Application of Modeling and Programming to Predict Plant Growth

Jiahao Zhang *, Zhiqing Wang, Xiaokun Yin

Faculty of Information Science and Engineering, Ocean University of China, Shandong, China, 266100

* Corresponding Author Email: zw_2003@163.com

Abstract. Plants in different regions have different growth rates due to various environmental factors. At the same time, the biomass of different plant communities is also different because of the biomass of different plants. Based on the application of the principle of plant growth, this paper analyzes the factors affecting the number of plants in the plant community, including precipitation, species number and pollution, and obtains the mathematical model of the change of the number of plants in the community over time under different environmental conditions. After drawing the mathematical model, the biological growth conditions in the environment can be simulated through mathematical modeling. The optimal plant species in the community for the maximum benefit can be obtained through programming forms such as genetic algorithm. The paper provides a reference result for the growth of plants under different environmental conditions.

Keywords: Differential equations, genetic algorithms, likelihood estimation.

1. Introduction

As plants in different regions have different adaptability to the environment, changes in environmental factors such as precipitation and pollution affect the growth of plants in a region. At the same time, the number of plants in a community also has an effect on the growth of a particular plant in the community. Therefore, we can establish a mathematical model of plant growth under different environmental conditions by considering factors affecting plant growth rate and plant quantity in the process of plant growth, such as precipitation, number of plants in the community, pollution degree, etc. According to this mathematical model, we can replace the actual situation of a certain region as the parameter. A complete plant growth model of the region can be obtained and programmed algorithms such as genetic algorithms can be used to predict and optimize plant growth conditions.

2. Changes of plant growth rate over time

2.1. Influences of precipitation factors on plants

2.1.1 Logistic model

When a species moves into a new ecosystem, its population changes. Assuming that the initial population of the species is less than the maximum capacity of the environment, the population will grow. There are two types of growth:

J-type growth: N(t)=n(p^t) if the species has no natural predators in the ecosystem and resources such as food space are abundant (ideal environment). Where,N(t) is the population number in year t,t is the time, and p is the annual growth rate (greater than 1). The shape of the image is J-shaped. Generally, J-shaped growth of the image is an ideal situation.

S-shaped growth: If the species has natural enemies in the ecosystem, food space and other resources are not sufficient (not ideal environment), then the growth function satisfies the logical Sith equation. This equation is the best mathematical model to describe the law of population growth under the condition of limited resources.

Logistic equation is differential equation:

$$\frac{dN}{dt} = r\frac{N(K-N)}{K} \tag{1}$$

Letter meaning: In the formula, N is the total number of population individuals, t is the time, r is the population growth potential index, and K is the maximum environmental capacity. Significance: When a species moves into a new ecosystem, its population changes. Assuming that the initial population of the species is less than the maximum capacity of the environment, the population will grow. In this ecosystem, the species has natural enemies, food, space and other resources are insufficient (not ideal environment), then the growth function satisfies the logistic equation, the image is S-shaped, this equation is the best mathematical model to describe the law of population growth under the condition of limited resources. [1]

2.1.2 Influence of precipitation on plant growth

Precipitation is an important factor affecting plant growth, and the precipitation in a region can affect the growth rate of plants. According to the relationship between plant growth and precipitation in grassland and woodland areas in northern China[2], we found that without considering the flood disaster caused by excessive precipitation and drowning plants, with the increase of precipitation, The plant growth rate increases, but when the precipitation exceeds a certain value, if the precipitation continues to increase, the influence of precipitation on the plant growth rate gradually approaches a certain value[3]. Therefore, we conclude that there is a logistic relationship between plant growth rate γ and precipitation.

2.2. Relationship between the number of plants and biomass in the community

We create a community competition concept after taking environmental considerations into account. The Logistic rule governs most natural population evolution under constrained environmental constraints. Assume that two populations that follow this law are in a competitive relationship. They slow down each other's population growth when they are in competition with one another. We may create a differential equation model of a community with two species based on this supposition[4]:

$$\frac{dx_1}{dt} = r_1 x_1\left(1 - \frac{x_1}{N_1} - \alpha_{12}\frac{x_2}{N_2} - \cdots - \alpha_{1n}\frac{x_n}{N_n}\right) \tag{2}$$

$$\frac{dx_2}{dt} = r_2 x_2\left(1 - \alpha_{21}\frac{x_1}{N_1} - \frac{x_2}{N_2} - \cdots - \alpha_{2n}\frac{x_n}{N_n}\right) \tag{3}$$

Through the above individual calculation formula, we can summarize a universal formula

$$\frac{dx_n}{dt} = r_n x_n\left(1 - \alpha_{n1}\frac{x_1}{N_1} - \alpha_{n2}\frac{x_2}{N_2} - \cdots - \frac{x_n}{N_n}\right) \tag{4}$$

On the other hand, we can expand the equation to the differential equation with i species based on this foundation:

$$\frac{dx_i(t)}{dt} = r_i(t)x_i(t)\left(1 - \frac{x_i(t)}{N_i} - \sum_{k\neq i}\alpha_k^i\frac{x_k(t)}{N_k}\right) \tag{5}$$

2.2.1 Improved differential equation combined with biomass

When analyzing how many different plant species are needed to benefit from the community, we can first confirm the contribution of different species to the community biomass, and establish an optimization model of the number of species and community biomass, and finally solve the problem with genetic algorithm.

So let's say we're looking at a plant community, and based on that, we can establish that there are n different species of plants in different species, and their biomass is. Thus, for commercial and

agricultural growing areas that need to maximize biomass, we can establish optimal models that combine the number of species with the total community biomass[5]. The objective function of this model is the maximum total biomass of the plant community in the long-term state, namely:

$$objfun = \lambda_1 * \sum_{i=1}^{\xi_1} x_i(t) + \lambda_2 * \sum_{i=\xi_1+1}^{\xi_1+1+\xi_2} x_i(t) + \ldots + \lambda_n * \sum_{i=\xi_1+1+\xi_2+1+\ldots+\xi_{n-1}}^{\xi_1+1+\xi_2+1+\ldots+\xi_{n-1}+\xi_n} x_i(t) \quad (6)$$

Since there is an upper limit of plant species, we assume that there are at most *K* plants in this community. In addition, since the initial succession period of the community is too long, we may assume that each plant has an initial value when t=0, so the constraint conditions of this optimization model are:

$$\xi_1 + \xi_2 + \xi_3 \leq K \quad (7)$$

Where, the governing equation, namely the equation of the transformation of the number of plants over time, is:

$$\frac{dx_i(t)}{dt} = r_i(t)x_i(t)\left(1 - \frac{x_i(t)}{N_i} - \sum_{k \neq i} \alpha_k^i \frac{x_k(t)}{N_k}\right) \quad (8)$$

2.2.2 Genetic Algorithm for Optimal Solution

Genetic algorithm (GA) is an optimization algorithm inspired by natural selection. It is a population-based search algorithm that exploits the concept of survival of the fittest. New populations are generated by repeated use of genetic operators on individuals present in the population. Chromosome representation, selection, crossover, mutation and fitness function calculation are the keys of genetic algorithm. The process of genetic algorithm is as follows. The population (Y) of n chromosomes is randomly initialized. Calculate the fitness of each chromosome in Y. The two chromosomes represent the selection of C1 and C2 from population Y based on fitness values.

Use a single point crossover operator with crossover probability (Cp) for C1 and C2 to produce a child O. A consistent mutation operator with mutation probability (Mp) is then used for the resulting offspring O to produce O '. The new offspring O 'are placed in the new population. Repeat selection, crossover, and mutation operations on the current population until the new population is complete.

The mathematical analysis of genetic algorithm is as follows: Genetic algorithm dynamically changes the search process through crossover probability and mutation probability to achieve the optimal solution. Genetic algorithms can modify the encoded genes. Genetic algorithms can evaluate multiple individuals and produce multiple optimal solutions. Therefore, GA has better global search capability. Offspring produced by chromosome cross parents are likely to abolish the fine genetic model parent chromosomes, cross formula is defined as the : R fell G lay down 2 G p ſ t = 3 G ð 1 Þ st, including G is algebra, G is the total number of evolutionary generation set by population. It can be seen from Equation (1) that R changes dynamically and increases with the increase of evolutionary algebra. In the initial phase of genetic algorithm, the similarity between individuals is very low. R-values should be low to ensure that the new population does not disrupt the good genetic schema of the individual. At the end of evolution, the similarity between individuals is very high, so should the R-value.[6]

According to the pattern theorem, the original pattern must be replaced by the modified pattern. In order to maintain the diversity of the population, the new model preserves the initial state of the population early in evolution. At the end of evolution, appropriate patterns are produced to prevent the excellent genetic patterns from being distorted. Algorithm 1 is the pseudocode of classical genetic algorithm. The pseudo-code of genetic algorithm is shown in the Table 1:

Table 1: Principle of genetic algorithm

GA

Input: maximum of biomass
Output: the number of herbs, the number of shrubs, and the number of trees.
BEGIN
t = 0 ; % Genetic algebra
Initialization P(t) ;
Calculate the fitness value of P(t) ; % Initialize a population or chromosome
while (the stop criterion is not met) do
begin
t = t + 1 ;
Choose P(t) from P(t-1) ; % Choose
Recombine P(t) ;
Calculate the fitness value of P(t) ; % Crossover and variation
END

2.3. Impacts of habitat pollution factors on plant growth

2.3.1 Changes of plant growth rate caused by pollution

First, on the question of pollution, we investigated the effects of soil pollution and air pollution on plant growth in temperate regions [7]. The survey results are shown in the table. We found that in the case of mild pollution, the growth rate of plants increased slightly, but as the pollution level increased, the growth rate of plants began to decrease.[8]

Based on the above survey results, the pollution degree was taken as the independent variable and the growth rate as the dependent variable for regression analysis. We believe that the function expressions of plant growth rate γ and pollution degree percentage μ should conform to the following forms:

$$r = \frac{a(\mu + c)}{(\mu + c)^2 + b} + d \tag{9}$$

2.3.2 Maximum likelihood estimation method to determine parameters

Maximum likelihood estimation is a statistical method which is used to find the parameters of the relative probability density function of a sample set. The method was first used by geneticist and statistician Sir Ronald Fisher between 1912 and 1922.

The maximum likelihood method explicitly uses a probabilistic model whose goal is to find phylogenetic trees that can produce observed data with high probability. Maximum likelihood method is a representative of a class of phylogenetic tree reconstruction methods based entirely on statistics. This method takes into account the probability of each nucleotide substitution in each sequence alignment.

For example, a conversion is about three times more likely to occur than a transpose. In a three-sequence alignment, if one of them is found to have a C, a T, and a G, it is reasonable to assume that the relationship between the C and T sequences is likely to be closer. The calculation of probability is complicated by the fact that the common ancestor sequence of the studied sequence is unknown; Since multiple substitutions may occur at one or more sites, and not all sites are independent of each other, the probability calculation is further complicated. Nevertheless, objective criteria can be used to calculate the probability of each locus, the probability of each possible tree representing the sequence relationship. Then, by definition, the tree with the largest sum of probabilities is most likely to be the phylogenetic tree that reflects the true situation.[9]

The principle of maximum likelihood estimation is as follows. Given a probability distribution D, assuming that its probability density function (continuous distribution) or probability aggregation

function (discrete distribution) is fD, and a distribution parameter θ, we can extract from this distribution a sample X1,X2, with n values... ,Xn, by using fD, we can calculate its probability:

$$P = (x_1, x_2, \dots, x_n) = f_D(x_1, x_2, \dots, x_n|\theta) \tag{10}$$

However, we may not know the value of theta, even though we know that the sampled data comes from the distribution D. So how can we estimate theta? A natural idea is to extract from this distribution a sample of n values X1,X2... ,Xn, and then use this sampled data to estimate θ [10].

Once we get it, we can find an estimate of theta from it. The maximum likelihood estimate looks for the most likely value about θ (that is, out of all possible θ values, look for a value that maximizes the "likelihood" of this sample). This approach is exactly different from other estimation methods, such as θ unbiased estimates, which do not necessarily output a most likely value, but rather a θ value that is neither overestimated nor underestimated.

To implement maximum likelihood estimation mathematically, we first define probability:

$$lik(\theta) = f_D(x_1, x_2, \dots, x_n|\theta) \tag{11}$$

And maximizes this function at all values of theta. The value that maximizes probability is known as θ's maximum likelihood estimate.

After investigation, and through the pollution rate and growth rate calculation, we can get the following results:

$$r = \frac{0.464(\mu + c)}{(\mu + 0.06)^2 + 0.103} + 0.461 \tag{12}$$

3. Conclusion

3.1. Relationship model between plant growth rate and precipitation

We have been able to confirm that the growth rate and precipitation of plants conform to the basic logistic model. It can be assumed that when precipitation is 0, the growth rate of plants is also 0. Therefore, we can obtain a logistic function that the growth rate is also 0 when precipitation is 0:

$$r_i(t) = \frac{\ell_i}{1 + e^{-kh(t)}} - \frac{\ell_i}{2} \tag{13}$$

Where is the maximum growth rate of species, k represents the slope of logistic curve, which reflects the relative drought resistance of plants in a certain area and the relationship between drought resistance of different plants. r is precipitation.

3.2. Growth model of a species in a given environment

According to the growth ability of plants under the interaction of different organisms in a certain environment obtained above, we can combine the growth rate model to determine the number of certain plants in the community:

$$\frac{dx_i(t)}{dt} = \left(\frac{\ell_i}{1 + e^{-kh(t)}} - \frac{\ell_i}{2}\right) x_i(t) \left(1 - \frac{x_i(t)}{N_i} - \sum_{k \neq i} \alpha_k^i \frac{x_k(t)}{N_k}\right) \tag{14}$$

Finally, after determining the changes in the number of plants under normal environmental conditions, we combined the above formula with the influence of pollution on plants, and it can be concluded that:

$$\frac{dx_i(t)}{dt} = \left(\frac{\ell_i}{1+e^{-kh(t)}} - \frac{\ell_i}{2}\right) x_i(t) \left(1 - \frac{x_i(t)}{N_i} - \sum_{k \neq i} \alpha_k^i \frac{x_k(t)}{N_k}\right) \left[r = \frac{0.464(\mu + c)}{(\mu + 0.06)^2 + 0.103} + 0.461 \right] \tag{15}$$

3.3. Model of plant biomass

Above, we have obtained the optimization objectives and constraints of the biomass of plants in a community and the number of different species of plants. Based on these, we can obtain the relationship between the biomass of a community and the number of different species of plants, as shown below:

$$objfun = \lambda_1 * \sum_{i=1}^{\xi_1} x_i(t) + \lambda_2 * \sum_{i=\xi_1+1}^{\xi_1+1+\xi_2} x_i(t) + \cdots + \lambda_n * \sum_{i=\xi_1+1+\xi_2+1+\cdots+\xi_{n-1}}^{\xi_1+1+\xi_2+1+\cdots+\xi_{n-1}+\xi_n} x_i(t) \quad (16)$$

In order to facilitate the use of genetic algorithm to solve the model, we first standardized the model, that is, reversed the objective function, so as to transform it into a problem of solving the maximum value. At the same time, according to the actual situation of the natural environment, we judged that if the data did not meet the constraint conditions in the iteration process, the results should produce a large deviation, so the penalty function method was used to deal with the constraint conditions. Since we need to obtain the minimum value of the objective function, when the decision variable does not meet the constraint conditions, a penalty factor δ will be added to the objective function, where δ is a very large negative number, so the value that does not meet the constraint conditions will be eliminated

At this point, we can get the general model formula as follows:

$$objfun = \lambda_1 * \sum_{i=1}^{\xi_1} x_i(t) + \lambda_2 * \sum_{i=\xi_1+1}^{\xi_1+1+\xi_2} x_i(t) + \ldots + \lambda_n * \sum_{i=\xi_1+1+\xi_2+1+\ldots+\xi_{n-1}}^{\xi_1+1+\xi_2+1+\ldots+\xi_{n-1}+\xi_n} x_i(t) + \begin{cases} 0, \xi_1 + \xi_2 + \xi_3 \le K \\ \delta, \xi_1 + \xi_2 + \xi_3 \ge K \end{cases} \quad (17)$$

References

[1] F. H. HSU, C. J. NELSON, W. S. CHOW, A Mathematical Model to Utilize the Logistic Function in Germination and Seedling Growth, Journal of Experimental Botany, Volume 35, Issue 11, November 1984, Pages 1629–1640, https://doi.org/10.1093/jxb/35.11.1629

[2] Holdrege, M. C., Beard, K. H., and Kulmatiski, A.. 2021. Woody plant growth increases with precipitation intensity in a cold semiarid system. Ecology 102(1):e03212. 10.1002/ecy.3212

[3] Xu, Z., Shimizu, H., Ito, S. et al. Effects of elevated CO2, warming and precipitation change on plant growth, photosynthesis and peroxidation in dominant species from North China grassland. Planta 239, 421–435 (2014).

[4] Carlos L Ballaré, Amy T Austin, Recalculating growth and defense strategies under competition: key roles of photoreceptors and jasmonates, Journal of Experimental Botany, Volume 70, Issue 13, 15 June 2019, Pages 3425–3434,

[5] Chen, Y.; Liu, Z.; Rao, X.; Wang, X.; Liang, C.; Lin, Y.; Zhou, L.; Cai, X.-a.; Fu, S. Carbon Storage and Allocation Pattern in Plant Biomass among Different Forest Plantation Stands in Guangdong, China. Forests 2015, 6, 794-808. https://doi.org/10.3390/f6030794

[6] Katoch, S., Chauhan, S.S. & Kumar, V. A review on genetic algorithm: past, present, and future. Multimed Tools Appl 80, 8091–8126 (2021).

[7] Shahid Iqbal, Jianchu Xu, Schaefer Douglas Allen, Sehroon Khan, Sadia Nadir, Muhammad Saleem Arif, Tahira Yasmeen,Unraveling consequences of soil micro- and nano-plastic pollution on soil-plant system: Implications for nitrogen (N) cycling and soil microbial activity,Chemosphere,Volume 260,2020,127578,ISSN 0045-6535,https://doi.org/10.1016/j.chemosphere.2020.127578.

[8] Yunus, M, and Iqbal, M. Plant response to air pollution. United Kingdom: N. p., 1996. Web.

[9] In Jae Myung,Tutorial on maximum likelihood estimation,Journal of Mathematical Psychology,Volume 47, Issue 1,2003,Pages 90-100,ISSN 0022-2496,https://doi.org/10.1016/S0022-2496(02)00028-7.

[10] White, Halbert. “Maximum Likelihood Estimation of Misspecified Models.” Econometrica, vol. 50, no. 1, 1982, pp. 1–25. JSTOR, https://doi.org/10.2307/1912526. Accessed 8 Apr. 2023.