Summary

The dandelion, officially named *Taraxacum officinale*, is known as an invasive species, which is likely to cause economic or environmental harm or harm to human health.

In this paper, we design a **Dandelion Spread Model** by using the Monte Carlo method.

Provided there is one dandelion at the edge of an open one-hectare plot of land, we need to predict the spread of dandelions over time.

We input 3 groups of random variables to our model, i.e., temperature, wind and humidity. Some details like equivalent temperature curve, dispersal distances under horizontal wind and updrafts, and random biological characteristics/behaviors are also taken into account.

We output all the locations of dandelions spreading on the land over time. In order to get stable results, we run the model with 2 phases. For each phase we run 1000 times. In the first phase, we determine optimal starting time. In the second phase, we fix the starting time and get the most probable outcome.

We apply our model to 6 locations and conclude the statistic characteristics. We also conduct a sensitivity analysis and discuss the strengths and weaknesses.

In this paper, we also devise a **Plant Impact Factor Model** to measure the impact of invasive plants on humans and the environment. We define a global impact factor and a local impact factor for different purposes.

We consider many attributes, including the plant's characteristics, its effects on humans and the environment, and its local adaptation, and use the Analytic Hierarchy Process to get the weights for each attribute.

We apply our model to 226 invasive plants and calculate their global impact factors. We also choose 4 plants and calculate their local impact factors at 3 locations. Likewise, we conduct a sensitivity analysis and discuss the strengths and weaknesses.

At the end of this paper, we write an article to elaborate on our viewpoint regarding invasive species. Rather than think they are enemies, we need to learn how to get along with them. To us human beings, they are just vigorous friends.

Contents

1 Introduction

1.1 Background

The dandelion is a very common plant with scientific name *Taraxacum officinale*, which is stemless, lactiferous and perennial[17]. We can find them at roadsides, in wild fields or at inconspicuous locations. Just imagine that these little herbs with beautiful yellow flowers have been evolving on this planet for about 30 million years[5] - even longer than we human beings!

But currently they are labeled as an invasive plant. **Invasive species** is a relative concept. It means a non-native species whose introduction does or is likely to cause economic or environmental harm or harm to human health[7]. In some regions, dandelions do have some traits that make them invasive. They can adapt to different environments. They have high growth rate and spread fast. They are likely to cuddle together and crowd out desirable plants, particularly in lawns. Also, they have deep taproots that enable them to pull available water out of the soil. It is also mentioned that dandelions could cause significant economic damage and infestation of other crops worldwide[17].

To disclose the reason that dandelions become invasive, researchers have done many studies on how the environment affects dandelions. Laboratory and field studies have discussed the impact of climatic factors on its growth. Laboratory studies have been conducted on the relationship between germination rate and ambiances, such as light and temperature[10]. Field studies have also been conducted on the same topic, providing credibility for the theoretical conclusions[24]. Studies have provided insights on the drought tolerance of dandelions[2]. In the meanwhile, soil pollution and geographic factors are also being considered[23].

Researchers have also paid great attention to dandelion dispersal, which is a major way for dandelion propagation. Past papers have analyzed the structure of dandelion seeds and the effects of humidity on their structure, which affect the dispersal speed[8]. Scientists have modeled the aerodynamics of the seeds and demonstrated the effects of wind speed and directions. It has also been proved that updrafts could not be neglected[18].

Beside this particular species, researchers have discussed invasive species more generally. Ecologists have used species native to the invaded range as "control species" in attempting to determine the traits associated with invasive plant species[12]. A protocol has been developed for categorizing non-native plants according to their negative impacts on biodiversity in a large area such as a state, nation, or ecological region[15]. Some online tools like PIMDAT[14] have been provided to help people make strategic decisions about invasive species control projects. Currently we also have Best Management Practices to prevent the introduction and spread of invasive species[1].

Public databases play an important role in the research of invasive species, especially for plants. By conferring to some database, we can find plant data to study biological attributes of the invasive species[19]. Another database for tracking introduced species provides a baseline for effective modeling of species trends and interactions, geospatially and temporally[20]. US Government's Open Data provides a lot of data for current invasive plants[3].

1.2 Problem Restatement

In this paper, we will solve 2 problems.

For the first problem, we will predict the spread of dandelions within a year (360 days). Suppose that a dandelion in its dispersal period happens to be adjacent to an open one-hectare plot of land. We need to create a mathematical model to describe how the dandelion will spread its seeds and how its offspring will scatter on the land. And we will have to consider the effects of various climatic conditions, such as temperate, arid, and tropical climates, on dandelion growth.

For the second problem, we will need to devise an "impact factor" for invasive species. We will formulate a mathematical model to determine this factor and apply it to dandelions and some other invasive plants. We also need to make sure the plants are invasive in the regions we choose.

2 Dandelion Spread Model

2.1 Assumptions and Justifications

1. **The initial dandelion is at the midpoint of the left edge of a one-hectare square plot of land. Seeds blown out of the area are neglected.**
   Justification: According to the problem, the initial dandelion is adjacent to an open one-hectare plot of land. In order to acquire maximum spread effect, we set the plant in the middle of the left edge and make the right side our spread area. The seeds drifting to the left of the edge will be neglected. We may imagine a river or a forest is at the left side, so seeds cannot spread back to our spread area.

2. **The field is planar.**
   Justification: We only consider the case of flat terrain. This helps us limit the spread to two dimensions. Flat terrain simplifies the calculation of dispersal distance since the wind can be at a constant height and the seed landing will not be affected by different topography.

3. **There are no other plants in the spread area.**
   Justification: Considering other plants would add complicity to our model, for example, we need to collect data of those plants and study the competitive interactions between plants. So, simplification is necessary.

   More specifically, we assume that no other plants hinder the spread of dandelions, which means

   a. Dandelion seeds are not intercepted by other plants;

   b. Dandelions do not have to compete with other plants.

4. **Wind direction is uniformly distributed from 0° to 360°.**
   Justification: Wind speed and direction can be affected by different topography, especially near the ground. At the height of dandelions, there could be vortices and updrafts that would greatly affect dispersal. Though we can obtain wind data from meteorological stations, the data may not be accurate for a small area. Therefore, we decided not to use the real-time wind direction data.

Moreover, though wind rose maps provide the probabilities for each direction, we cannot decide the relative wind direction to the area without knowing the edge orientation of our spread area. So we assume wind direction is uniformly distributed from 0° to 360°. In this way, all directions have same odds and we can pin the dandelion on an arbitrary edge.

5. **Dandelion number always increases.**
   Justification: More specifically, we assume that there are none of the followings:

   a. Human intervention;

   b. Natural enemies;

   c. Extreme weather conditions.

   This assumption simplifies our model. Without considering above conditions, it will be an carefree place for dandelion growth. And if we notice that the dandelion is a perennial herb, it is safe to assume the number will always increase.

2.2   Model Overview

We will use a Monte Carlo simulation model to solve the problem.

In reality, we find that sometimes dandelions grow in clusters, and at other times they are very sparse. Actually, the spread process is affected by many random factors, including climate conditions, wind speed and direction, topographical features and so on. In order to eliminate this randomness, we need a method to find statistical patterns from repeating the process many times. Candidates are neural networks, deep learning, stochastic processes, etc. Among them, the Monte Carlo method becomes an option due to its simplicity. Its underlying concept is to introduce randomness to solve problems. It relies on repeated random sampling to obtain numerical results. We will simulate dandelion spread process by going through the process many times, observing the results from a statistical perspective and drawing some conclusions.

We consider five environmental factors as input parameters, as shown in Table 1.

Table 1 Environmental factors

Symbol	Description	Unit
μ_T	Mean temperature	°C
σ_T	Standard deviation of temperature	°C
μ_W	Mean wind speed	m/s
σ_W	Standard deviation of wind speed	m/s
μ_H	Mean humidity	%

And we will do the following steps to consider the environmental influence on the dandelion spread process:

- Construct a temperature curve in Section 2.3;

- Model dispersal mechanisms in Section 2.4;

- Introduce an adaptation factor k in Section 2.5.

Then our model will simulate the spread of dandelions by tracking every single dandelion according to a life cycle pattern, which we will model in Section 2.6.

We will choose 6 locations in the US. For each location, we have one set of 5 input environmental factors. We will run our model in 2 phases. In the first phase, we will run 1000 times to get the optimal starting date for each location. In the second phase, we will fix the starting date we get from the first phase and run another 1000 times to get the final outcome.

We choose to run 1000 times because we find that the data with 1000 runs will exhibit an approximately normal distribution with a small margin of error for the mean.

As the output of our model, we get all the dandelion locations for each day within 360 days.

In order to depict the statistical characteristics of dandelion spread, we conclude all the locations into 2 variables – one is the **number** of dandelions and the other is the **mean distance** of dandelions from the initial one. The first variable reflects spread intensity and the second one reflects spread range.

2.3   Temperature Curve

We assume that the temperature follows a sinusoidal curve, as shown in Fig.1. We only consider locations in the northern hemisphere, so we further assume that the lowest temperature is reached on Jan.1st, when time $t = 0$. We observe that the period of the curve is 360 (for convenience, we set each month to 30 days). Thus, we write the equation as

$$T = -A \cos \frac{2\pi}{360} t + B.$$
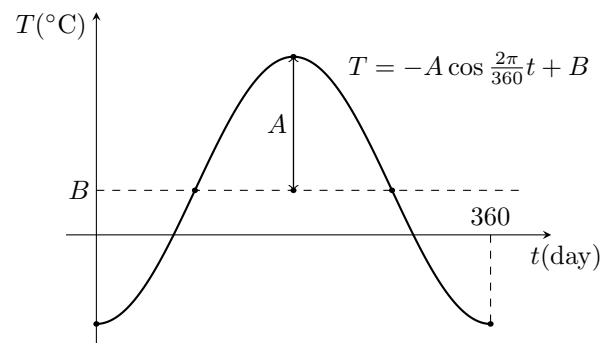
Fig. 1 Temperature curve

The following conditions should be satisfied:

$$\mu_T = \frac{1}{360} \int_0^{360} \left(-A \cos \frac{2\pi}{360} t + B\right) \mathrm{d}t,$$

$$\sigma_T^2 = \frac{1}{360} \int_0^{360} \left[\left(-A \cos \frac{2\pi}{360} t + B\right) - \mu_T\right]^2 \mathrm{d}t.$$

Solving for A and B, we get

$$T = -\sqrt{2}\sigma_T \cos \frac{2\pi}{360} t + \mu_T. \tag{1}$$

2.4   Dispersal Mechanisms

For the **direction** of seed dispersal, we assume that the seed always follows the direction of the wind, which is uniformly distributed from 0° to 360° according to Assumption 4.

For the **distance** of seed dispersal, we consider two types of wind, horizontal winds and up-drafts[18] (see Fig.2). The seed travels either a normal distance s_N due to horizontal winds or a long distance s_L due to updrafts. We set dispersal distance as the greater one between s_N and s_L.

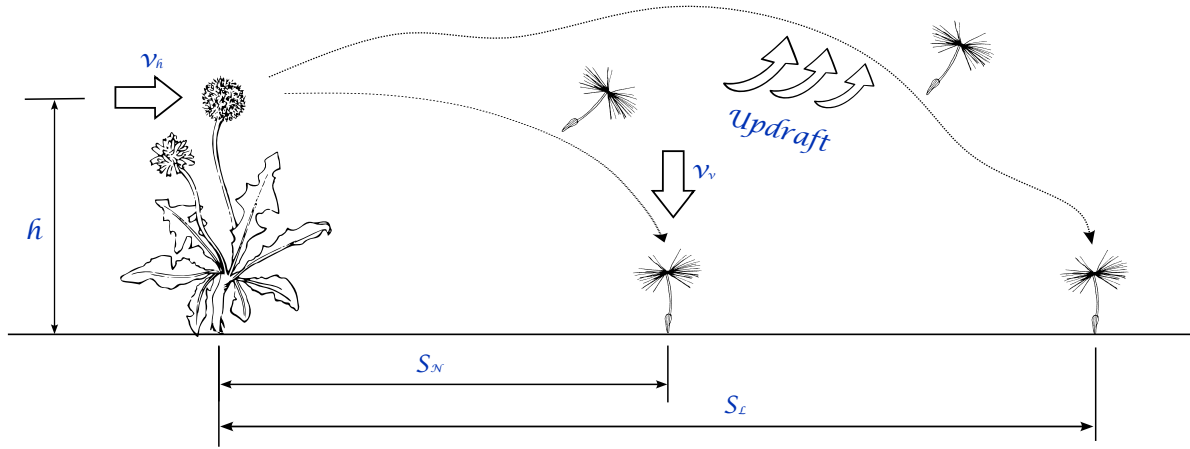**Fig. 2** Dandelion dispersal distance

2.4.1 Normal Distance s_N

In the case of horizontal wind, due to large air drag on the seed's pappus, we assume the seed moves at a constant velocity in a statistical sense[18]. we have

$$\frac{s_N}{v_h} = \frac{h}{v_v}, \tag{2}$$

where:

   h is the height of the puffball, which is evaluated in Section 2.6;

   v_v is the vertical component of the velocity of the seed, which is a constant value, 0.4 m/s[18].

   v_h is the speed of the horizontal wind, which follows a normal distribution whose mean equals μ_W and standard deviation equals σ_W;

2.4.2 Long Distance s_L

In the case of updraft, we suppose that s_L follows an exponential distribution[13] with cumulative distribution function $\mathrm{P}(s_L < s) = 1 - ae^{-bs}$. We know $\mathrm{P}(s_L < 0) = 0$. We also know that the probability of a seed being blown further than 10 meters is 0.5%[18]. Solving for a and b, we get

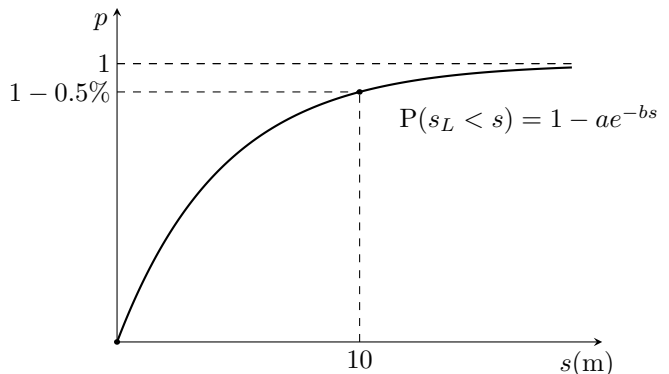$$\mathrm{P}(s_L < s) = 1 - e^{-0.53s}. \tag{3}$$

Fig. 3 Cumulative distribution function of long-distance dispersal

2.4.3 Dandelion Density

The leafs of dandelions have a radius of around 15cm. If a seed lands without falling on the leaf of another dandelion, the dandelion density should be less than $(1/0.15)^2 \approx 45$ plants/m². If a seed arrives at a location whose vicinity has a density greater than this maximum value, the seed fails to land, which means it is not counted in our model.

2.5 Adaptation Factor

We introduce an adaptation factor k which is evaluated using temperature T and mean humidity μ_H. Dandelions prefer temperate climates and relatively high humidity. We divide temperature into three ranks, which indicate if the temperature is suitable for the dandelion. We also divide humidity into three ranks, which indicate if the humidity is suitable for the dandelion. On a given day, we calculate k by computing the arithmetic mean of the rank for temperature and humidity and scaling k so that $0 \leq k \leq 1$.

We use k to determine the mean value of some parameters of dandelions, such as the height of the puffball. To allow for some variation, we let the parameters follow normal distributions whose

$$\mu_{\text{parameter}} = \left(\text{worst} + \frac{\text{best} - \text{worst}}{4} \right) + k \frac{\text{best} - \text{worst}}{2},$$
$$\sigma_{\text{parameter}} = \frac{\text{best} - \text{worst}}{12}. \tag{4}$$

The worst and best values for relevant parameters will be given in Section 2.6. If the parameter exceeds the limits, it is set to the limit.

2.6 Dandelion Life Cycle

A dandelion can transfer between 6 stages: **seed**, **developing**, **dispersal**, **inter-dispersal**, **hold**, and **dormancy**.
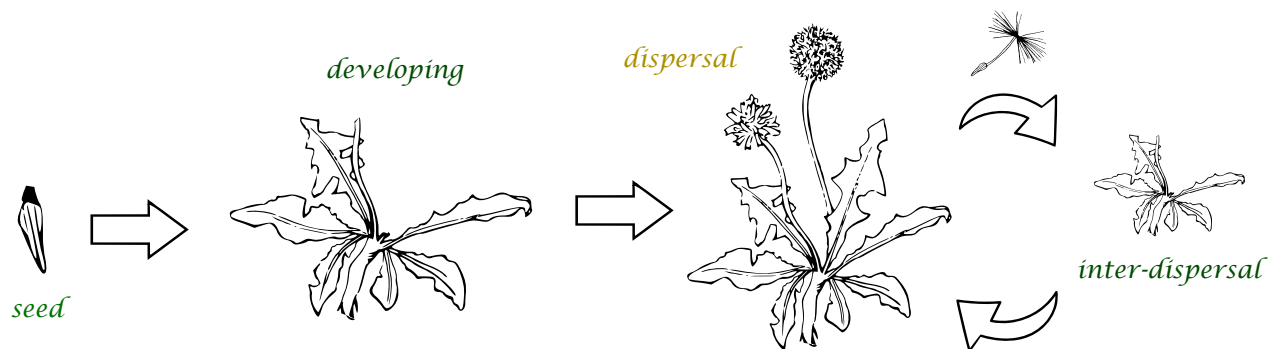

Fig. 4 Dandelion life cycle

- Out of all **seed**s, only 2% germinate successfully[9]. These seeds spend 3 days growing into the developing stage[17].

- The **developing** stage is the time from sprouting to seed ripening. It takes a sprout 60–95 days to grow into a flower, and additional 9–12 days for the flower to ripen[4]. The exact duration of this stage is determined by the adaptation factor k (worst = 107 days, best = 69 days) per Section 2.5.

- The **dispersal** stage lasts 10 days. A dandelion plant can produce around 10 seed heads per year and it usually flowers twice, so it produces 5 heads every time it flowers. The number of seeds on each head is determined by k (worst = 150, best = 200)[11]. It takes each head 2 days to disperse its seeds. The height of the puffball, h, is determined by k (worst = 12cm, best = 40cm)[6]. The seeds are dispersed per the method stated in Section 2.4.

- The **inter-dispersal** stage is the stage between two dispersal stages of a plant. In this stage, dandelions regrow their scapes and flowers. Since most dandelions that flowered in spring usually flower again in fall[17], the duration of this stage is set to be half a year, which is 180 days.

We also consider two additional stages, **hold** and **dormancy**.

- When the temperature T is greater than 24°C, only 12.5% of the flowers bloom and disperse their seeds[25]. We let dandelions enter dispersal by probability 12.5% and move the rest into the **hold** stage. In this stage, flowers are held in buds. As soon as T falls below 24°C, flowers can bloom and disperse seeds by 100%.

- Dandelions enter the **dormancy** stage when the temperature T falls below 0°C. When the temperature is low, it is likely to snow and all life activities of dandelions are paused. As soon as T rises above 0°C, dandelions resume its previous stage.

2.7 Algorithm

Monte Carlo simulation algorithm for dandelion spread

INPUT: run times n and environment factors μ_T, σ_T, μ_W, σ_W and μ_H

---

for 1 to n **do**
    for all t in days of one year (360) **do**
        $T \leftarrow$ evaluate current temperature per Eq.1
        $k \leftarrow$ evaluate adaptation factor per Section 2.5
        for *dandelion* in current all dandelions **do**
            if $T < 0$ **then** *previous stage* $\leftarrow$ *current stage*, *current stage* $\leftarrow$ **Dormancy**
            if *current stage* is **Dormancy** and $T > 0$ **then** *current stage* $\leftarrow$ *previous stage*
            if *current stage* is **Hold** and $T < 24$ **then** *current stage* $\leftarrow$ *previous stage*
            if *current stage* is **Developing** or *current stage* is **Inter-dispersal then**
                if days of *current stage* reaches the next stage **then**
                    if $T > 24$ **then**
                        current stage $\leftarrow$ **Dispersal** with probability 0.125
                      if Dispersal succeeds **then**
                        current stage $\leftarrow$ **Dispersal**
                      else
                        previous stage $\leftarrow$ *current stage*

$\qquad$ *current stage* ← **Hold**
$\qquad$ **end if**
$\qquad$ **else**
$\qquad$ *current stage* ← **Dispersal**
$\qquad$ **end if**
$\qquad$ **end if**
$\qquad$ **end if**
if *current stage* is **Seed then**
$\qquad$ **if** days of *current stage* reaches the next stage **then** *current stage* ← **Developing**
end if
if *current stage* is **Dispersal then**
$\qquad$ determine the seed count that needs to be dispersed per Section 2.6
$\qquad$ **for all** *seed* in dispersing seeds **do**
$\qquad$ calculate dispersal distance per Section 2.4
$\qquad$ calculate land location and add it to *locations*
$\qquad$ generate a *new dandelion* with stage **Seed**
$\qquad$ **end for**
$\qquad$ **if** days of *current stage* reaches the next stage **then** *current stage* ← **Inter-dispersal**
$\qquad$ **end if**
$\qquad$ **end for**
$\qquad$ **end for**
end for
calculate *numbers* and *mean distances* from *locations*
Return *locations* , *numbers* and *mean distances*

OUTPUT: dandelion spread *locations*, *numbers* and *mean distances* over days

---

2.8   Dandelion Spread Results

2.8.1   Locations

We choose 6 locations in the US so that we have a variety of temperature and wind conditions. AK is cold; CA, DC, and KS are warm; and FL and HI are relatively hot. AK, DC, and KS have distinct seasons; CA and FL show moderate temperature fluctuations; and HI has very stable temperature. The wind conditions also exhibit different patterns. DC has notably large wind speed. The values of the five environmental factors are listed in Table 4.

Table 4 Environmental factors for six locations[21]

State	Abbreviation	μ_T (°C)	σ_T (°C)	μ_W (m/s)	σ_W (m/s)	μ_H (%)
Alaska	AK	-0.05	9.08	7.27	1.42	81.46
California	CA	16.20	4.99	6.02	0.90	80.36
District of Columbia	DC	12.64	8.63	13.97	9.51	77.49
Florida	FL	22.11	4.76	6.51	1.71	77.05
Hawaii	HI	22.75	1.40	6.23	0.69	74.64
Kansas	KS	12.58	9.98	8.58	1.55	79.37

2.8.2   First Phase Results

We run the model for 2 phases.

In the first phase, we determine the optimal starting month of simulation. For all 6 locations, we run the model with 12 cases that start respectively on the first day of each month, and get the number and mean distance for each case and each location.

As an example, the results for Florida are shown in Fig.5. The dandelion numbers vary greatly among the case, but the mean distances are approximately the same. We can see that for Florida, we get the maximum dandelion number when the starting date is May 1st. So we will use May as the starting month in the next phase.
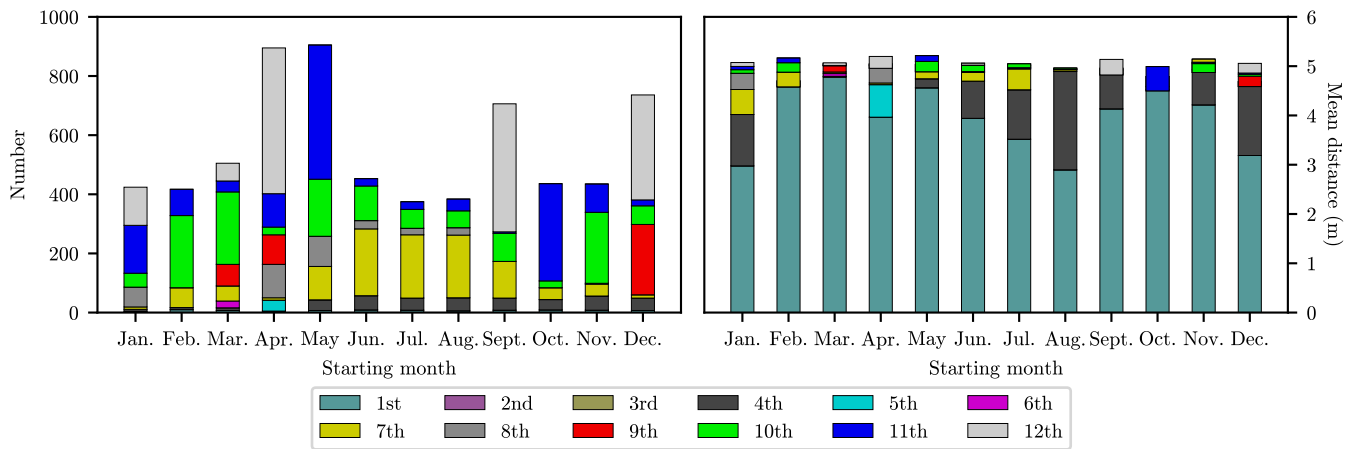

Fig. 5 Number and mean distance in Florida when simulation starts at different months

Table 5 lists all the data from the run of the first phase. The headers indicate the starting months, and the numbers in the table are the dandelion numbers at the end of 12 months. The numbers corresponding to the optimal starting months are marked in blue. In this way, we can determine that the optimal starting month is September for AK, August for CA, September for DC, May for FL, September for HI, and August for KS.

Table 5 Number at six locations when simulation starts at different dates

State	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
AK	74	79	70	35	44	75	43	49	**82**	72	74	75
CA	547	567	545	532	585	561	544	**596**	564	566	581	563
DC	1611	1589	1609	1563	1680	1657	1673	1733	**1808**	1696	1677	1658
FL	424	417	505	895	**905**	453	375	384	706	436	435	736
HI	519	329	610	593	586	596	308	543	**618**	383	385	600
KS	823	276	268	405	548	274	266	**1067**	287	288	950	834

2.8.3   Second Phase Results

In the second phase, we run the model 1000 times for each location beginning at the optimal starting month from the first phase. The resulting frequency histogram for Florida is shown in Fig.6. We can see that both the number and the mean distance follow an approximately normal distribution.
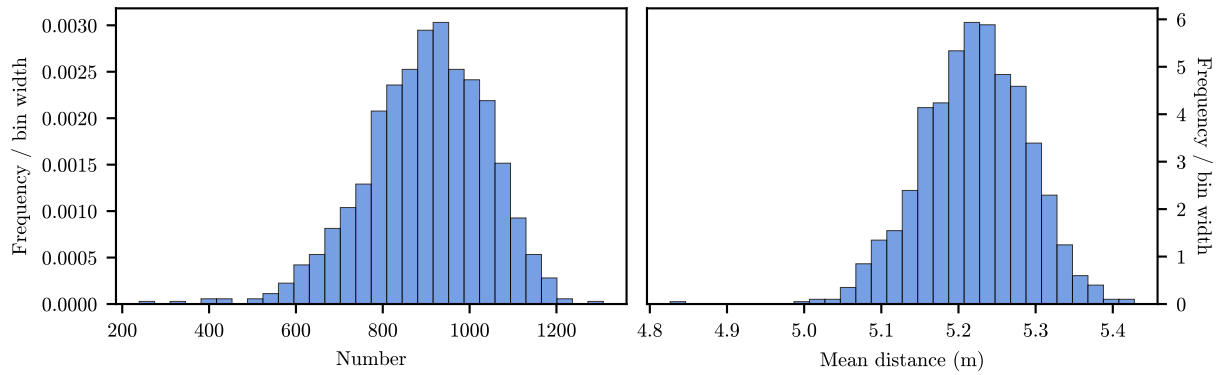
Fig. 6 Frequency distribution of number and mean distance in Florida

We calculated several statistical characteristics for each location, as shown in Table 6 and Table 7. We examine the statistics for dandelion number. The confidence intervals of the means are relatively small for all locations, which implies the means are quite accurate. The standard deviations and ranges are large, which shows that the spread of dandelions is easily affected by random factors such as fluctuations in the environment. For all locations except California, the distributions are platykurtic. For all locations except Alaska, the distributions are skewed to the left. Their masses are concentrated on the right, so the dandelion numbers are slightly more likely to be greater than the means rather than smaller. However, because of the unfavorable environmental conditions in Alaska, the number is more likely to be smaller than the mean.

Table 6 Descriptive statistics of number at six locations

Location	Mean	Standard Deviation	Kurtosis	Skewness	Range	Median	Confidence Level (95.0%)
AK	72.7	26.5	-0.11	0.38	150	71	1.64
CA	595.1	82.48	3.78	-1.32	744	609	5.11
DC	1780	484.9	-0.04	-0.03	3274	1772.5	30.1
FL	886.4	140.4	1.04	-0.72	981	905	8.71
HI	607.5	81.9	1.12	-0.79	565	617	5.08
KS	1066	186.4	1.05	-0.72	1403	1096	11.6

Table 7 Descriptive Statistics of the mean distance at six locations

Location	Mean	Standard Deviation	Kurtosis	Skewness	Range	Median	Confidence Level (95.0%)
AK	4.29	0.18	0.44	0.25	1.28	4.29	0.01
CA	4.52	0.09	0.51	-0.53	0.67	4.52	0.01
DC	11.84	0.21	1.5	-0.51	1.93	11.85	0.01
FL	5.21	0.07	1.91	-0.64	0.66	5.21	0.00
HI	4.99	0.08	0.138	-0.27	0.5	4.99	0.00
KS	6.52	0.05	0.8	-0.36	0.44	6.52	0.00

Out of the 1000 runs, we select the run that has dandelion number closest to the mean value. The positions and current stages of the dandelions at the end of 12 months are shown in Fig.7. For all locations except the District of Columbia, most dandelions grow in a semicircular ring centered at the origin. Alaska is apparently unsuitable for dandelions to spread. In California, Hawaii, Florida, and Kansas, dandelions all spread to a moderate range.
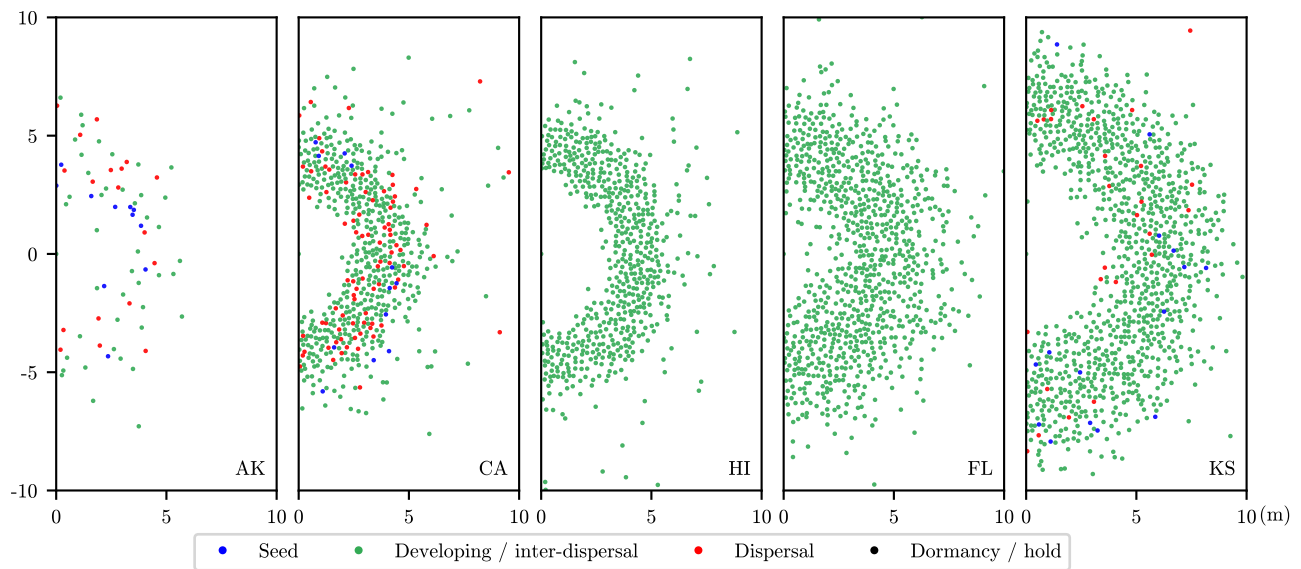
Fig. 7 The spread at the end of 12 months at different locations
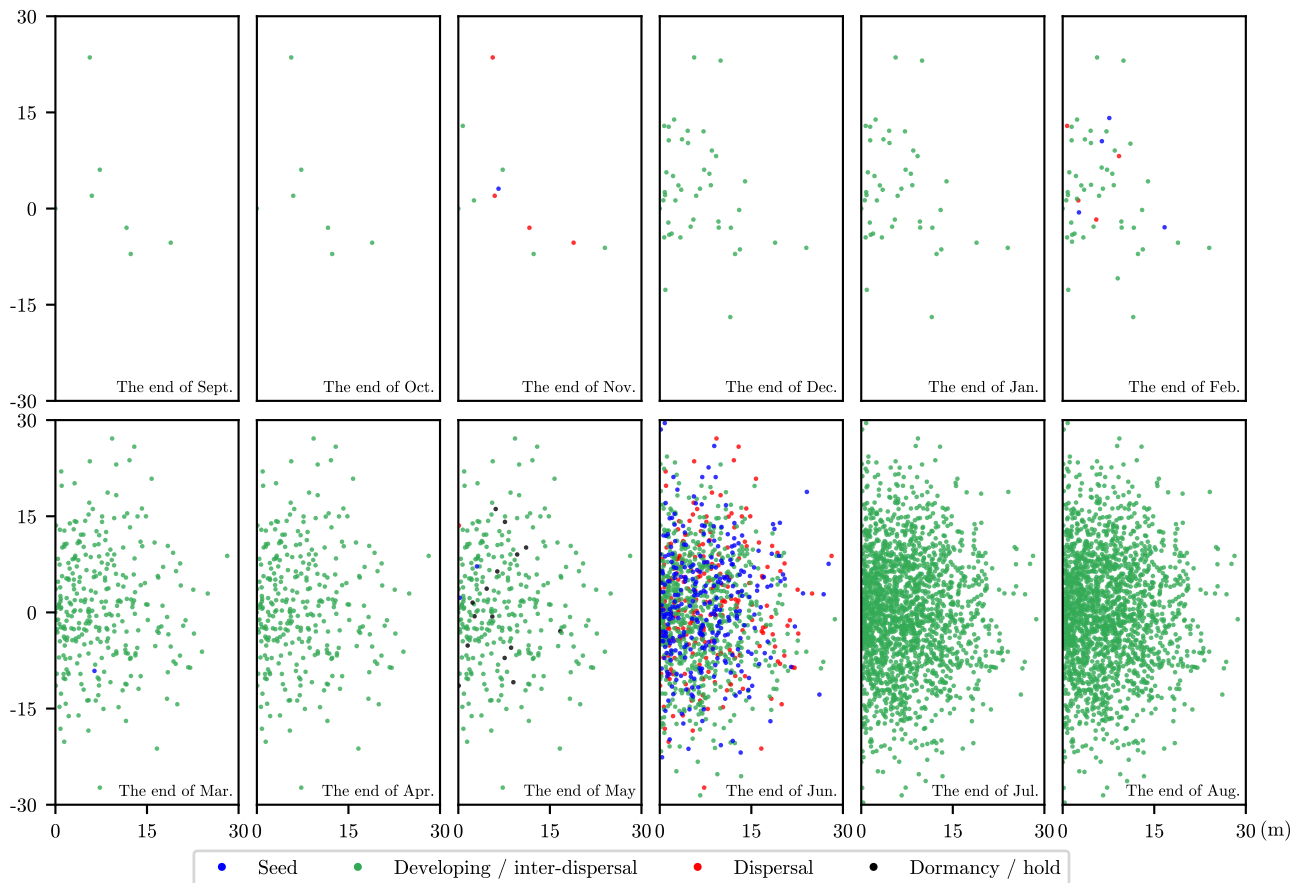

Fig. 8 The spread during 12 months in the District of Columbia

The District of Columbia has the most dandelions at the end of 12 months. We plot dandelion spread at the end of each month at this location (see Fig.8) so that it can be clearly seen how the dandelions spread in the course of time. Compared to other locations, the District of Columnbia has greater wind speed so that the dandelions spread much further and they form a semicircle instead of a ring. The closer to the origin, the denser the plants. Additionally, there is a sudden increase every three months, which is caused by the flowering of a new generation of dandelions.

Finally, we draw a graph of the natural logarithm of dandelion number in Fig.9. The lines have several step changes and platforms. The dandelion number increases exponentially at the end of each fixed period. This conclusion coincides with our observation in Fig.8. Again, Alaska is different. Dandelions in Alaska are dormant in the first three months of a year, and eventually produce fewer offspring than other locations.
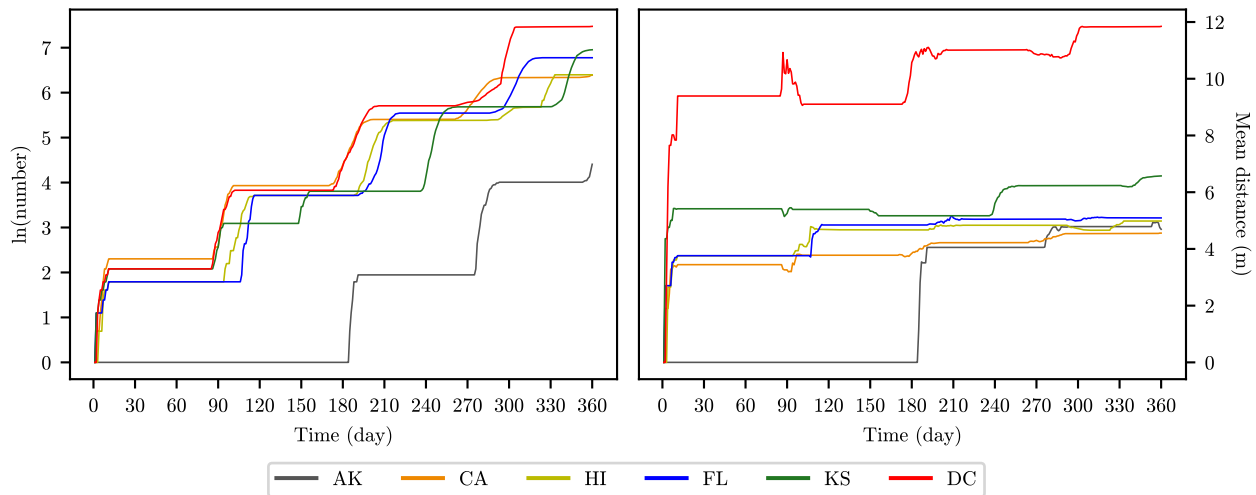

Fig. 9 Number and mean distance over time

2.9   Sensitivity Analysis

We conduct a sensitivity analysis on the five environmental factors from Table 1 by varying one of them while keeping the other four fixed.

Fig.10 shows the results for temperature. When μ_T falls below 7°C or rises above 23°C, dandelion number decreases sharply. This is because dandelions are transferred to the dormancy or hold stages at certain temperatures. The number is very sensitive to extreme temperatures, but only moderately sensitive to mild temperatures. The mean distance increases with μ_T up to 19°C, where it temporarily levels off and begins to fall at 23°C. It has medium sensitivity to μ_T.

Both number and mean distance decrease as σ_T increases. They decrease relatively slowly below $\sigma_T = 5$°C and rapidly above that temperature. However, the overall sensitivity is not high for either number or mean distance.

Fig.11 shows the results for wind speed. The number and mean distance have positive correlations with μ_W and σ_W and change drastically when the two factors vary. When μ_W increases, seeds are dispersed farther and land further apart from each other. When σ_W increases, there is greater variation in dispersal distance. In both cases it is less likely for seeds to land on other plants or exceed the density, making dandelion number rise quickly.
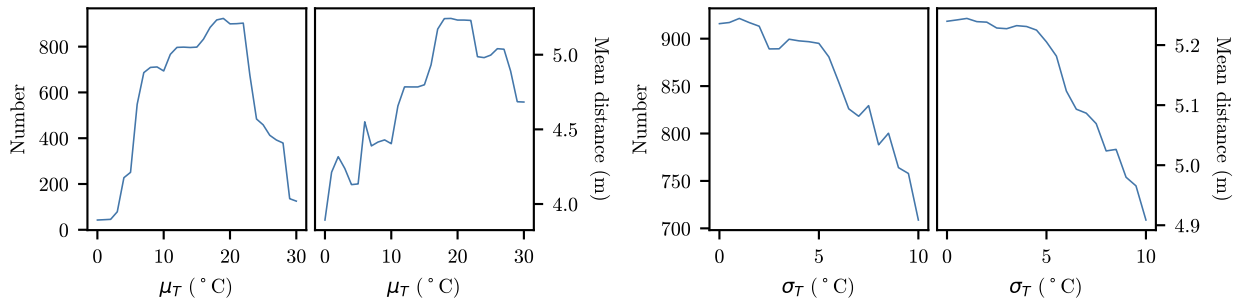
Fig. 10 Sensitivity analysis: temperature
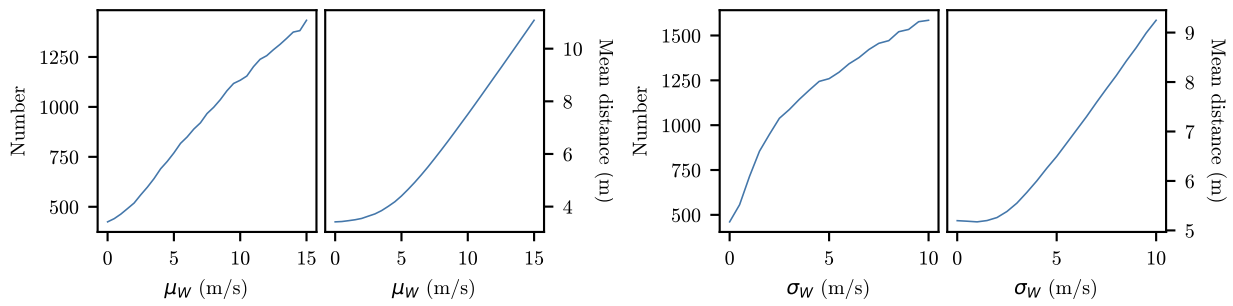
Fig. 11 Sensitivity analysis: wind speed

Fig.12 shows the results for humidity. There are three levels of number and mean distance corresponding to different humidity. This is because of the three ranks for humidity that are used in the calculation of the adaptation factor k (see Section 2.5).
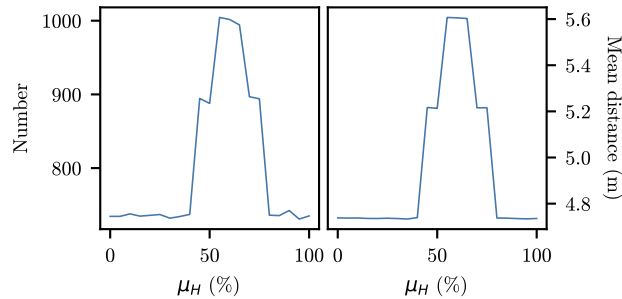
Fig. 12 Sensitivity analysis: humidity

2.10   Strengths and Weaknesses

Strengths

- Our model is **realistic**. We use data for real locations. And we consider how environmental factors and random variations contribute to dandelion spread. Besides, we track every plant during the entire spread process.
- Our model is **easy to understand and implement**. The simulation method is very intuitive, and the model only requires five input parameters that are easily acquired from meteorological databases.

- Our model is **extensible**. We can easily consider other factors by adding a component to the model. For example, if we want to consider the effect of rain, we can determine the effects of rain on the different stages of dandelions and add this new component to our current model.

Weaknesses

- Our model is **time-intensive**. To get a fairly accurate result, we have to run 1000 times for every location. Compared to other possible models, the efficiency of our model is relatively low.
- The **adaptation factor** may be an over-simplification. It may not represent the reality very accurately. If we can acquire the relationship between dandelion parameters and climate conditions, we will be able to substitute the adaptation factor with some more accurate components.
- Our model **misses certain factors**. We do not consider rain, which stops flowers from blooming and hinders seed dispersal. Other missed factors include interactions with other plants and human intervention. Also, regarding long dispersal distance, the wind speed is not considered. These can be added as new components in the future.

3 Plant Impact Factor Model

3.1 Assumptions and Justifications

1. **A plant's impact has a positive correlation with its biological characteristics.**
   Justification: Study shows that the invasiveness of a plant species is closely related to its performance-related traits[22]. Thus we can think of biological characteristics as fairly reliable indicators on the invasiveness, which is the plant's ability to cause harm to the environment and to humans.

2. **A plant's impact can be lessened by its benefits.**
   Justification: If plants are edible for animals or beneficial for the environment in some other ways, it is likely that the environment will be subject to less harm overall. Likewise, if plants can be made into products or utilized in other ways, they can benefit the economy and cause less harm to humans.

3. **At locations with higher population density, the impact of a plant is greater.**
   Justification: At places with higher population density, it is likely that people try harder to control undesirable species. This leads to more expenditure, which means the impact is greater. At places with low population density, there is less human intervention. Non-native plants may reach an equilibrium with native ones, thus causing less impact.

3.2 Model Description

3.2.1 Model Overview

The **impact factor** is a measure of adverse impact that an invasive plant can make on the environment and human beings.

We define two types of impact factors, namely **global impact factor** and **local impact factor**.

With global impact factor, we can compare the invasiveness between species in general perspective. For example, by using global impact factor, we can tell which one has more invasiveness between dandelion and clover, by and large.

Local impact factor does more refined work. By introducing more local adaptation attributes, we can tell the difference of invasiveness between locations for one species . We can also compare the invasiveness for different species in different places.

The impact factor is determined by a group of impact attributes. We break down all the attributes into 3 categories. **Category A** is related to plant characteristics, **Category B** is related to human and environment and **Category C** is related to plant's local adaptation. Global impact factor considers attributes from both Category A and Category B. Local impact factor considers attributes from all 3 categories.

Table 8 lists all the variables we will use in this model.

Table 8 Variable definition

Symbol	Definition
IF_G	Global impact factor
IF_L	Local impact factor
ρ	Invasion level coefficient
a_i	Attribute value for A_i
b_j	Attribute value for B_j
c_k	Attribute value for C_k
α_i	Weight for attribute A_i
β_j	Weight for attribute B_j
γ_k	Weight for attribute C_k

The formulas for calculating global impact factor and local impact factor are as follows:

$$\text{IF}_G = 100\rho \left(\sum \alpha_i a_i + \sum \beta_j b_j \right), \tag{5}$$

$$\text{IF}_L = 100\rho \left(\sum \alpha_i a_i + \sum \beta_j b_j + \sum \gamma_k c_k \right). \tag{6}$$

- **Invasion level coefficient** ρ is 0.9 if a plant is less widespread invasive; otherwise, it is 1.0.
- **Attribute values** a_i, b_j and c_k are determined per Section 3.2.2.
- **Attribute weights** α_i, β_j and γ_k are determined per Section 3.2.3.

3.2.2 Attribute values

Every attribute contains a list of options described in words than in numbers. In order to leverage Eq.5 and Eq.6, we need to map the options from words to numbers per Table 9.

Table 9 Attributes values

Symbol	Attributes	a_i, b_i and c_i Evaluation
Category A - Plant Characteristics		
A_1	Duration	$a_1 = 0.5$ (Annual), 0.75 (Biennial), 1 (Perennial)
A_2	Growing Habit	$a_2 = 0$ (Tree), 0.25 (Shrub), 0.5 (Vine), 0.75 (Graminoid), 1 (Forb/herb)
A_3	Growth Rate	The growth rate after successful establishment. $a_3 = 0.5$ (Slow), 0.75 (Moderate), 1 (Rapid)
A_4	Lifespan	$a_4 = 0.25$ (when $a_1 = 0.5$ or 0.75), 0.5 (Short), 0.75 (Moderate), 1 (Long)
A_5	Fertility Requirement	Relative level of nutrition (N, P, K) required for normal growth and development. $a_5 = 0.5$ (Low), 0.75 (Medium), 1 (High)
A_6	Fruit/Seed Abundance	The number of seeds produced. $a_6 = 0.25$ (None), 0.5 (Low), 0.75 (Medium), 1 (High)
A_7	Propagated Methods	The number of propagation methods is n_7. The methods can be Propagated by Bare Root, by Bulb, by Container, by Corm, by Cuttings, by Seed, by Sod, by Sprigs, or by Tubers. $a_7 = 0.25$ ($n_7 = 1$), 0.5 ($n_7 = 2$), 0.75 ($n_7 = 3$), 1 ($n_7 \geq 4$)
A_8	Seed Spread Rate	The capability of the plant to spread through its seed production. $a_8 = 0.25$ (None), 0.5 (Slow), 0.75 (Moderate), 1 (Rapid)
A_9	Seedling Vigor	The expected seedling survival percentage of the plant. $a_9 = 0.5$ (Low), 0.75 (Medium), 1 (High)
Category B - Human and Environment		
B_1	Toxicity	The relative toxicity of the plant to either humans or livestock. $b_1 = 0$ (None), 0.5 (Slight), 0.75 (Moderate), 1 (Severe)
B_2	Product	The level of the plant known to be suitable for multiple types of products. $b_2 = 0$ (Copiousness), 0.25 (Many), 0.5 (Some), 0.75 (few), 1 (None)
B_3	Palatable Animal	The relative palatability of this plant to browsing animals or to grazing animals. $b_3 = 0$ (High), 0.5 (Moderate), 0.75 (Low), 1 (None)
B_4	Palatable Human	The plant produce berries, nuts, seeds, or fruits are palatable to humans. $b_4 = 0.5$ (Yes), 1 (No)
B_5	Commercial Availability	The plant propagules are in the commercial marketplace. $b_5 = 0.5$ (Yes), 1 (No)
Category C - Location		
C_1	Soil Adaptation	The soil adaptation level. $c_1 = 0$ (Low), 0.5 (Medium), 1 (High)
C_2	Temperature Adaptation	The temperature adaptation level. $c_2 = 0$ (Low), 0.5 (Medium), 1 (High)
C_3	Humid Adaptation	The humid adaptation level. $c_3 = 0$ (Low), 0.5 (Medium), 1 (High)
C_4	Population Density	The level of population density. $c_4 = 0$ (Low), 0.5 (Medium), 1 (High)

Attributes B_2 (Product), B_3 (Palatable Animal), B_4 (Palatable Human), and B_5 (Commercial Availability) are special since they measure benefits rather than harms. Per Assumption 2, the greater the benefit, the less the impact. Therefore, when the palatability is higher, or when the plant has more uses, the corresponding values are smaller.

3.2.3 Attribute weights

We use the Analytic Hierarchy Process (AHP) to obtain the weight vector. When it is not easy to construct a weight vector by conferring to research results, AHP is an option. Since AHP makes pairwise comparisons and has a method to evaluate the rationality for those comparisons, it can reach a relatively reasonable result.

We make comparisons between attributes according to Saaty's fundamental scale of absolute numbers[16]. When two attributes are equally important, we get 1; when one attribute is moderately more important than the other, we get 3; when one attribute is strongly more important than the other, we get 5; when one attribute is much strongly more important than the other, we get 7; when one attribute is extremely more important than the other, we get 9. Number 2, 4, 6 and 8 are used to express intermediate values.

The resulting comparison matrix is shown in Table 10.

Table 10 AHP comparison matrix

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	B_1	B_2	B_3	B_4	B_5	C_1	C_2	C_3	C_4
A_1	1	1/3	1/7	1/3	1	1/7	1/4	1/7	1/5	1/9	1/5	1/3	1/4	1/4	1/5	1/5	1/5	1/3
A_2	3	1	1/5	1	3	1/5	1/2	1/5	1/3	1/7	1/3	1	1/2	1/2	1/3	1/3	1/3	1
A_3	7	5	1	5	7	1	4	1	3	1/3	3	5	4	4	3	3	3	5
A_4	3	1	1/5	1	3	1/5	1/2	1/5	1/3	1/7	1/3	1	1/2	1/2	1/3	1/3	1/3	1
A_5	1	1/3	1/7	1/3	1	1/7	1/4	1/7	1/5	1/9	1/5	1/3	1/4	1/4	1/5	1/5	1/5	1/3
A_6	7	5	1	5	7	1	4	1	3	1/3	3	5	4	4	3	3	3	5
A_7	4	2	1/4	2	4	1/4	1	1/4	1/2	1/4	1/2	2	1	1	1/2	1/2	1/2	2
A_8	7	5	1	5	7	1	4	1	3	1/3	3	5	4	4	3	3	3	5
A_9	5	3	1/3	3	5	1/3	2	1/3	1	1/5	1	3	2	2	1	1	1	3
B_1	9	7	3	7	7	3	4	3	5	1	5	7	4	4	5	5	5	7
B_2	5	3	1/3	3	5	1/3	2	1/3	1	1/5	1	3	2	2	1	1	1	3
B_3	3	1	1/5	1	3	1/5	1/2	1/5	1/3	1/7	1/3	1	1/2	1/2	1/3	1/3	1/3	1
B_4	4	2	1/4	2	4	1/4	1	1/4	1/2	1/4	1/2	2	1	1	1/2	1/2	1/2	2
B_5	4	2	1/4	2	4	1/4	1	1/4	1/2	1/4	1/2	2	1	1	1/2	1/2	1/2	2
C_1	5	3	1/3	3	5	1/3	2	1/3	1	1/5	1	3	2	2	1	1	1	3
C_2	5	3	1/3	3	5	1/3	2	1/3	1	1/5	1	3	2	2	1	1	1	3
C_3	5	3	1/3	3	5	1/3	2	1/3	1	1/5	1	3	2	2	1	1	1	3
C_4	3	1	1/5	1	3	1/5	1/2	1/5	1/3	1/7	1/3	1	1/2	1/2	1/3	1/3	1/3	1

Note that attribute B_1 (Toxicity) is a predominant attribute in our model. Toxic plants can poison the local environment and lead to a large expenditure for humans to control them. Attributes A_3 (Growth Rate), A_6 (Fruit/Seed Abundance), A_8 (Seed Spread Rate) are also important because they determine the spread rate of a plant when it is introduced to a new place.

For the global impact factor, the comparison matrix is comprised of the numbers from first 14 rows and first 14 columns. For the local impact factor, the comparison matrix is comprised of all the numbers in the table.

We conduct a consistency test on the global matrix and the local one. The resulting CRs are small enough for both matrices, so both matrices pass the test.

Table 11 Consistency test

Impact Factor Type	Attribute Number n	Max Eigenvalue λ_{max}	$CI = \dfrac{\lambda_{max} - n}{n - 1}$	RI	$CR = \dfrac{CI}{RI}$	CR < 0.1?
Global	14	14.501	0.0386	1.49	0.0259	Yes
Local	18	18.561	0.0330	1.49	0.0221	Yes

For each matrix, we normalize the columns so that each column adds up to 1. Then, we calculate the sum of each row to get a sum vector. Last, we divide the sum vector by n (the attribute number in Table 11) to get the final weight vector.

The weight results are shown in Table 12.

Table 12 Weights results

Type	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
Global	0.014	0.027	0.134	0.027	0.014	0.134	0.044	0.134	0.066
Local	0.011	0.021	0.113	0.021	0.011	0.113	0.034	0.113	0.052

	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2	γ_3	γ_4
Global	0.227	0.066	0.027	0.044	0.044	-	-	-	-
Local	0.192	0.052	0.021	0.034	0.034	0.052	0.052	0.052	0.021

3.2.4 Algorithm

Analytic Hierarchy Process Model for Impact Factor

INPUT: All the attribute options and **Invasive level coefficient** ρ

global value vector ← evaluate all the **Attribute values** from inputted attribute options per Table 9
local value vector ← evaluate all the **Attribute values** from inputted attribute options per Table 9
global weight vector ← Table 12
local weight vector ← Table 12
IF_G ← 100ρ times the dot product of *global weight vector* and *global value vector*
IF_L ← 100ρ times the dot product of *local weight vector* and *local value vector*

OUTPUT: IF_G and IF_L

3.3 Results

3.3.1 Global Impact Factor Results

We downloaded a database containing a list of invasive plants from US-RIIS[20] and chose 226 plants from it. These plants belong to 3 classes, 26 orders and 53 families.

In order to get all the plant attributes, we developed a web crawler to grab data from USDA Plants Database[19]. Thus, we have all the attribute options ready for each plant.

Then we implement this model to calculate the global impact factors for all 226 plants and rank them according to the factor. Table 15 shows some results.

Table 15 Plant Rankings

Rank	Scientific Name	Vernacular Name	Duration	Growing Habit	IF_G
1	*Melilotus officinalis*	Sweetclover	Perennial	Forb/herb	84.5
2	*Senecio vulgaris*	Old-man-in-the-spring	Biennial	Forb/herb	83.3
3	*Ailanthus altissima*	Tree of heaven	Perennial	Tree	81.8
4	*Crotalaria spectabilis*	Showy rattlebox	Annual	Forb/herb	81.7
5	*Ranunculus repens*	Creeping buttercup	Perennial	Forb/herb	80.8
6	*Melilotus indicus*	Annual yellow sweetclover	Annual	Forb/herb	80.0
7	*Digitalis purpurea*	Purple foxglove	Biennial	Forb/herb	79.9
8	*Sorghum halepense*	Johnsongrass	Perennial	Graminoid	78.9
9	*Schedonorus arundinaceus*	Tall fescue	Perennial	Graminoid	76.8
10	*Vicia sativa*	Garden vetch	Annual	Forb/herb	73.8
⋮					
26	*Taraxacum officinale*	Common dandelion	Perennial	Forb/herb	66.9
⋮					

We observe that half of the top-10 plants are perennials and 7 of them are forbs/herbs. Intriguingly, the tree of heaven, which is a tree, has made it to the third place. This is mostly because it is toxic and grows rapidly.

Dandelions have a relatively high ranking but are not among the top species. This is because dandelions are edible and have medicinal and culinary use, which lowers the global impact factor.
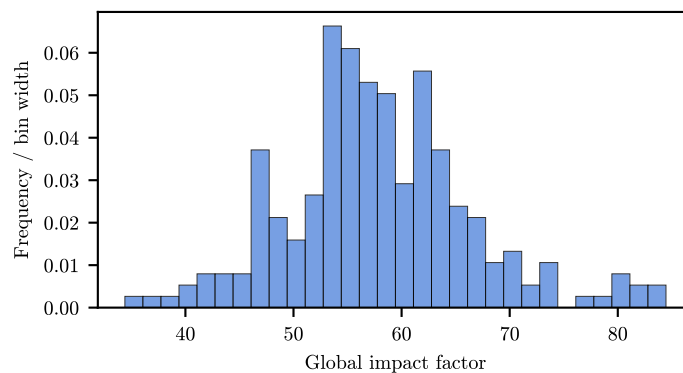

Fig. 13 Frequency distribution of global impact factor

Table 16 Descriptive statistics of the global impact factor

Mean	Standard Deviation	Kurtosis	Skewness	Range	Median	Confidence Level (95.0%)
57.9	8.65	0.81	0.41	50.51	57.25	1.13

We draw a histogram by using 226 global factors IF_G (see Fig.13), and calculate several statistical characteristics for the distribution (see Table 16).

IF_G is roughly concentrated around the mean and the median but does not exactly follow a normal distribution. The distribution is skewed to the right according to the table, meaning that there is a longer right tail. However, we can observe a left tail as well on the histogram. Neither tail is very long. There are no global impact factors below 30 or above 90.

3.3.2 Local Impact Factor Results

For calculating local impact factor IF_L,

- First, we choose 3 locations Alaska (AK), Florida (FL), and Nevada (NV). Alaska is extremely cold, Florida is warm and humid, and Nevada is hot and dry;

- Then, we choose 4 plants that are invasive at these locations (there are two exceptions: no trees of heaven in Alaska and no yellow salsifies in Nevada). Sweetclover ranks the first in global impact factor; the tree of heaven is the highest ranking tree in terms of global impact factor (rank 3); yellow salsify has the highest global impact factor among nontoxic plants (rank 21); also, the dandelion is not left out (rank 26);

- Last, we prepare all the attributes options with regards to **Category C** and implement the model to get local impact factor IF_L.

The results are shown in Fig.14.

For each plant, their impact would be greater at locations with more favorable environments, and vice versa.

For different plants, their local impact factor ranking roughly conforms to the global one, but there are exceptions. For example, though sweetclover has a IF_G higher than that of the tree of heaven, the IF_L for sweetclover in Alaska is lower than that for the tree of heaven in Florida.
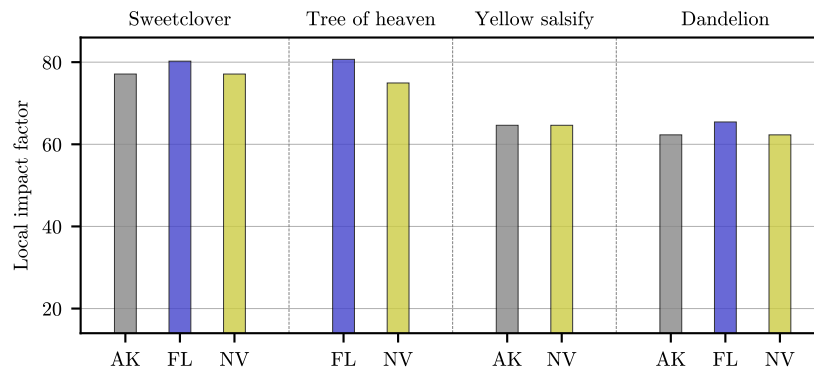
Fig. 14 Local impact factor

3.4 Sensitivity Analysis

In Eq.5 and Eq.6, the impact factors are evaluated as 100ρ multiplied by a linear combination of attribute values. We discuss the model's sensitivity as follows:

- For variable ρ, when other variables do not change, changing ρ from 1.0 to 0.9 will lead to a 90% discount to current IF value;

- For one certain attribute value, when other variables do not change, the change of impact factor is in proportion with the change of the attribute value. The formula is as follows:

$$\Delta\text{IF} = 100\rho \cdot w\Delta v$$

Where:

ΔIF is the change of impact factor value;
Δv is the change of the attribute value;
w is the attribute weight corresponding to the attribute.

For example, if only attribute B_1 (Toxicity) changes from "Slight" to "Severe", the change of attribute value Δv equals 0.5 (Table 10). Considering that the weight of B_1 for global impact factor is $\beta_1 = 0.227$ and for local impact factor is $\beta_1 = 0.192$ (Table 12), we will conclude that global impact factor will increase 11.4ρ and local impact factor will increase 9.6ρ.

3.5 Strengths and Weaknesses

Strengths

- Our model is **easy to implement**. For one plant, we can design a questionnaire to collect its attribute options and use our model to calculate the impact factors quickly.

Weaknesses

- Our model is **subjective**. The weights are subjectively determined. If we could involve more research results in determining the attribute weights, our model could be more accurate.

- Our attribute design may be **uncomprehensive**. It is possible that we have overlooked some other important attributes.

Our Vigorous Friend

It is always a dreamlike scene to see dandelion seeds flying on a summer day. All those little parachutes drift in the air, slowly and gently, through a beam of light, like angels from the sky.

We know how children love dandelions. They never tire of blowing puffballs and running after the parachutes. But we also know how adults dislike them. The dandelions cause allergy, grow with our crops, and spread like mad. Most of the time, we label them as invasive herbs that need great effort to take care of.

Whether we like them or not, dandelions spread all over the world. They can adapt to almost all kinds of environments, regardless of what the soil quality, humidity and temperature are. They grow fast, have abundant seeds and propagate without relying on other creatures. They never tire of dispersing, germinating and flowering. Looking at them, we have to admit they are so vigorous and full of vitality.

That is why they cannot stop "invading". But we need to be aware that "invasion" should be a neutral word than a derogatory one for nature species. When we talk about species invading, it should be no difference from talking about the emergence, prosperity, mutation or even extinction of species. These are all natural processes. Study has proved that without human intervention, even invasive species can reach an equilibrium with others. At some points, the invasive species could develop very fast and occupy a lot of resources, but gradually, when other species go into competition sufficiently, all the species will get along in harmony and share our planet peacefully.

We believe it is the same for our human beings – we will get along well with all the species, including invasive ones.

Sometimes it is we humans who flap the butterfly's wing. Brazilian red eared turtles were a hit in the pet market. However, some people purchased them and released them in the wild due to their compassion. Soon the turtles dominated the aquatic ecosystem with their tenacious vitality. Humans bring species from here to there and artificially raise them. Humans sometimes introduce species which do not bring harm initially but do eventually. We humans break the balance of nature and make the whole ecological system blink.

The invasive species are just who they are. Most of them have a lot of merits to human beings. They can provide fodder for our livestock, or they can be food or medicine. They are so lovely when we learn to get along with them. Think about dandelions. They have never been labeled invasive until recently. But they have been evolving in this planet for about 30 million years! Long enough to prove their innocence. It is not fair to treat them like an enemy. They are just old friends, sometimes sweet, sometimes petulant, and after all they are just very vigorous and full of vitality. We are so lucky that they are around – children will definitely agree with us, and how about adults?

References

[1] *Best management practices (bmps) to prevent the introduction and spread of invasive species.* [Online]. Available: `https://www.forestasyst.org/invasive_species.cfm`.

[2] M. T. Brock and C. Galen, "Drought tolerance in the alpine dandelion," *Taraxacum ceratophorum*, pp. 1311–1321, 2005.

[3] *Current invasive plants (feature layer).* [Online]. Available: `https://catalog.data.gov/dataset/current-invasive-plants-feature-layer-9c83e`.

[4] *Dandelion.* [Online]. Available: `https://www.gardenorganic.org.uk/expert-advice/garden-management/weeds/weed-datasheets/dandelion`.

[5] W. K. C. M. G. Elaine Anderson, *Dandelions*, 2012. [Online]. Available: `https://web.archive.org/web/20120626161645/http://gardening.wsu.edu/column/05-04-03.htm`.

[6] *Everything you need to know about growing dandelions*, 2018. [Online]. Available: `https://veggiegardenseeds.com.au/products/dandelion-taraxacum-officinale-herb-x-100-seeds`.

[7] *Executive order 13112*, 1999. [Online]. Available: `https://www.govinfo.gov/content/pkg/FR-1999-02-08/pdf/99-3184.pdf`.

[8] D. Greene and E. Johnson, "A model of wind dispersal of winged or plumed seeds," *Ecology*, vol. 70, no. 2, pp. 339–347, 1989.

[9] A. Honek, Z. Martinkova, and P. Saska, "Post-dispersal predation of taraxacum officinale (dandelion) seed," *Journal of Ecology*, vol. 93, no. 2, pp. 345–352, 2005.

[10] W. Letchamo and A. Gosselin, "Light, temperature and duration of storage govern the germination and emergence of taraxacum officinale seed," *Journal of Horticultural Science*, vol. 71, no. 3, pp. 373–377, 1996.

[11] *Meet a plant: Dandelion.* [Online]. Available: `https://gardens.duke.edu/sites/default/files/Duke%20Gardens%20-%20Meet%20a%20Plant%20Dandelion.pdf`.

[12] N. Z. Muth and M. Pigliucci, "Traits of invasives reconsidered: Phenotypic comparisons of introduced invasive and introduced noninvasive plant species within two closely related clades," *American Journal of Botany*, vol. 93, no. 2, pp. 188–196, 2006.

[13] A. Okubo and S. A. Levin, "A theoretical framework for data analysis of wind dispersal of seeds and pollen," *Ecology*, vol. 70, no. 2, pp. 329–338, 1989.

[14] *Pimdat.* [Online]. Available: `https://www.imapinvasives.org/decision-analysis-tool`.

[15] J. M. Randall, L. E. Morse, N. Benton, R. Hiebert, S. Lu, and T. Killeffer, "The invasive species assessment protocol: A tool for creating regional and national lists of invasive nonnative plants that negatively impact biodiversity," *Invasive Plant Science and Management*, vol. 1, no. 1, pp. 36–49, 2008.

[16] T. L. Saaty, "Decision making with the analytic hierarchy process," *International journal of services sciences*, vol. 1, no. 1, pp. 83–98, 2008.

[17] S. Stewart-Wade, S. Neumann, L. Collins, and G. Boland, "The biology of canadian weeds. 117. taraxacum officinale gh weber ex wiggers," *Canadian journal of plant science*, vol. 82, no. 4, pp. 825–853, 2002.

[18] O. Tackenberg, P. Poschlod, and S. Kahmen, "Dandelion seed dispersal: The horizontal wind speed does not matter for long-distance dispersal-it is updraft!" *Plant Biology*, vol. 5, no. 05, pp. 451–454, 2003.

[19] *United states department of agriculture.* [Online]. Available: `https://plants.usda.gov/home`.

[20] *United states register of introduced and invasive species (us-riis).* [Online]. Available: `https://www.sciencebase.gov/catalog/item/5b911a5ce4b0702d0e808588`.

[21] *Usa location information - usa.com.* [Online]. Available: `http://www.usa.com/`.

[22] M. Van Kleunen, E. Weber, and M. Fischer, "A meta-analysis of trait differences between invasive and non-invasive plant species," *Ecology letters*, vol. 13, no. 2, pp. 235–245, 2010.

[23] K. J. Verhoeven and A. Biere, "Geographic parthenogenesis and plant-enemy interactions in the common dandelion," *BMC evolutionary biology*, vol. 13, pp. 1–8, 2013.

[24] K. Yoneda, H. Momose, and S. Kubota, "Effects of daylength and temperature on flowering in juvenile and adult phalaenopsis plants," *Journal of the Japanese Society for Horticultural Science*, vol. 60, no. 3, pp. 651–657, 1991.

[25] F. Yoshie, "Effects of day length and temperature on flowering in taraxacum platycarpum and t. officinale," *Bulletin of the Institute of Natural Sciences, Senshu University*, no. 51, pp. 53–62, 2020.