

A Study on the Prediction of Dandelion Seed Dispersal Based on Gaussian Probability Distribution

Jiachen Zhong^{*,†}, Guozheng Zhu[†], Huangzi Yan[†], Yutong Yan[†]

Chengdu Shishi High School, Chengdu, China, 610015

* Corresponding Author Email: ying2526@163.com

† These authors contributed equally.

Abstract. The "Spread Prediction Model" presented in the study focuses on forecasting the dispersion of dandelion seeds across a landscape using a Gaussian Probability Distribution. Initially assuming dandelions start from a central point in a 1-hectare area divided into 10,000 units, the model employs a two-dimensional Gaussian function to simulate seed spread influenced by wind direction and strength. Factors such as germination rates, seed production per plant, and survival rates across different climates (temperate, tropical, arid) are integrated into growth cycle simulations. These simulations, conducted in stages from seed dispersal to plant growth, are validated through sensitivity analysis of key parameters. Finally, the model's results are visualized using heat maps to demonstrate seed spread patterns over time and under varying environmental conditions.

Keywords: Gaussian probability distribution; Dandelion seed dispersal; Predictive Modelling.

1. Introduction

In today's field of ecology and environmental management, it is important to predict and understand the dispersal process of plant seeds in different geographical environments. Dandelion (*Taraxacum officinale*), a widely distributed herbaceous plant, has a significant impact on ecosystems as its light-flying seeds are able to spread rapidly over a wide area driven by wind. In order to effectively manage and control the dispersal of dandelion, researchers need to develop reliable mathematical models to simulate and predict the dispersal paths and number of its seeds [1-2].

The aim of this paper is to explore a prediction model for dandelion seed dispersal based on Gaussian probability distribution. The core of the problem lies in how to use mathematical models to accurately describe the dispersal pattern of dandelion seeds under the influence of wind force and direction. The modelling idea proposed in this paper is to divide the ground surface into fine cells and use a two-dimensional Gaussian distribution function to simulate the seed dispersal process. By considering factors such as growth cycle, seed production and survival rate under different climatic conditions, we will develop a comprehensive predictive model to reveal the dispersal dynamics of dandelion under different environmental conditions [3]. This study will provide a powerful tool for ecological researchers and environmental managers to optimise dandelion management and control strategies in order to maintain stable and healthy ecosystems [4-5].

2. Spread Prediction Model

2.1. Gaussian Probability Distribution

One hectare of open standard square area has a dandelion in the float stage in the geometric center. Because we need to predict the spread of the holistic land as accurately as possible, dividing it up into the smallest possible units of equal area can help us obtain more accurate predictions. In order to facilitate the calculation and to show the subsequent results, this paper arrives at the following division method: 1 hectare is equal to 10,000m², we divided it into 100*100 units, that is, 10,000 pieces unit land of 1m².

Because of dandelions have a certain height, and the usual wind direction is parallel to ground level. Considering the fact that there are various uncontrollable factors in the actual situation, and this paper is not able to take all the variables into account, we assumed and constructed an ideal system for the probability model as the first step as the assumption mentioned [6]. The probability of the initial dandelion radiating outward, propagating and landing is roughly expressed by the Gaussian probability density function.

Derivative condition has shown that the farther away from the plant, the probability that seeds from the same plant will land in the same area is greatly reduced according to the properties of the Gaussian probability density function. The actual situation is carried out on a two-dimensional plane, so a two-dimensional Gaussian distribution probability density function involving the plane is used. Introduced a probability density(Z) of two-dimensional continuous random variables (x, y), shown as follows:

$$Z = f(x, y) = \frac{1}{2\pi G_x G_y} \exp\left(-\frac{1}{2(1-\rho)^2} \left[\frac{x-M_x}{G_x} - 2\rho \left(\frac{x-M_x}{G_x} \right) \left(\frac{y-M_y}{G_y} \right) + \left(\frac{y-M_y}{G_y} \right)^2 \right]^2\right) \quad (1)$$

Where M_x and M_y represent the mean, G_x and G_y represent the standard deviation, and ρ represents the correlation coefficient. When $\rho=0$, that is, x and y are independent of each other, the equation can be expressed as follows:

$$Z = f(x, y) = \frac{1}{2\pi G^2} e^{\left(-\frac{(x-M_x)^2 + (y-M_y)^2}{2G^2}\right)} \quad (2)$$

In order to obtain the value of the final seed number and its distribution according to the probability density function, that is, the probability distribution function, this paper obtains the conclusion of the probability accumulation in the probability function according to the definition of the probability distribution. Similarly, the probability density function represents the rate of change of the probability, so we can obtain the distribution of the number of seeds in a specific region through the definite integral of the probability density [7-8].

Here we use P to represent the distribution function. Since it is a two-dimensional normal distribution, we can integrate about x -axis and about y -axis first, which are double integrals. The result will be the probability of seeds spreading over a certain area. The formula shown as follows:

$$P = \iint z \, dx dy \quad (3)$$

In summary, we have obtained the probability of the number of seeds in a certain area. A short relational expression can express the number of seeds in the region (N_g) by the probability of obtaining seeds (P_g) and the number of seeds spread(n_s):

$$N_{seed \, get} = P \times N_{seed \, set} \quad (4)$$

As mentioned above, the land is divided equally by $100 \times 100 = 10,000$ units of land of $1m^2$. We use a 100-order matrix to represent the 10,000 plots (rows i : 1~100, columns j : 1~100), and specific plots can be represented by Land $ij(L_{ij})$.

Similarly, As mentioned in the previous assumptions, the original dandelion is spreading out right in the middle, but the dandelion actually spreads out and successfully lands, germinates, and produces new seeds still needs to be taken into account. Therefore, we finally assumed that the number of seeds

spread by L_{ij} from one field to each field is N_{ij} . The amount of each propagation of all 10,000 land L_{ij} can be expressed by summation formula:

The conditions in the spread need to take into account the wind direction, so it is necessary to add bias to the original two-dimensional normal function. In the case of actual analysis, one additional assumption is required [9-10]: The overall direction of the wind is the same.

So the blowing distance of its wind needs to consider the comprehensive wind during the propagation period. The whole distribution function only changes in the means M_x and M_y , and the rest stays the same. At the same time we need to set a random coordinate interval for this moving coordinate. In order to fit the actual situation, this region conforms to the normal distribution.

According to many seasonal climate conditions, because it is a temperate monsoon climate. In the current climate, the duration of summer is relatively longer, so here we take the initial wind direction as southeast wind. In a tropical monsoon climate, the spring and autumn seasons last longer, so we take the northwest winds that dominate in spring and autumn. In drought conditions, the spring lasts longer, so we choose southeast winds.

2.2. Growth Cycle

Because we actually need to predict the spread of dandelions, the growth of dandelions after successful dispersal is also one of the factors that needs to be taken into account. Under the hypothetical conditions, the overall spread process can be roughly expressed as figure 1:

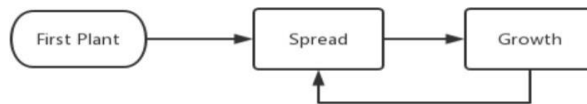


Figure 1. The relationship map of dandelion growth cycle

This is the cycle. The growth process can be further subdivided into smaller stages.

Before germination, it is still the seed stage. And then that is the stage of the first plant. This stage can be divided into seedlings, plants, flowers can spread in the float stage. Then there is the stage of the second plant, which is the same as the first plant. If we want to add the influence of environmental factors on its growth stage, we divide the process into seed, flowering, and seed setting (with the ability to spread). After the seeds are successfully planted, the environment, the quality of the seeds themselves and other factors will produce a certain number of deaths. So survival rates need to be considered from seed to successful flowering. After flowering, it is still affected by environmental factors, and there will be a certain death rate and the number of seeds. The environment will affect the overall growth time and the survival rate of each stage. The mind map is as figure 2:

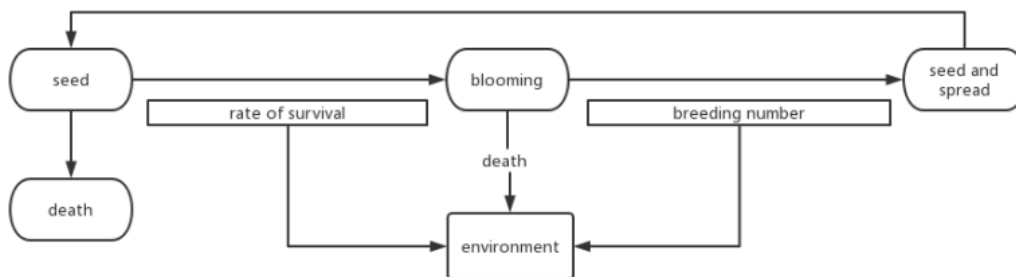


Figure 2. The relationship map of dandelion growth cycle

Table 1. Matrix diagram of impact weight on the ecosystem

	Germination Rate	Number of Seeds	Growth	Survival Rate
Temperate	95%	10*250	2	99%
Tropic	180%	8*200	3	90%
Desert/Arid	75%	5*150	4	50%

According to table 1, in the temperate climate environment (low latitude area), there are more rain, more storms, sufficient light, high humidity, and obvious changes in rainfall and storms during seasonal change;

In the northern region, there is less rain and less storms, and these changes are not obvious when the seasons change, but the obvious change is the temperature, and the temperature changes in the middle latitude region are distinct.

The effects of temperature on the Germination Rate (GR%), Germination Indicator (GI) and Vigor Indicator (VI) of dandelion seeds were considered. The germination rate of dandelion seeds was affected by varying temperature culture at 20 ~ 30°C. The normal germination rate was 95%, and because the temperate climate is not extreme, the survival rate is 99%. The number of seeds per dandelion is 10 times 250. (10 is the number of flowers per dandelion) the growth time is generally about February.

In the tropical climate environment, the temperature is generally higher throughout the year, and the four seasons are not obvious. The general temperature is above 25°C when the seed germination rate is about 80 percent. Because of the warmer temperatures, each dandelion plant will produce about eight flowers and each flower will produce 200 seeds. The growth time is generally around March. Due to the generally higher precipitation, root rot is easy to occur during germination, so the plant survival rate is about 50 percent. The growth time is generally in March.

For desert and other arid areas, the precipitation is very little and the temperature difference between day and night is not conducive to the spread and growth of plants. The germination rate is only about 75%. Each dandelion has about 5 flowers and 150 seeds per flower. The growth period is concentrated around April and the plant survival rate is only about 50%.

In this model of an one-stage plant from growth to successful spread, we assume that:

Ideally, all seeds should be able to complete one stage of growth at the same time. The number of seeds obtained in one stage can be subtracted from the number of seeds obtained in the previous stage by the total number of seeds obtained. In the actual situation, the survival rate of the seed (α_{seed}) and the survival rate of the plant (α_{plant}) need to be considered. As mentioned above, we are analyzing the growth of seeds by unit stage Δt , so this paper classifies the number of seeds of an unit stage as $\frac{\Delta N_{seedT}}{\Delta t}$. The above formula represents the number of seeds at the end of the current period. So we use the total number of seeds obtained minus the number of seeds obtained after the end of the previous phase to get the number of the current phase. ΔN_{seedT} indicates the number of seeds obtained during this period, and $N_{seed(T-1)}$ is the number of seeds obtained in the previous cycle. $N_{seedallset}$ represents the number of seeds dispersed.

$$\frac{\Delta N_{seed T}}{\Delta t} = \alpha_{seed} * N_{seed all set} - N_{seed(T-1)} \quad (5)$$

In this formula, we need to calculate the number of plants in the current stage by subtracting the number of surviving plants in the previous stage from the number of seeds obtained in the previous stage. Because it takes a period of time from seed to grow to a propagating plant, we used ΔN_{plantT} represents the number of plants grown in that period, and N_{plantT} is the number of plants grown in the previous cycle.

$$\frac{\Delta N_{plant T}}{\Delta t} = N_{seed all get} + \alpha_{plant} N_{plant (T-1)} \quad (6)$$

Therefore, we are able to represent the number of seeds (N_{ij}) that spread in a specific land ($Land_{ij}$). Obviously, we are also be able to calculate the number of seeds of all.

$$N_{seed} = \sum_{i=1}^{100} \sum_{j=1}^{100} N_{ij} \quad (7)$$

Similarly, the number of seeds produced can be expressed by multiplying the number of flowers by the number of seeds produced by each flower:

$$N_{seed\ set} = N_{flower} * N_{flower's\ seed} \quad (8)$$

2.3. Model Checking

The total number of dandelions is affected by the survival rate brought about by the local environment. We started off with the assumption that each plant produces an average of 500 seeds, and we reduced the survival rate from 85% to 50%, reducing each time by 1%, to figure out the difference in the total number of dandelions, and compare the difference with the original case.

Then change the number of seeds that each plant can produce, from 500 to 300, reducing each time by 10. Plot the changes again and compare the changes after changing the parameters, will show the importance of the number of seeds grown and the survival rate to the overall dandelion reproduction, if the two differences are similar and the difference is not large, then results are regard as stable.

We set 2 parameters, starting at 50% and increasing n by 0.5% each time until it reaches 100%. Connect to every one of them. And in the same way, we get z in the same way.

Compare the two curves of n and z, and get the result.

We need to measure dandelion fan winds and travel distances in temperate, tropical and arid regions. Now set a temperate monsoon climate at 40° north latitude, and a tropical monsoon climate at 20° north latitude. Then take 43° north latitude for instance. We use the average wind this year to estimate how far the dandelions can travel by different winds. The sensitivities of population size to germination rate, rate of survival and number of species per plan is shown in figure 3.

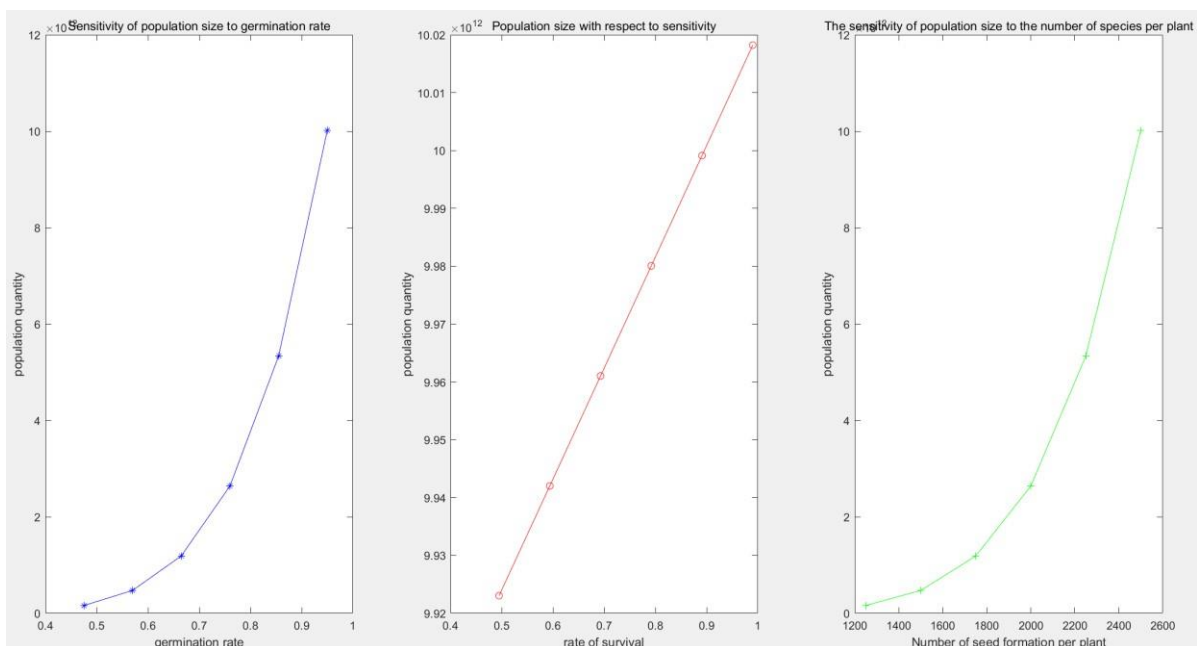


Figure 3. The sensitivities of population size to germination rate, rate of survival and number of species per plant

The degree of variation in the output of the model to determine its sensitivity to changes in system parameters or surrounding conditions. Therefore we made this sensitivity test of germination rate, rate of survival and the number of per plant to determine whether the model is reliable or not to use.

Generally, it can be seen from the figure that the influence of the changes of germination rate and Number of seed formation per unit on the final population number is exponential. Meanwhile, it is found that the influence of rate of survival on the population tree is linear. We found that the variation gap between germination rate figure and number of seed formation per plant figure was similar. It can be seen that the germination rate and Number of seed formation per unit were similar in the calculation of population size, while the rate of survival had the greatest impact. It can be seen that its weight is the highest.

2.4. Programming and Conclusion

In the process of programming, as mentioned above, we divided the whole land into $100 \times 100 = 10000$ square land units. So we treat it as a matrix in the program, and we end up using a matrix to represent the overall spread of the whole land. In the program, we also generally use the idea of modeling the formula to complete the programming content. As mentioned above, we assume that the process of propagation is instantaneous, and that the complete growth of the plant and the complete dispersal of the seed are completed in one stage. So in the first stage, at the end of the first propagation, the distribution of the plant is the one in the center, and the seeds are the result of the first superposition of the distribution functions. In the second stage, the distribution of the plants is the original plant plus the number of seeds in the previous section. The seeds of the previous stage have already grown into plants at this stage, so the number of plants at this stage should be multiplied by the distribution function.

As mentioned above, we determine the general wind direction and wind power in a year according to different climatic conditions, and therefore the degree of deviation of the function caused by the wind direction. According to the data we have queried, the germination rate of seeds, the survival rate of plants and the survival rate of plants under different climate conditions are further obtained, and their values are substituted into the formula of the program, as shown in figure 4 to figure 6. The results of the procedure are represented in part by heat maps, the spread of the seeds at different stages.

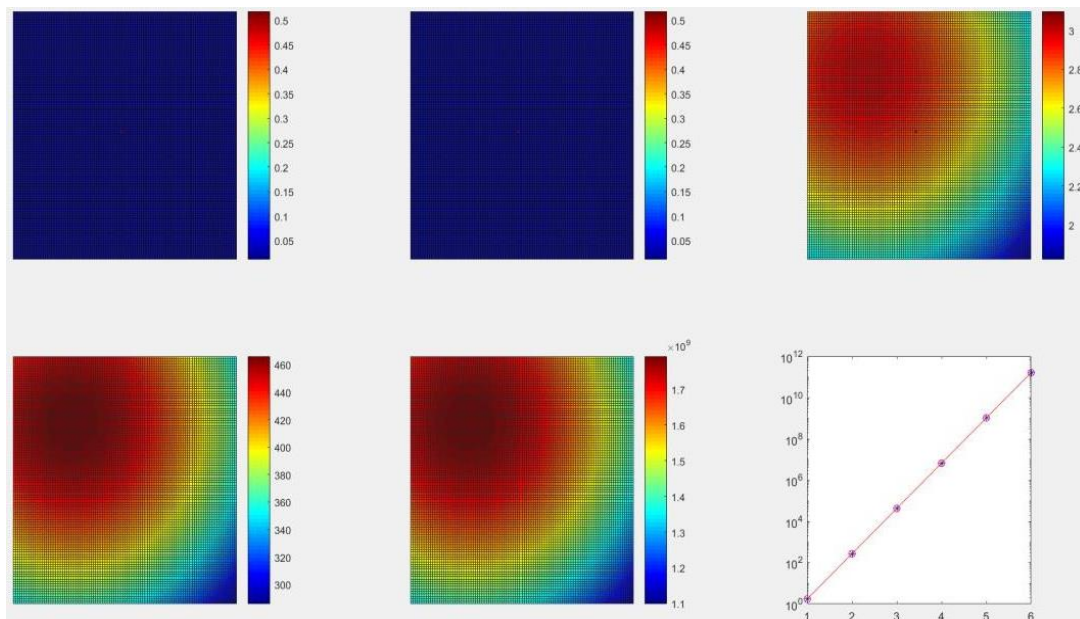


Figure 4. The results indicate the spread of dandelion seeds in dry climate

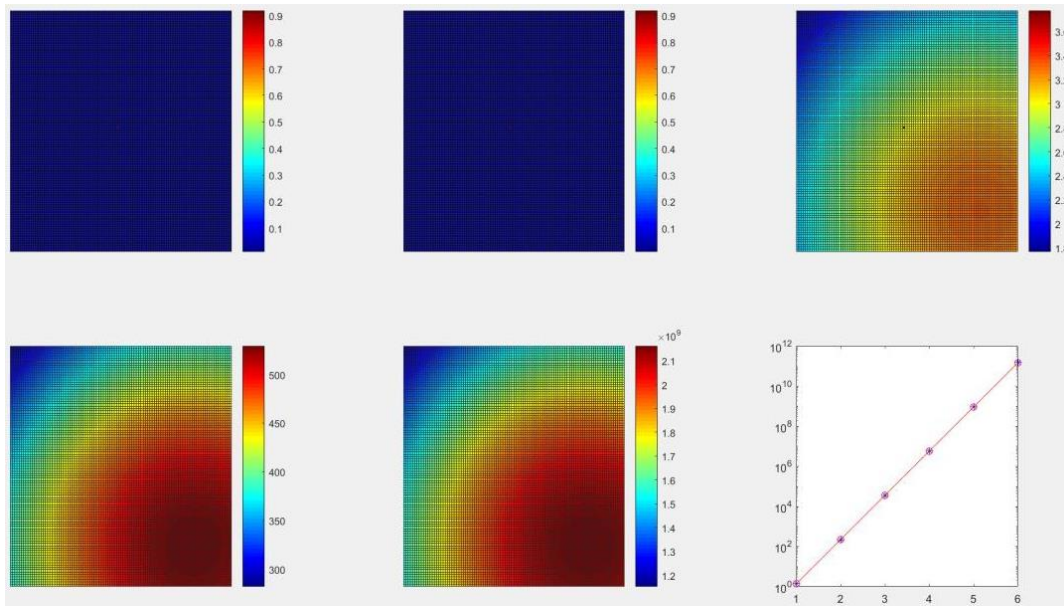


Figure 5. The results indicate the spread of dandelion seeds in tropical monsoon climate

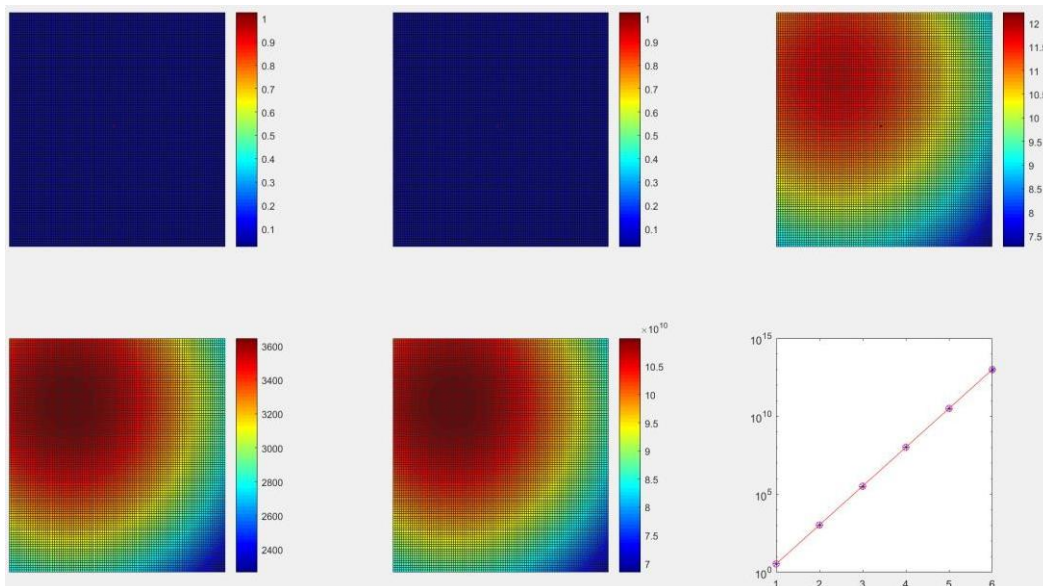


Figure 6. The results indicate the spread of dandelion seeds in temperate monsoon climate

From left to right and from top to bottom, the spread of the first five stages is shown respectively. The five charts represent the spread in January, February, March, June and December. The graph at the bottom right represents the overall trend over the six stages by logarithmic axes. The horizontal coordinate represents the number of stages, while the vertical coordinate represents the number of seeds in the population. There will be a large gap in the number of the two from the beginning, so the final effect of the image is very poor readability. Since the seed number of the ordinate is represented by an exponential function with a base of 10, we use the logarithmic function as the ordinate, so that the difference between the value expressed by the ordinate and the horizontal coordinate will not be too large, which can have better readability.

As mentioned above, we determine the general wind direction and wind power in a year according to different climatic conditions, and therefore the degree of deviation of the function caused by the wind direction. According to the data we have queried, the germination rate of seeds, the survival rate of plants and the survival rate of plants under different climate conditions are further obtained, and their values are substituted into the formula of the program. The results of the procedure are represented in part by heat maps, i.e.

The spread of the seeds at different stages: From left to right and from top to bottom, the spread of the first five stages is shown respectively. The five charts represent the spread in January, February, March, June and December. The graph at the bottom right represents the overall trend over the six stages by logarithmic axes. The horizontal coordinate represents the number of stages, while the vertical coordinate represents the number of seeds in the population. There will be a large gap in the number of the two from the beginning, so the final effect of the image is very poor readability. Since the seed number of the ordinate is represented by an exponential function with a base of 10, we use the logarithmic function as the ordinate, so that the difference between the value expressed by the ordinate and the horizontal coordinate will not be too large, which can have better readability.

3. Conclusion

In this study, we successfully established a mathematical model to predict the seed dispersal of dandelion based on Gaussian probability distribution, and deeply analysed the seed dispersal pattern and its influencing factors under different climatic conditions. Through the application of the model, we draw the following conclusions: Firstly, the model effectively simulated the dispersal path of dandelion seeds on the surface. By considering the effects of wind and wind direction on seed dispersal, we were able to accurately predict the distribution of seeds in different seasons and geographical locations. Second, sensitivity analyses of the model showed that seed germination rate, seed yield per plant, and plant survival rate had significant effects on population size. Changes in these parameters had important effects on the growth and dispersal rate of dandelion populations. Finally, the dispersal pattern of dandelion seeds in different climatic environments was demonstrated by heat maps, which provided a scientific basis and reference for the development of ecological management and control strategies.

In summary, this study not only provides a new perspective for an in-depth understanding of the dandelion seed dispersal process, but also provides reliable mathematical tools and methods for predicting and managing this ecological process. Future studies can further optimise the model and consider the influence of more ecological factors to improve the accuracy and applicability of the prediction.

References

- [1] Liao Huixuan, Zhou Ting, Chen Baoming et al. Ecological control of invasive alien plants [J]. Journal of Sun Yat-sen University (Natural Sciences), 2021, 60 (04): 1- 11.
- [2] Yu Haixin. Adaptation analysis of 4 small invasive pests and rapid detection of 3 invasive organisms by LAMP [D]. Tarim University, 2023.
- [3] Yang Nan-Wan, ZHANG Gui-Fen, LIU Wan-Xue. "Alien Invasive Species Management" column guide [J]. Plant protection, 2022, 48 (06): 1.
- [4] Wang Bingbing. Research on the propagation model of exotic organisms [D]. Zhejiang University of Technology, 2019.
- [5] RAO Hongjiang, Xiao Lixian, HE Yongtai et al. Maximum power point tracking of photovoltaic system based on dandelion optimization algorithm [J]. Electric Power Science and Engineering, 2023, 39 (09): 46-54.
- [6] ZHANG Qian, Xue Xiaoqiang, Xu Yanqiang. Optimization of finite element Model of Two- tower Cable-stayed Bridge based on Sensitivity Analysis [J]. Highway Traffic Technology, 2023, 39 (05): 104-112.
- [7] MA Wenshan. Application of definite integral and double integral in area calculation [J]. Journal of West Fujian Vocational and Technical College, 2022, 24 (02): 116-120.
- [8] He Hongying, Zhang Shi. Coordinates in the plane of the double integral of numerical algorithm [J]. Journal of mianyang normal university, 2020, 33 (11): 6 11-17.
- [9] Chi Jinsheng, Yang Dawei, Mao Lin. Heat map generation algorithm for weakly supervised semantic segmentation with multi-category labels [J]. Journal of Dalian Minzu University, 2023, 25 (01): 40-46.
- [10] Wu Chunming, Lian Liying, Zhao Liquan. Improved CenterNet object detection algorithm with introduction of joint point heat map [J]. Journal of Beijing Jiaotong University, 2022, 46 (02): 20- 28.