **we simulate the dispersion of three plants**

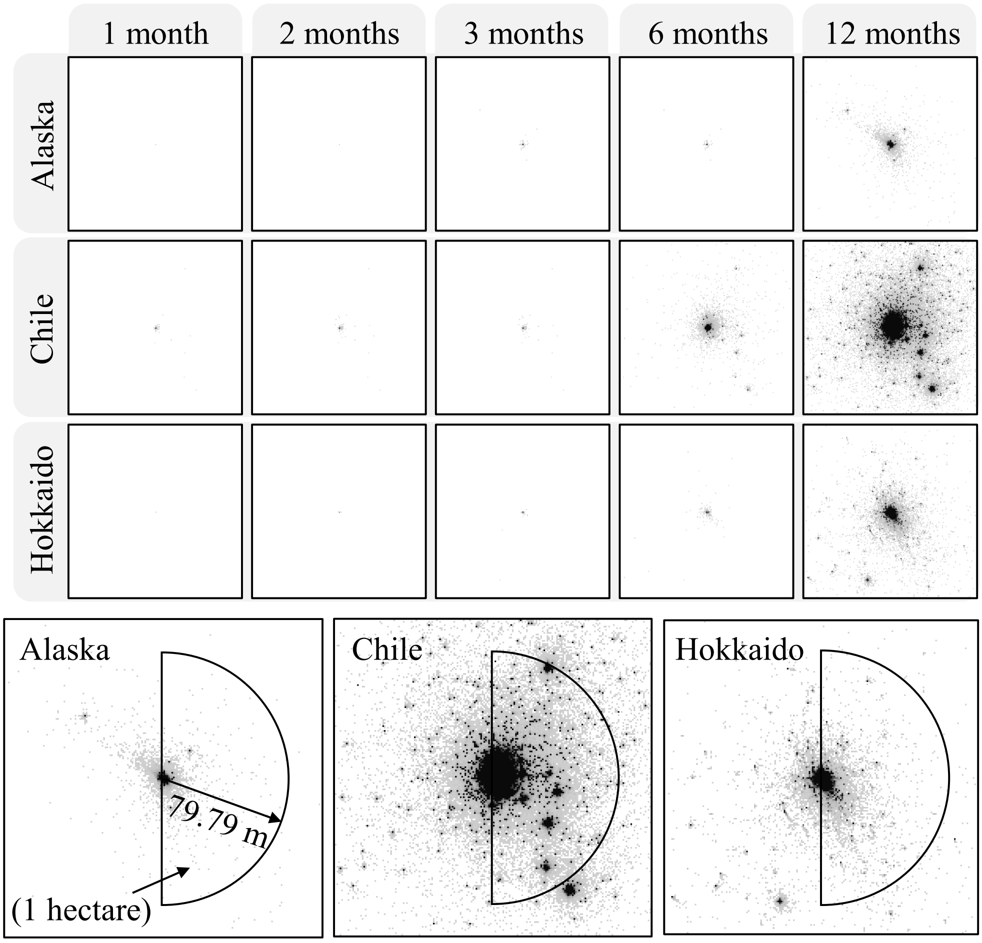

Figure 5.6: **Top**: Dispersion of dandelions at 1, 2, 3, 6, and 12 months at different locations. **Bottom left**: The 360th day of Alaska. **Bottom middle**: The 360th day of Chile. **Bottom right**: The 360th day of Hokkaido. The area of the semicircle is 1 hectare, in which the coverage rates are calculated.

in 14 locations as shown in Table 4.2 and 4.3. Therefore, the sensitivity analysis can be directly performed by analyzing the relationship between influential variables (annual average values of each location) and the coverage rate. As shown in Fig. 5.9, all four influential variables are effective in our model and have a positive correlation with the coverage rate, especially temperature and soil pH, to whom the model is most sensitive. One of the reasons for the high correlation of temperature is that germination is also related to the temperature while not related to others.

5.3.3 Advantages and disadvantages

The advantages of our model are: (1) The entire life cycle, including germination, growth, and seed dispersion, is formulated. (2) The distribution of plants can be explicitly simulated using the proposed Cellular Automata algorithm. (3) The Cellular Automata algorithm can be (and indeed is) implemented in a parallel way, which significantly reduces computational time. (4) The results can accurately reveal the influences of climate, geography, and properties of plants on the coverage of invasive plants, which are robustly verified by sensitivity analysis.

Disadvantages can be concluded as (1) Although the effect of temperature on germination is calibrated from experimental results, the formulations still contain empirical functions, which can all be determined by sufficient experiments for each of the investigated plants. (2) The computational time is still considerable although in parallel because repeated simulations should be done under all considered situations for the model with uncertainty.

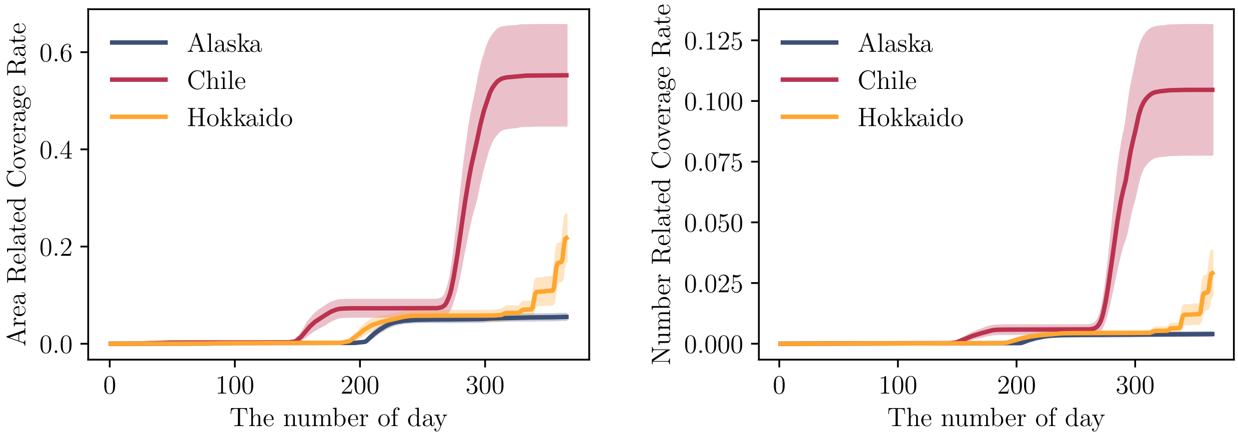

Figure 5.7: Coverage rate in different locations. Shadows represent the biased standard deviation of five repeated simulations.

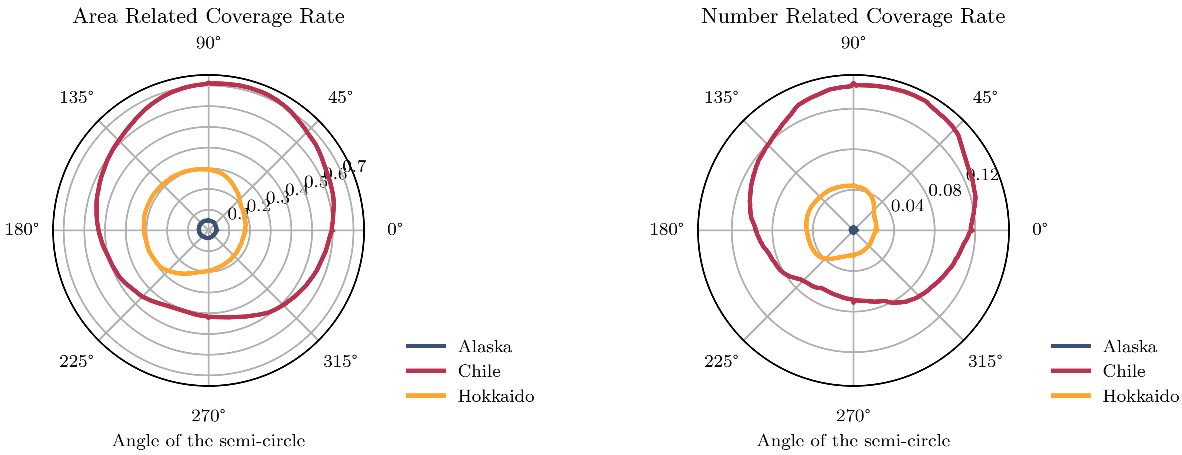

Figure 5.8: The effect of the angle of the investigated semi-circle. The results are from one of the repeated simulations.

6 Modeling the impact of invasive species

In this part, we choose two other invasive plants, the Centaurea solstitialis and Solidago canadensis. In our work, we mainly discuss their influences in the fourteen places (same as before). Another sensitivity analysis is conducted to confirm our model of the 'impact factor' of invasive species.

6.1 Invasive Species Indicators (ISI)

To comprehensively demonstrate the impact of invasive species on the ecological environment and human society while considering the data accessibility, after multiple rounds of detailed discussions, an indicator system named the Invasive Species Indicators (ISI) is established, where relevant parameters describing the spreading trend and the social impact are included. Generally, ISI consists of 7 indicator clusters, with a total number of 23 indicators to measure the impact of invasive species.

The indicator cluster of growth environment is influential in the spread of invasive species, where the growth coefficients, the Multi-Scale Topographic Position Index (mTPI) describing the terrain slope (geographical barriers), and the similarity of environmental conditions to the origin are scored subjectively based on the environmental data acquired from Google Earth Engine.

Due to their correlations to biodiversity and the ecological resistance to invasive species, ecological conditions including the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) which reflect the vegetation coverage are considered, besides a subjective

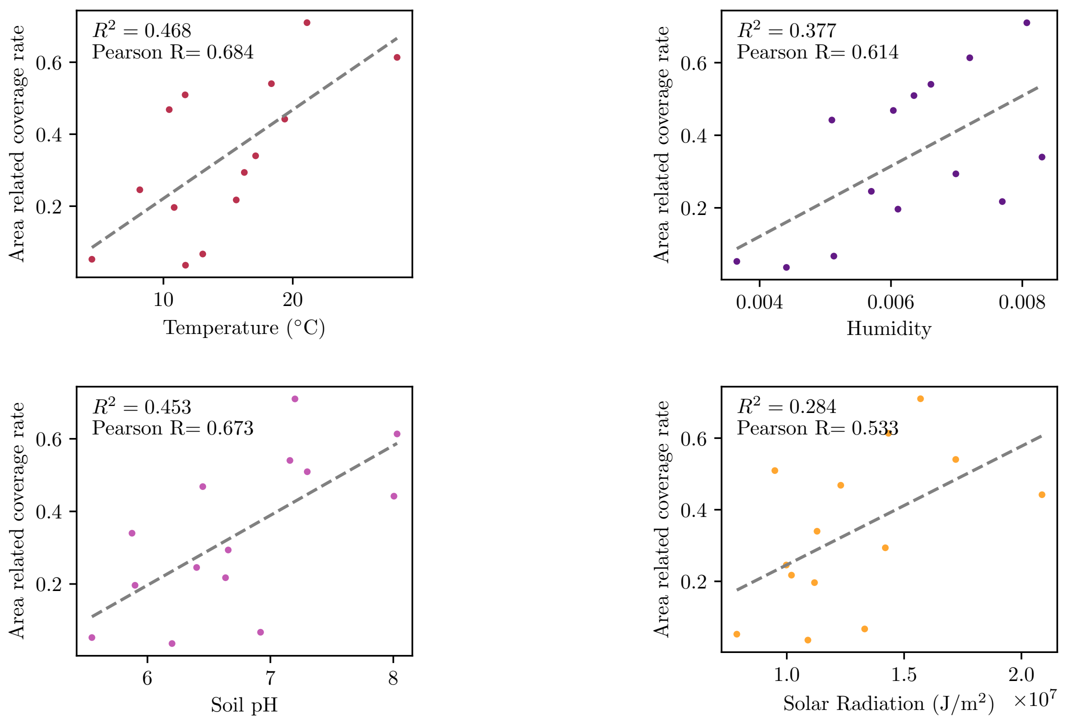

Figure 5.9: The relation between the coverage rate and four influential variables.

score of ecosystem incompatibility, which is decided based on extensive literature review [27].

Since a full quantitative consideration of the potential impact on biodiversity is impossible to realize due to the limited time and resources, we use the method of evaluating the current biodiversity vulnerability of each region to substitute the process. Here the topographic diversity indicates the ability of the habitat to support high plant diversity [24], the percentage of threatened plants [22], the loss of vegetated land [23], and the percentage of protected areas [25] are presented to represent the pressure and invasion-proof capability of the land and are accessed from the OECD database.

To quantify the risk caused to agricultural production, a similar measure is taken, with the percentage of cropland coverage and the percentage of forest coverage as indicators of productive vulnerability.

Also, the coverage rate results are obtained from simulations of 14 locations in problem 1 to quantify the propagation of the plant. The coverage rate is obtained from the entire simulation domain (200 m$\times$ 200 m) instead of a 1-hectare semi-circle to eliminate the impact of different wind directions on the spread of invasive species.

Apart from the above, social effects should also be taken into account. To quantify the effect of the edibility and medicinal availability of invasive plants, the proportion of positive literature is accessed. We formulate two search query templates for PubMed as

Query 1: <BOTANICAL NAME> AND (medicine OR food) NOT disease NOT risk;

Query 2: <BOTANICAL NAME> AND (disease OR risk) NOT medicine NOT food.

The number of retrieved articles n_{pos} through Query 1 is regarded as the number of positive literature, while the number of articles n_{neg} through Query 2 is the number of negative literature. Accessed on 6th November 2023, $n_{\text{pos}} = 194$ and $n_{\text{neg}} = 39$ are for Taraxacum officinale (dandelion), $n_{\text{pos}} = 40$ and $n_{\text{neg}} = 6$ are for Solidago canadensis, and $n_{\text{pos}} = 21$ and $n_{\text{neg}} = 14$ are for Centaurea solstitialis. The proportion of positive literature is calculated by $n_{\text{pos}}/(n_{\text{pos}} + n_{\text{neg}})$ for each of the plants. Indeed, the retrieved articles using Query 1 from PubMed may also contain negative articles and vice versa, which might be room for improvement in the modeling.

Besides these, the equivalent first-year agricultural impact is designed as

$$E = \frac{P \cdot r_{\text{prop}} \cdot S_{\text{area}}}{S_{\text{country}}}, \tag{6.1}$$

where P is the annual productional value of agriculture, forestry, and fishing of the country, r_{prop} is the propagation rate, S_{area} is the resolution of the environmental data (55.659 km^2), and S_{country} is the area of the country. The subjective score of economic dependence on the ecological environment (which is decided based on the estimated proportion of ecological-related economic output value) is also introduced here to evaluate the importance of environmental quality. Another fact that affects people's opinion on the invasive plant is the ornamental value of the plant, which is completely subjectively decided based on our group discussion.

Finally, existing social conditions of the regions are considered as well, including the coverage of urban areas and the Human Development Index, for they may partly support the intensity and willingness the control invasive species driven by the country's government.

The full list of our ISI system is shown in Table 6.1, where the abbreviations that will be used below are illustrated. Here, the indicators with positive effects mean that they can reduce the negative impact of invasive plants, and those with negative effects will make the condition worse.

Table 6.1: Invasive Species Indicators

Category	Variable	Abbreviation	Effect	ANP weight	Data source
Growth environment	α_s (Sunshine growth coefficient)	Gro1	-	0.0374	/
	α_h (Humidity growth coefficient)	Gro2	-	0.0381	/
	α_{pH} (pH growth coefficient)	Gro3	-	0.0615	/
	α_T (Temperature growth coefficient)	Gro4	-	0.0423	/
	mTPI (Multi-Scale Topographic Position Index)	Gro5	+	0.0168	ERGo
	Similarity of environmental conditions to origin	Gro6	-	0.0242	Subjective
Ecological conditions	NDVI (Normalized Difference Vegetation Index)	Eco1	+	0.0138	Google Earth Engine
	EVI (Enhanced Vegetation Index)	Eco2	+	0.0238	Google Earth Engine
	Ecosystem incompatibility	Eco3	+	0.1122	Investigation-based subjective decision
Biodiversity vulnerability	Topographic Diversity	Bio1	+	0.0264	EGRo
	Percentage of threatened plants	Bio2	-	0.0222	OECD database
	Loss of natural and semi-natural vegetated land since 2004	Bio3	-	0.0693	OECD database
	Percentage of protected Areas	Bio4	+	0.0183	OECD database
Productive vulnerability	Percentage of cropland cover	Prod1	-	0.1144	GEE
	Percentage of forest cover	Prod2	-	0.0268	World Bank database
Propagation	Number related coverage rate in 12 months	Prop1	-	0.0613	CA modeling results
	Number related coverage rate in 6 months	Prop2	-	0.0225	CA modeling results
Social impact	Percentage of positive literature	Soc1	+	0.0475	PubMed
	Relative first-year Agricultural & forestry Impact	Soc2	-	0.0448	World Bank (Processed)
	Ornamental value	Soc3	+	0.0193	Subjective decision
	Economic dependence on ecological environment	Soc4	-	0.0276	Investigation-based subjective decision
Existing social conditions	HDI (Human Development Index)	Ext1	+	0.0155	World Bank
	Percentage of urban coverage	Ext2	+	0.1127	World Bank

6.2 Modeling of the impact factor

6.2.1 Weighting of metrics using Analytic Network Process (ANP)

To better manually control the weight relationship between various indicators according to prior knowledge, the ANP was used to conduct the indicator weighting process due to its universality and flexibility, as well as the tolerance for non-independence and categorical complexity of indicators, which is why we choose ANP over the Analytic Hierarchy Process (AHP), whose results may vary due to the relationship between indicators.

Based on the connections of indicators (as shown in Fig. 6.1) and the prioritization of each connection as well as the indicator categories (as shown in Table 6.1), computational results of the ANP model can be generated by the *Super Decisions* (SD) software through an interactive procedure that requires us to do pairwise importance comparison of indicators subjectively. Then the software automatically analyzes the subjectively given relative importance among categories and indicators to obtain weights of each indicator without further operating assistance.

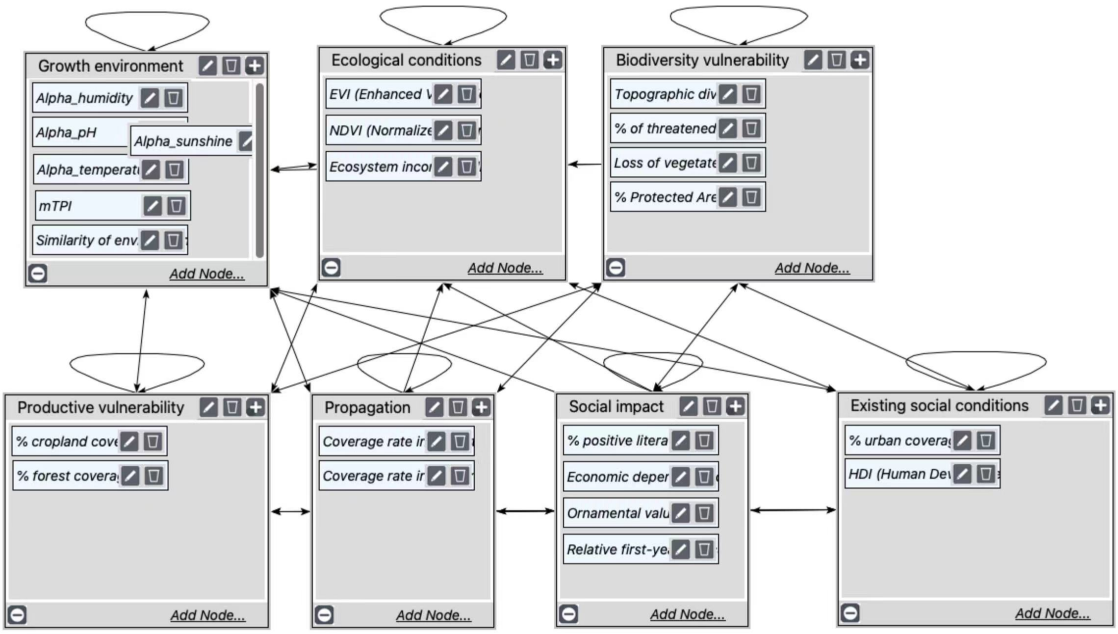

Figure 6.1: The connections of indicators used in the ANP modeling in the Super Decisions software.

6.2.2 Integration of indicators

The commonly used Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) method is taken here to conduct the overall integration of values under multiple indicators. Below is the process.

Step 1: Decision matrix construction.

After the data collection process, an evaluation matrix $(x_{ij})_{m\times n}$ of m alternatives (the selected places) and n indicators (ISIs) is constructed, where x_{ij} is the value of the i-th alternative under the j-th indicator.

Step 2: Data normalization.

To better quantify and integrate data under various indicators, a data normalization process is executed. Using the Min-Max normalization with homotropization (transforming positive factors into negative ones in this case):

$$\overline{x}_{ij} = \begin{cases} \dfrac{\max_i(x_{ij}) - x_{ij}}{\max_i(x_{ij}) - \min_i(x_{ij})}, \ j \in I_+ \\ \dfrac{x_{ij} - \min_i(x_{ij})}{\max_i(x_{ij}) - \min_i(x_{ij})}, \ j \in I_- \end{cases} \tag{6.2}$$

a normalized data matrix $\overline{x} = (\overline{x}_{ij})_{m\times n}$ can be obtained. I_+ is the set of all positive indicators, and I_- is the set of all negative indicators, here $I_+ \bigcup I_- = I$, where I is the set of all the ISI.

Step 3: Weighting the normalized decision matrix.

$$\hat{x}_{ij} = \overline{x}_{ij} \cdot w_j, \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n \tag{6.3}$$

where w_j is the weight of the ISIs obtained from ANP.

Step 4: Determination of ideal solutions.

The positive ideal solution S_+ and the negative ideal solution S_- can be determined as:

$$\begin{aligned} S^+ &= (\max_i s_{i1}, \max_i s_{i2}, \ldots, \max_i s_{in}), \ i \in I_+, \\ S^- &= (\min_i s_{i1}, \min_i s_{i2}, \ldots, \min_i s_{in}), i \in I_-, \end{aligned} \tag{6.4}$$

whose components are denoted by s_j^+ and s_j^-, respectively.
Step 5: Calculation of the distance between each of the alternatives (data points) and the positive ideal solutions.

$$D_i^+ = \sqrt{\sum_{j=1}^{n}(s_{ij} - s_j^+)^2}, \quad i = 1, 2, \ldots, m \tag{6.5}$$

Step 6: Calculation of the distances between the alternatives and the negative ideal solutions.

$$D_i^- = \sqrt{\sum_{j=1}^{n}(s_{ij} - s_j^-)^2}, \quad i = 1, 2, \ldots, m \tag{6.6}$$

where D_i^+ and D_i^- are L^2-norm distances from the alternative i to the positive and negative ideal solutions.
Step 7: Final Scoring.

$$f_i = \frac{D_i^-}{D_i^- + D_i^+} \tag{6.7}$$

here $0 \leq f_i \leq 1$.
Step 8: Ranking of the alternatives.

The ranking of alternatives can be acquired according to f_i $(i = 1, 2, \ldots, m)$.

6.2.3 Determining the impact factor

The more data points we have, the more information we can provide to the judgment of ANP and the key calculation of ideal solutions. With the convenient powerful data extraction method from Google Earth Engine and other abundant datasets as well as the simulation, we are able to build up a dataset with complete ISI records and fourteen locations. Each of the locations is used to perform simulations for three plants, summing up to a dataset with 42 data points.

Note that each of the investigated plants is reported to have invaded 3 or 4 of these locations, which is not used during the modeling so that can be an evaluation of the accuracy and effectiveness of the modeling approach.

In our work, after normalization, **we will split the 42 data points into the training set used to obtain TOSIS parameters, and the testing test used to validate the performance of the model to predict the impact factor**. The training set contains 14 records of dandelion at 14 locations. The testing set contains 28 records of Centaurea solstitialis and Solidago canadensis at 14 locations, respectively.

The final impact factor of each data point that has all ISIs recorded or simulated, either new data points or data points in the dataset, can be evaluated under the TOPSIS framework:

$$\text{Impact Factor (IF)} = \frac{D^-}{D^- + D^+}, \tag{6.8}$$

where D^- and D^+ are the distance of a data record to the negative ideal solution and the positive ideal solution obtained from the training set.

6.3 Results and discussion

6.3.1 Results of weights from ANP

Weights and ranks of each ISI from ANP analysis are shown in Table 6.1 and illustrated in Fig. 6.2. The results can partially reflect our tendency when judging invasive plants and their impact.

The percentage of cropland cover (Prod1) reflects the current status of agricultural development, which is directly and primarily relevant to the extent of destruction caused by invasive species, thus its weight when considering invasion is the highest among all the 23 indicators. Meanwhile, the percentage of urban coverage (Ext2) shows the urbanization degree of the investigated location. If the location is highly urbanized, the invasive plant will be immediately observed, and proper actions will soon be taken by city or plant management departments, besides providing the spread of invasive species with greater geographic barriers. The ecosystem incompatibility (Eco3) aims to evaluate the potential ecological limitations on the spread and living of the invasive species, and its importance is significant in both perspectives determining the impact of invasive species, which prompts us to elevate its unique importance despite being a subjective indicator.

Although individual indicators from the growth environment category do not rank high, the sum of all these indicators occupies over 20 percent of the total contribution. The growth ratio affects the dispersion period, and further influences the final coverage.

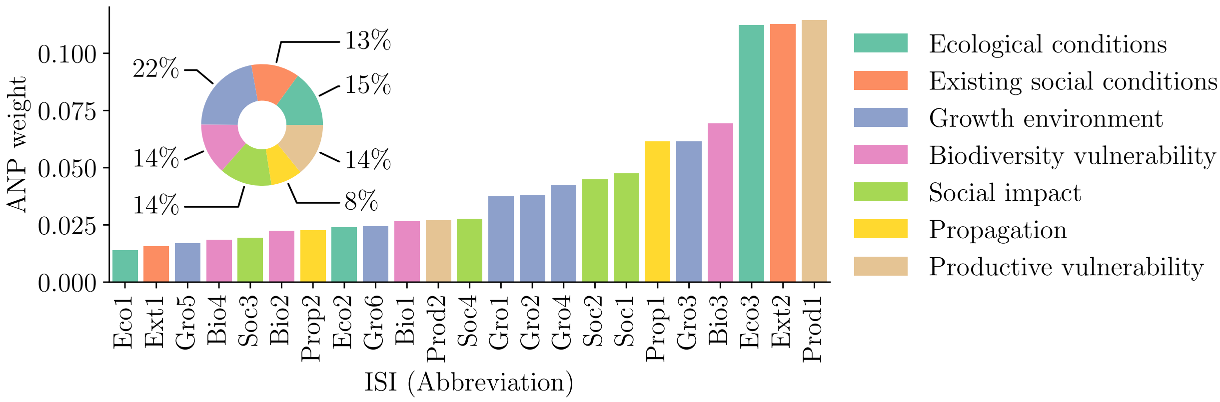

Figure 6.2: Weights of ISIs from ANP and proportion of each category.

6.3.2 Results of the Impact Factor

Using the training set (data points of dandelions) and weights from ISIs, the TOPSIS is performed, and the normalized positive and negative ideal solutions can be obtained for further usage of the model. The parameters obtained from TOPSIS including the ideal positive solution and the ideal negative solution are applied for inference on the testing set. Results of impact factors are shown in Fig. 6.3. For dandelions, TOPSIS accurately identifies two out of three locations that have reported the invasion of dandelions, based on our carefully selected objective and subjective indicators and well-established routine for weights of indicators. The predicted Impact Factor for the other two plants in all locations also highlights the locations that have been invaded, suggesting the good generalization ability of our model and the Impact Factor.

From the results of the three species above, it can be seen that in our established impact factor evaluation system, 0.5 can be considered a perfect benchmark for determining whether a species should be considered an invasive species in a particular region. This result has significant potential implications in the field of invasive species identification and prevention. As for the accuracy and universality of this standard, we weren't able to obtain large-scale instance validation during the short period of work, which will be carried out in more detailedly in our future work based on deeper investigation.

6.3.3 Sensitivity analysis

To evaluate the sensitivity of the TOPSIS model, we perform the partial dependency analysis. For each indicator, while fixing the values of others, the value of all data points of this indicator changes from its lowest (0 after MinMax normalization) to its highest value (1 after normalization) in a fixed step, and the evaluated impact factors are obtained during the process to build a partial relationship between the average impact factor with the indicator. The results for all ISIs are shown in Fig. 6.4. The built model is sensitive to changes in ecological conditions, existing social conditions,

Figure 6.3: Results of Impact Factors for three investigated plants.

and productive vulnerability, mostly because of Eco3, Ext2, and Prod1, which is the reflection of our weights shown in Fig. 6.2.

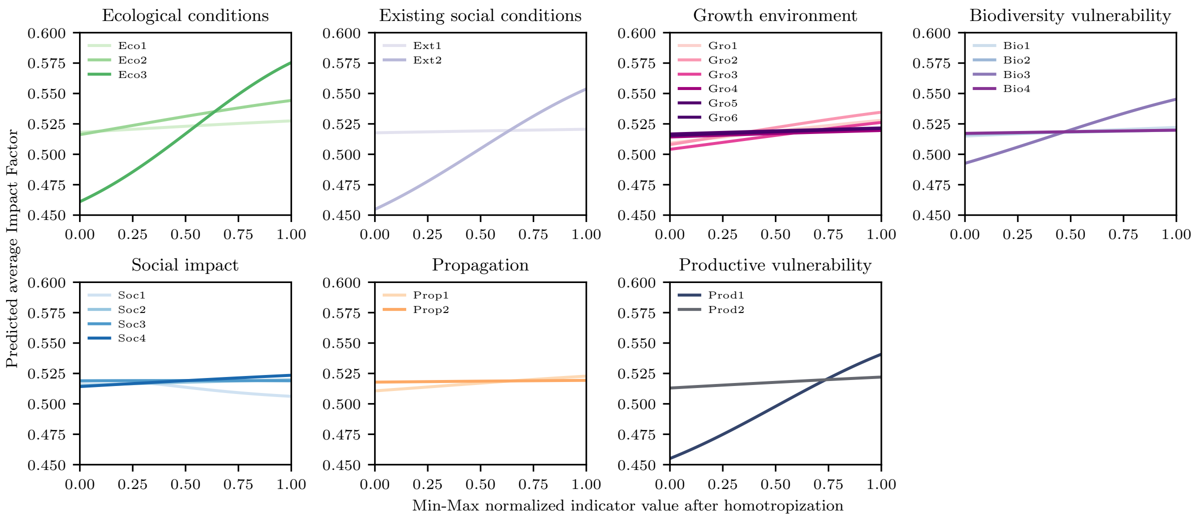

Figure 6.4: Partial dependency plot.

6.3.4 Advantages and disadvantages

In modeling the impact factor of invasive species, we use ANP to quantify subjective judgments and TOPSIS to construct the final model. Due to the sufficiency and objectiveness of the indicator system (ISI), the model has achieved great overall stability and certainty, while maintaining the final results of the same species in different regions to be of a certain degree of differentiation. The impact factor (the TOPSIS score) of a potential invasive species has a critical value of 0.5, which perfectly matches the midpoint of the ideal solutions.

Nevertheless, due to limitations in time and resources, our analysis of the potential economic and social impacts of invasive species did not use simulation predictions under actual models but used estimates based on the current situation instead. In addition, the scoring of subjective evaluation indicators may also be avoided by setting quantitative methods respectively, and hiring experts for centralized discussions may help.

7 Conclusion

In our study, we build a model to explore the spread of dandelion and two other invasive plants. We made some assumptions at the first of our study. We consider different factors influencing germination, growth, and the spread of the dandelion. For this part, we mainly discuss the influence of temperature, sunlight, and rainfall. By building a Cellular Automata model, we can visualize the spreading of dandelions in a given period. Then to decide the contradictory influence the invasive plants have on local areas, we make up a set of evaluating factors. Through normalizing the data and building the weighting metrics, we can express the invasive ability and impact of a single plant by using an impact factor. Our impact factor built from ANP and TOPSIS can accurately evaluate the invasion risk of a location concerning the plant. We use two other plants that spread seeds through the wind: Centaurea solstitialis and Solidago canadensis in our simulations and 14 worldwide locations to analyze the sensitivity, validate the effectiveness of the model, and provide solid data source of indicators for the impact factor evaluation.

By building the spreading model and evaluating factors, we can know the spreading pattern and invasive ability of different plants. In this case, we can have a deeper understanding of the invasive plants and make a better method to prevent them from destroying the ecosystem and economics of the local area. What's more, we can also make good use of these plants, and contribute to the increase of living conditions of local citizens.

Limited by time and computing resources, we simplify influential factors and the simulation. The results will be more promising with sufficient resources. Some influential factors are modeled using empirical functions, which can be further determined through detailed experiments.

References

[1] P. R. MEZYNSKI and D. F. COLE: "Germination of Dandelion Seed on a Thermogradient Plate" Weed Science, Volume 22, Issue 5, September (1974):506-507.

[2] Hoya, A., Shibaike, H., Morita, T. et al. Germination characteristics of native Japanese dandelion autopolyploids and their putative diploid parent species. J Plant Res 120, 139–147 (2007).

[3] https://hort.extension.wisc.edu/articles/dandelion-taraxacum-officinale/

[4] Molina-Montenegro, Marco A., et al. "Higher plasticity in ecophysiological traits enhances the performance and invasion success of Taraxacum officinale (dandelion) in alpine environments." Biological Invasions 14 (2012): 21-33.

[5] Nathan, Ran, Uriel N. Safriel, and Imanuel Noy-Meir. "Field validation and sensitivity analysis of a mechanistic model for tree seed dispersal by wind." Ecology 82.2 (2001): 374-388.

[6] https://www.westcoastseeds.com/blogs/wcs-academy/how-to-grow-dandelions

[7] https://www.johnnyseeds.com/growers-library/herbs/dandelion/dandelion-key-growing-information.html

[8] https://www.nies.go.jp/biodiversity/invasive/DB/detail/80640e.html

[9] http://intl.ce.cn/gjzh/200605/24/t20060524_7074354.shtml

[10] https://home.nps.gov/kefj/blogs/fighting-the-insidious-invasion.htm

[11] Yonggang, Chen, Shiming, Wu, et al. Forest Pest and Disease 3 (2017): 36-37. (in Chinese)

[12] Julia, Roxana, David W. Holland, and Joseph Guenthner. "Assessing the economic impact of invasive species: the case of yellow starthistle (Centaurea solsitialis L.) in the rangelands of Idaho, USA." Journal of Environmental Management 85.4 (2007): 876-882.

[13] McIver, James, Robbin Thorp, and Karen Erickson. "Pollinators of the invasive plant, yellow starthistle (Centaurea solstitialis), in north-eastern Oregon, USA." Weed biology and management 9.2 (2009): 137-145.

[14] Renkai, Yan, et al., Hubei Plant Protection 200 (2023): 6-7. (in Chinese)

[15] Gusev, A. P. "The impact of invasive Canadian goldenrod (Solidago canadensis L.) on regenerative succession in old fields (the Southeast of Belarus)." Russian Journal of Biological Invasions 6 (2015): 74-77.

[16] Wang, Congyan, et al. "Canada goldenrod invasion cause significant shifts in the taxonomic diversity and community stability of plant communities in heterogeneous landscapes in urban ecosystems in East China." Ecological Engineering 127 (2019): 504-509.

[17] Hoya, Akihiko, et al. "Germination characteristics of native Japanese dandelion autopolyploids and their putative diploid parent species." Journal of Plant Research 120 (2007): 139-147.

[18] https://zhidao.baidu.com/question/568520260144584444.html (in Chinese)

[19] Sun, B.-H.; Guo, X.-L. Aerodynamic Shape and Drag Scaling Law of a Flexible Fibre in a Flowing Medium. Theoretical and Applied Mechanics Letters 2023, 13, 100397, doi:10.1016/j.taml.2022.100397.

[20] Iyer, V., Gaensbauer, H., Daniel, T.L. et al. Wind dispersal of battery-free wireless devices. Nature 603, 427–433 (2022). https://doi.org/10.1038/s41586-021-04363-9

[21] WMO Guide to Meteorological Instruments and Methods of Observation WMO-No. 8 page I.5-13

[22] https://data.oecd.org/biodiver/threatened-species.htm

[23] https://data.oecd.org/biodiver/land-cover-change.htm

[24] https://developers.google.com/earth-engine/datasets/catalog/CSP_ERGo_1_0_Global_SRTM_topoDiversity#description

[25] https://data.oecd.org/biodiver/protected-areas.htm#indicator-chart

[26] https://databank.worldbank.org/metadataglossary/africa-development-indicators/series/UNDP.HDI.XD

[27] Maheshwari, Shamoni, and Daniel A. Barbash. "The genetics of hybrid incompatibilities." Annual review of genetics 45 (2011): 331-355.