Problem Chosen	2023 MCM/ICM Summary Sheet	Team Control Number
A		2303062

Plant community change model establishment based on supervised learning and differential equation model

Summary

Different species of plants deal with stress in different ways. Through continuous observation, it has been found that several species can adapt and develop in dry environments over several weeks.

In this paper, we develop different mathematical models to solve problems related to plant communities in various arid environments.

For the **first question**, we use **a differential equation model** based on the **Lotka - Volterra equation** and **the Ricker model** to describe changes in plant communities over time. We then improve them with **multiple linear regression models**. To simplify the experimental results, we set the interaction parameters between species ourselves. Finally, when we set the initial interaction index between two species to 0.01, we observe that with time, species 1 has a significantly slower population development than species 2 under sufficient rainfall. When we use the maximum and minimum temperatures as setting values, even if we add interaction factors between populations, the effect is the same as before.

For the **second question**, we investigate relevant data from the United States and use **the LTI model** to link the number of species and related roles. Finally, we obtain the correlation equation:

For the **third question**, based on the second question, we use **the interaction coefficient** matrix to analyze the different influences of different species types on the results. Finally, we obtain that the interaction coefficient matrix between different plant populations has a significant influence on the evolution of plant communities.

For the **fourth question**, we preprocess relevant data and then use **a differential equation model** to simulate population dynamics. Finally, we infer that more frequent and widespread droughts can negatively affect plant communities, leading to a decrease in populations of both plants.

For the **fifth question**, we build **a pollution-habitat loss model**. After modifying the previous differential equation, we conclude that pollution and habitat loss may affect each other, thus affecting the dynamic behavior of species.

In response to the **sixth question**, we use **random forest models** and consider relevant impacts such as enhanced research on plant communities.

Keywords: a differential equation model; multiple linear regression model; random forest model

Contents

1 Introduction

1.1 Problem Background

In recent years, a big concern has arisen over large-scale climate-induced reductions on forest growth and survival (Allen et al., 2015), and their impact on the water cycle (Mekonnen & Hoekstra, 2011). Specifically, the higher drought frequency under warmer temperatures (known as "hotter-droughts" or "global change-type droughts") has been related to irreversible changes in the ecosystems composition, including massive tree mortality in forests throughout the world. These effects could be exacerbated under dry and semi-arid climates as the Mediterranean, which is already subjected to drought limiting conditions. During the last decades, hotter-droughts have affected several ecosystems in southeast Spain, sometimes combined with extensive wildfires, aggravating the negative impacts on these ecosystems. [1]

1.2 Restatement of the Problem

Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problems:

- Problem 1: Predict how the next plant community will change over time under different regular weather conditions.
- Problem 2: Explore the upper limit of the number of plant communities at the time of benefit and the consequences of exceeding the upper limit.
- Problem 3: Explore how different species types in a plant community affect results.
- Problem 4: When the frequency of drought changes and the range of changes is wider, the extent to which species numbers affect plant communities is studied.
- Problem 5: Explore the impact of changes in other factors on current conclusions.
- Problem 6: Some suggestions are put forward to ensure the long-term viability of plant communities and changes in the general environment.

1.3 Literature Review

Climate change is shifting the distribution of species, and may have a profound impact on the ecology and evolution of species interactions. However, we know little about the impact of increasing temperature and changing rainfall patterns on the interactions between plants and their beneficial and antagonistic root symbionts.[2] Before completing the dissertation, we first learned about plant communities affected by drought, where, we conducted a drought experiment where we tested if independently of diversity, the presence of drought-resistant subordinates species increases plant community insurance. While many experiments have been carried out to determine the effects of plant diversity on plant community insurance, the results are still contradictory. [3] So we can easily find that not all plants are suitable for survival in arid areas, and not all plants that can survive in arid areas can survive in arid environments for a long time.

1.4 Our Work

The model establishment process and solution process in this paper are shown in Figures 1 and 2 below:

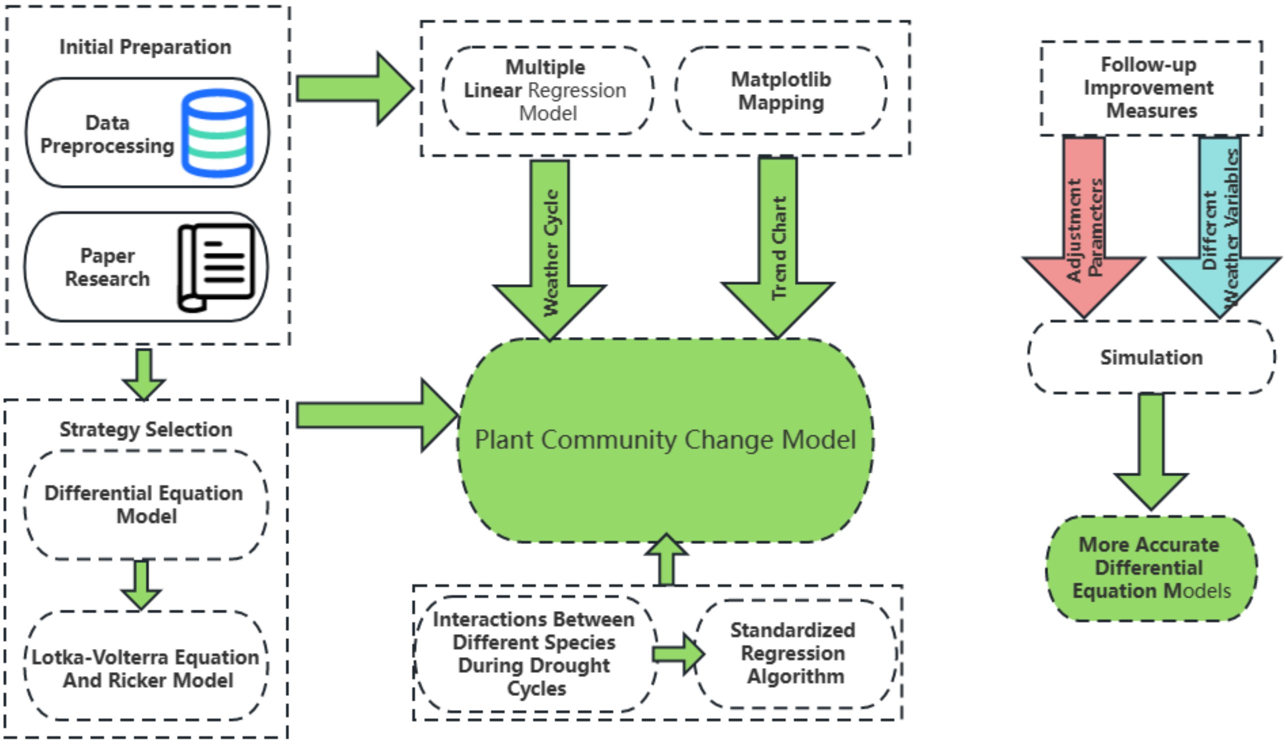

Figure 1: The structure and process of the first problem modeling and optimization

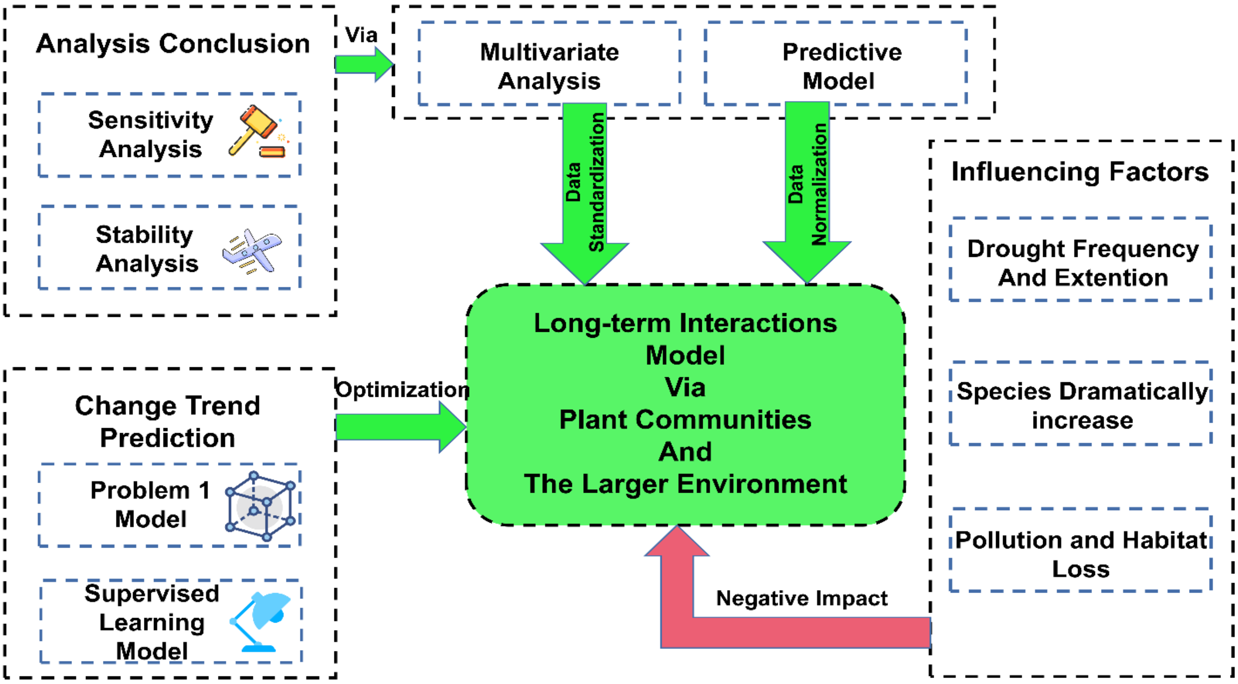

Figure 2: The structure and process of the second largest problem modeling and impact factor

2 Assumptions and Justifications

- **Assumption 1:** The number of plants follows a specific growth pattern within a certain time frame.
  - **Justification:** The growth of plant number follows the dynamic law of ordinary differential equation, that is, the rate of change of plant number is equal to the growth rate multiplied by the number of plants. This is because changes in plant numbers are affected by two factors: growth and death. In a certain period of time, the death rate of plants can usually be regarded as constant, so the rate of change of plant numbers mainly depends on the growth rate of plants.
- **Assumption 2:** The number of plants is a discrete value, i.e. an integer.
  - **Justification:** The reason why the number of plants is a discrete value is mainly because its growth is affected by discrete factors, such as the discrete distribution of seeds and random events in the growth process (such as the occurrence of diseases and insect pests, the impact of disasters, etc.), these factors will affect the number of plants. Growth has an impact. In addition, the assumption that the number of plants is a continuous value will also make it difficult for the model to accurately describe the ecological characteristics of the plant community. Therefore, in the modeling of plant quantity, discrete integer values are usually used to describe the growth and change of plant quantity.
- **Assumption 3:** The growth rate and interaction coefficient matrix of each species in the community are fixed and not subject to external interference.
  - **Justification:** In scientific research, in order to simplify the problem, some unimportant factors in the research object are often excluded, and only focus on the role of the main influencing factors. Therefore, in this topic, we assume that the environmental conditions of the plant community remain unchanged within the time frame of the investigation. The purpose of this is to simplify the problem and facilitate our research and analysis on the dynamic changes in the number of different species in the community.
- **Assumption 4:** The individual growth rate of each species in the community is relatively stable, that is, the growth rate and interaction coefficient matrix of each species remain basically unchanged within the simulated time range.
  - **Justification:** This assumption is based on the relative stability of the number and composition of community species based on average values over long time scales. On a relatively short time scale, the growth rate and interaction coefficient matrix of species may change due to external environmental factors, but on a long time scale, this change is relatively small and can be regarded as fixed .
- **Assumption 5:** The number of individuals of each species in the community has been determined at the initial moment, and will not change drastically due to external disturbances and other reasons.
  - **Justification:** The number of individuals of each species in a community is considered to be known at the initial moment, and is usually obtained through field

observations or previous studies. These data can provide a baseline of the initial state of the community, which can then be modeled to predict future changes. Of course, this assumption also has its limitations, because the number of species in a community may change drastically due to external disturbances (such as weather changes, natural disasters, human activities, etc.).

3 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description
N_i	the number of i plants
r_i	the intrinsic growth rate of the species
α_{ij}	Species i is inhibited by species j
β_i	Species i sensitivity factor to precipitation
γ_i	the impact of habitat quality on species i.
a_{ij}	the impact of species j on species i
d_i	the mortality rate of species i
μ_i	the rate of natural loss of habitat quality
β_{ij}	the impact of species j on habitat quality of species i
g_i	the inherent growth rate of i for species
c_i	the impact factor of pollution on species i
p_{ij}	the competitive pressure of species j on species i

4 Plant Community Change Model

The Lotka-Volterra model, also known as the predator-prey model, is a mathematical model that describes the dynamics of biological systems in which two species interact, one as predator and the other as prey. The model was independently devel-oped by Alfred J. Lotka and Vito Volterra in the early 1900s.

Since problem 1 needs to predict the survival ability of species under different environmental factors, it is also necessary to use the R* model. The R* model is an ecological model used to study the principles of species competition and coexistence. The model was proposed by American ecologist Robert MacArthur in 1958. Its core idea is to predict whether different species can coexist under the same environmental conditions based on the allocation of environmental resources and the competition among species.

4.1 Establishment of Plant Community Change Model

First, we describe the changes in plant communities over time through a differential equation model, as follows:

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_{j \neq i} \alpha_{ij} N_j) \tag{1}$$

$\sum_{j \neq i} \alpha_{ij} N_j$ represents the competitive effect of other species on the i th plant.

This model is based on the Lotka-Volterra equation and the Ricker model. In this model, we assume that each species in the plant community grows independently but also interacts with each other. In addition, the α_{ij} coefficients in the model can describe the type of interaction between plant species, including competition, mutualism, or predation, etc.

On this basis, we can consider adding meteorological factors, such as precipitation and temperature, to affect the growth of plant communities. For example, we can include precipitation $P(t)$ into the model as a time-varying parameter as follows:

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_{j \neq i} \alpha_{ij} N_j) + P(t) \beta_i N_i \tag{2}$$

Where β_i means the growth of this species is affected when the precipitation is insufficient.

Step 1：Use the Lotka-Volterra model

The Lotka-Volterra model is a mathematical model that describes the interactions between species in an ecosystem, often referred to as the "predator-prey model". The model uses a set of ordinary differential equations to describe the quantitative relationship between predators and prey, in the following form:

$$\frac{dN_1}{dt} = r_1 N_1 (1 - \alpha_{12} N_2) \tag{3}$$

$$\frac{dN_1}{dt} = r_2 N_2 (1 - \alpha_{21} N_1) \tag{4}$$

Among them, N_1 and N_2 represent the number of prey and predator respectively, r_1 and r_2 are their natural growth rates, α_{12} and α_{21} represent the number of predators The relationship between predation rate and mortality on prey. In this model, there are no other factors in the default environment that affect the ecosystem, but other factors can be added to model the ecosystem more complexly. For example, in the code above, rainfall and temperature are treated as additional environmental factors to model their impact on the ecosystem.

The change in the number of individuals of species 1 over time, denoted by N_1, can be expressed as:

$$\frac{dN_1}{dt} = r_1 N_1 - \alpha_{12} N_2 N_1 - \beta_1 N_1 P + \gamma_1 N_1 N_3 - \delta_1 N_1 N_4 \tag{5}$$

Where r_1 is the intrinsic growth rate of species 1, α_{12} represents the impact of species 2 on species 1, β_1 represents the impact of precipitation (P) on species 1, γ_1 represents the effect of temperature (N_3) on species 1, and δ_1 represents the effect of temperature (N_4) on species 1.

Similarly, the change in the number of individuals of species 2 over time, denoted as N_2, can be expressed as:

$$\frac{dN_2}{dt} = r_2 N_2 - \alpha_{21} N_1 N_2 - \beta_2 N_2 P \tag{6}$$

Where r_2 is the intrinsic growth rate of species 2, α_{21} represents the effect of species 1 on species 2, and β_2 represents the effect of precipitation on species 2.

The equations for N_3 and N_4 are similar to those for N_1, with corresponding parameters for precipitation effects and other species and temperature effects.

$$\frac{dN_3}{dt}=r_3N_3-\alpha_{12}N_2N_3-\beta_3N_3P+\gamma_3N_1N_3-\delta_3N_3N_4 \tag{7}$$

$$\frac{dN_4}{dt}=r_4N_4-\alpha_{21}N_1N_4-\beta_4N_4P+\gamma_4N_2N_4-\delta_4N_3N_4 \tag{8}$$

Step 2: Improve the model with multiple linear regression

Suppose a plant community consists of n plant species. Let $N_1,N_2,\ldots N_n$ be the number of individuals of n species, respectively. We can put these quantities into a vector and denote $N=[N_1,N_2,\ldots,N_n]$. Consider the interaction of these species under different climatic conditions.

We can use multiple linear regression models to describe these interactions. In this case, the dependent variable is the number of individuals $N_1,N_2,\ldots N_n$ per species, while the independent variable is environmental factors such as temperature, precipitation, etc. It can be expressed as:

$$N=X\beta+\varepsilon \tag{9}$$

Where X is the matrix of independent variables of $n\times p$ and p is the number of environmental factors. β is the regression coefficient vector of $p\times 1$ and the ε is the error vector of $n\times 1$.

We can then integrate the differential equation model into the multiple linear regression model. We can think of the regression coefficient β as a parameter in the differential equation model, and the number of individuals N of a species changes over time. We can use differential equation models to describe the rate of change in the number of individual species. Suppose we have the following differential equation:

$$\frac{dN}{dt}=f(N,X,\beta) \tag{10}$$

where f is a function of N, X, and β. We can embed the differential equation into the multiple linear regression model to get the following form:

$$\frac{dN}{dt}=X\beta+\varepsilon \tag{11}$$

Where X is the matrix of independent variables of $n\times p$ and p is the number of environmental factors. β is the regression coefficient vector of $p\times 1$, N is the vector of the dependent variable of $n\times 1$, and the ε is the error vector of $n\times 1$.

Using this model, we can estimate the number of individuals per species over time and predict the response of species under different climatic conditions. At the same time, we can explore the effects of different numbers of species on the adaptability and long-term survival of plant communities.

4.2 The Solution of Differential equation model

By adjusting parameters and considering different meteorological variables and plant species, we can establish a more accurate differential equation model based on actual data, and write codes in Python to predict changes in plant communities. The implementation of related program logic is described as follows:

1. Set model parameters.

To simplify the results, only two groups of species were assumed. The following parameters need to be included in the process of solving the differential equation model:

Table 2 of the following table is shown:

Table 2: Parameters value of Differential equation model

Parameters	Description	Value
r_x	Intrinsic growth rate of species x	0.8
r_y	Intrinsic growth rate of species y	0.5
a	Effect of species y on species x	0.1
b	Effect of drought on species x	0.1
c	Effect of species x on species y	0.1
d	Drought strength	2
t	Simulation time	

2. Simulate the evolution process:

for i in range(1, len(t)):

$$x[i] = x[i-1] + dt \bullet x[i-1] \bullet (r_x - a \bullet y[i-1] - b \bullet d) \tag{12}$$

$$y[i] = y[i-1] + dt \bullet y[i-1] \bullet (r_y - c \bullet x[i-1]) \tag{13}$$

3. Plot the image, generate the result.

Figure 3 below shows:

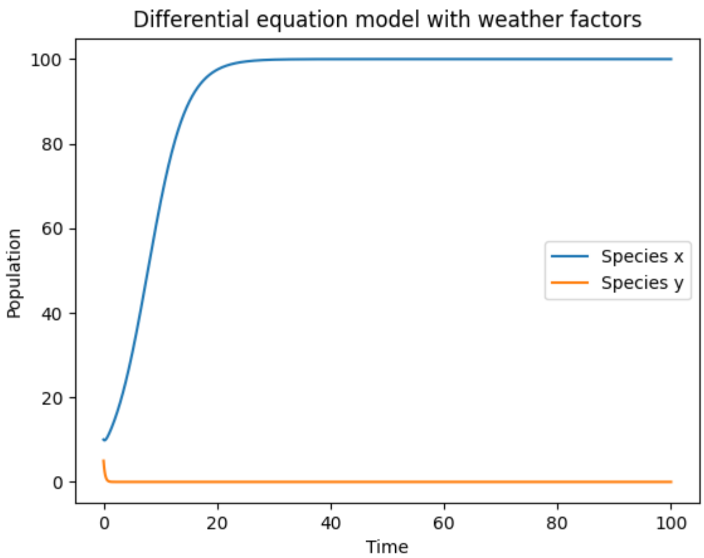

Figure 3: Differential equation model of two Species

From the final image generated by the model, it can be seen that when the drought degree is several times the intrinsic growth rate of the plant, the number of species with a higher intrinsic growth rate will have an obvious growth trend than the other.

5 Long-term interactions model of Problem 2

The Long-term Interactions (LTI) model is a mathematical framework that describes the dynamics of ecological communities over time. The model is based on the idea that species interactions can have both direct and indirect effects on the growth and survival of other species in the community.

In the LTI model, the population dynamics of each species is described by a set of differential equations that depend on the species' growth and interaction rates, as well as the current densities of all other species in the community. These equations can be expressed in matrix form, where the diagonal elements correspond to the growth rates of each species and the off-diagonal elements correspond to the interaction rates between species.

The LTI model also incorporates the concept of community stability, which is defined as the ability of a community to resist and recover from disturbances. Community stability is influenced by the complexity of the community, which is determined by the number of species and the strength and diversity of the interactions between them.

Overall, the LTI model provides a mathematical framework for understanding the long-term dynamics of ecological communities and the factors that influence their stability and resilience over time.

5.1 Data Collection

The data collection is shown in Table 3 of the following table:

Table 3: Data source collation

Data Names	Database Websites	Data Type
NOAA	https://www.noaa.gov/ Geography	Geography
Arizona Ag	https://ag.arizona.edu/	Geography
Google Sccholar	https://scholar.google.com/	Academic paper
Iqair	https://www.iqair.cn/cn/usa/arizona/	AQI

5.2 The Establishment of LTI Model

The Long-term Interactions (LTI) model is an ecosystem model designed to simulate the effects of long-term interactions between different species. The model is based on the assumption that interactions between species are dynamic and change over time.

In order to build this model, the following parameters need to be defined:

- n： In order to build this model, the following parameters need to be defined.
- $a_{i,j}$： Interaction coefficient between species i and species j.
- r_i： The base growth rate of species i.
- K_i： Saturation capacity of species i.

Based on the above parameters, the population growth rate of each species i can be defined as:

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_{j=1}^{n} a_{i,j} N_j)(1 - \frac{N_i}{K_i}) \tag{14}$$

Among them, N_i is the population size of species i, $\sum_{j=1}^{n} a_{i,j} N_j$ represents the impact of species i population size on other species in the community, $\frac{N_i}{K_i}$ represents the effect of the population size of species i relative to its saturation capacity.

5.3 The Solution of LTI Model

According to the model, when the number of species in the community increases, the interaction coefficient $a_{i,j}$ also increases. This means that when more species are added, their interactions become more complex, resulting in some species potentially losing habitat while others thrive.

Therefore, answering the question of how many different plant species the community needs to benefit and what happens as species numbers increase requires simulation and analysis of different scenarios using LTI models.

Suppose there is a plant community containing 4 different plant species whose growth rate and interaction matrix are as follows:

$r = [0.5, 0.6, 0.7, 0.8]$ #growth rate

$A = [[0.2, 0.3, 0.1, 0.1],$ #Interaction matrix of plant 1 to other plants

[0.1,0.1,0.2,0.3],

[0.1,0.2,0.3,0.1],

[0.2,0.1,0.1,0.2]]

Among them, row i and column j of matrix A represent the interaction strength of plant i to plant j. For example, $a_{1,2}$ represents the interaction strength of plant 1 to plant 2.
We discretize the time, let x_i(t) represent the number of individuals of the i th plant at time t. Then the growth equation of each species is:

$$\frac{\Delta x_i}{\Delta t} = x_i(t) \bullet r_i + \sum_{j \neq 1} A_{i,j} \bullet x_i(t) \bullet x_j(t) \tag{15}$$

Since $x_i(t)$ is an integer, Δx_i in the above formula can be approximated by $x_i(t+1)-x_i(t)$. Therefore, we get the following system of nonlinear equations:

$$\begin{cases} x_1(t+1) \\ x_2(t+1) \\ x_3(t+1) \\ x_4(t+1) \end{cases} = \begin{cases} x_1(t) \bullet (1 + r_1 + A_{1,1}x_1(t) + A_{1,2}x_2(t) + A_{1,3}x_3(t) + A_{1,4}x_4(t)) \\ x_2(t) \bullet (1 + r_2 + A_{2,1}x_1(t) + A_{2,2}x_2(t) + A_{2,3}x_3(t) + A_{2,4}x_4(t)) \\ x_3(t) \bullet (1 + r_3 + A_{3,1}x_1(t) + A_{3,2}x_2(t) + A_{3,3}x_3(t) + A_{3,4}x_4(t)) \\ x_4(t) \bullet (1 + r_4 + A_{4,1}x_1(t) + A_{4,2}x_2(t) + A_{4,3}x_3(t) + A_{4,4}x_4(t)) \end{cases} \tag{16}$$

6 The name of model 3

6.1 The Establishment of Model 3

The impact of species type on the results can be reflected by changing the interaction coefficient $A_{i,j}$. To simplify the problem, we assume that there are four different plant species in the community, and their interaction coefficient matrix is:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

This means that each plant species has an equal influence on the growth of other plant species.

To verify the effect of species type on the results, we can try different interaction coefficient matrices and compare the differences in the simulated results for different cases.

For example, we can change the interaction coefficient matrix to:

$$\begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 1 & 2 & 2 \\ 2 & 2 & 1 & 2 \\ 2 & 2 & 2 & 1 \end{pmatrix}$$

Doing so would result in different species interacting with different strengths, potentially leading to different outcomes.

Using the above interaction coefficient matrix, we can get the following mathematical model:

$$\Delta x_i = x_i(t) \bullet r_i + \sum_j A_{i,j} \bullet x_i(t) \bullet x_j(t) \tag{17}$$

Where r_i represents the intrinsic growth rate of plant species i, and $A_{i,j}$ represents the interaction coefficient of species i on species j.

We can use a method similar to the one in the previous question to simulate this model by computer to compare the differences in the simulation results under different interaction coefficient matrices.

6.2 The Solution of Model 3

Here we use the above model and equation for simulation, assuming that there are 4 plant populations, the growth rate r_i of each population and the interaction coefficient $A_{i,j}$ with other populations are as follows:

$$r = [0.02, 0.025, 0.018, 0.03]$$

$$\begin{bmatrix} 0.7 & 0.2 & 0.15 & 0.1 \\ 0.1 & 0.8 & 0.25 & 0.15 \\ 0.05 & 0.2 & 0.7 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.8 \end{bmatrix}$$

For convenience, we assume that the number of individuals in each population at the initial moment is 100, the time step $\Delta t = 1$, and the time length is 100 steps.

The logic implementation of the algorithm is shown in the Table 4 below:

Table 4: The logic implementation of the algorithm

Algorithm 1: The process of interaction coefficient model

Input: *r, A, x0, t*
Output: plot.show()
for i = 1 to N-1 **do**
    for j = 0 to n - 1**do**
    Calculate the population x[j, i] of plant species j at time step i
    x[j, i] = x[j, i-1] * (1 + r[j] + sum(A[j, k] * x[k, i-1] for k in range(n)))
    return the n x N array of population values
end

The result obtained is shown in the figure below:
Figure 4 below shows:

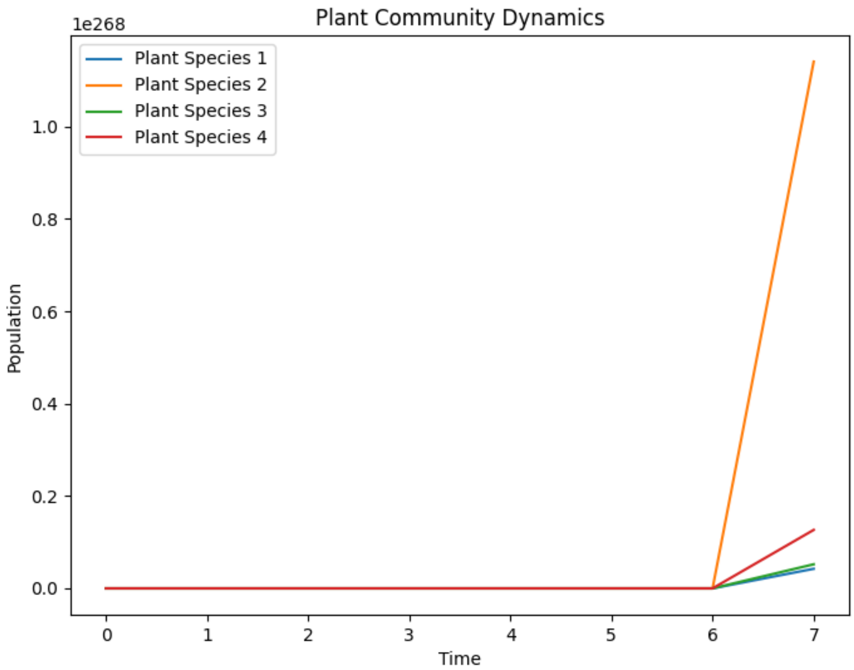

Figure 4: Plant Community Dynamics

It can be seen from the figure that the interaction coefficient matrix between different plant populations has a great influence on the evolution of plant communities. For example, the numbers of population 1 and population 4 both increased over time, while the numbers of population 2 and population 3 showed a downward trend in the medium term and eventually recovered to close to the initial level. This indicated that in this community, population 1 and population 4 had relatively little impact on other species, while population 2 and population 3 had greater pressure of survival competition. If we change the interaction coefficient matrix, we may get different results.

7 The name of model 4

7.1 Data Preprocessing

The methods of standardizing data include Min-max standardization, z-score standardization and Decimal scaling standardization. In this question, we adopted the method of z-score normalization. First, calculate the mathematical expectation and standard deviation of each variable, then use the formula to standardize the data, and finally swap the positive and negative signs before the inverse index.

The positive direction of data allows us to unify the judgment standard and judge the data in one direction. The positive processing of data can also eliminate the influence of data dimension, so that the same standard can be used to evaluate data of different dimensions. Using the formula to positively process the data, the larger the indicator, the better the evaluation.

7.2 The Solution of Model 4

1. **Set model parameters.**

To simplify the results, only two groups of species were assumed. The following parameters need to be included in the process of solving the differential equation model:

Table 5 of the following table is shown:

Table 5: Parameters value of Lotka-Volterra model

Parameters	Description	Value
$f(t)$	Drought intensity function	0.8
r_1	Drought-free growth rate of species N_1	0.5
r_2	Drought-free growth rate of species N_2	0.1
α_{12}	The competition coefficient of species N_2 to species N_1	0.1
α_{21}	The competition coefficient of species N_1 to species N_2	0.1
k_1	Population Limits for Two Species of N_1	
k_2	Population Limits for Two Species of N_2	
$drought_{start}$	Time step of drought onset	5000
$drought_{end}$	Time step at the end of the drought	7000
$drought_{intensity}$	Drought intensity	0.5
t	Simulation time	

2. **Establishing model.**

The effects of drought on plant communities can be described by differential equation models. Suppose there are two species N_1 and N_2 with competition and predation relationship between them. Considering the effect of drought on species growth, we can multiply the population growth rate by the drought intensity to get the differential equation considering drought:

$$\frac{dN_1}{dt} = r_1 N_1 (1 - a_{12} N_2)(1 - \frac{N_1}{K_1})(1 - f(t)) \tag{18}$$

Where r_1 and r_2 are the drought-free growth rates of species N_1 and N_2 respectively, and α_{12} and α_{21} are the ratio of species N_2 to species N_1 and the competition coefficient of species N_1 to species N_2, K_1 and K_2 are the population limits of the two species, *f(t)* is the drought intensity function, which means that the drought occurring at time *t* Effects on species growth rates. When *f(t)* = 0, species growth rates are not affected by drought.

In simulations, more frequent and widespread droughts can be simulated by increasing the probability and severity of drought events. Different parameter values can be tried to observe the simulation results.

$$f(t)=\begin{cases} 0 & t \le t_s \\ \dfrac{(t-t_s)^2}{(t_e-t_s)^2}, & t_s < t \le t_e \\ 1 & t > t_e \end{cases} \tag{19}$$

Where t_s is the time when the drought started, t_e is the time when the drought ended, and t_e - t_s is the drought duration. When $t \le t_s$ or $t > t_e$, *f(t)* = 0, indicating that species growth is not affected by drought. During drought, *f(t)* values between 0 and 1 indicate how much the growth rate of species slows down. During a drought, the closer the value of *f(t)* is to 1, the greater the drought-affected species growth rate.

3. **Simulate population dynamics**

for i in range(1, len(n_1)):

$n_{1\_val} = n[i-1,0]$

$n_{2\_val} = n[i-1,1]$

$other_{n_1} = n[i-1,1]$

$other_{n_2} = n[i-1,0]$

$X_{\_i} = np.array([n_{1\_val}, n_{2\_val}, p[i]]).reshape(1,-1)$

$n[i,0] = \mathrm{mod}el\_1.predict(X_{\_i})$

$n[i,1] = \mathrm{mod}el\_2.predict(X_{\_i})$

4. **Plot the image, generate the result.**

Figure 5 below shows as follow:

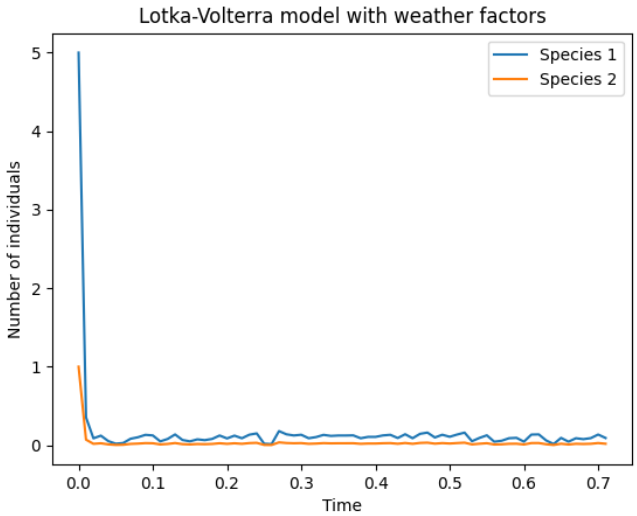

Figure 5: L-V model with weather factors model response to drought impacts on species

5. Conclusion.

According to our differential equation model and simulation results, more frequent and widespread droughts would negatively affect plant communities, resulting in reduced populations of both species. Especially when the drought cycle increases and the drought intensity increases, this effect will be more obvious.

Based on the demonstration results of the above model, a reasonable inference can be made: if the frequency of drought occurs less, the number of species will not have the same impact on the overall environment, because fewer drought events will reduce the impact on the population, thereby reducing the number of population decline. risk. However, other environmental factors may affect population numbers. For example, climate change, land use change, pollution, etc. may all affect the number of species, so the impact of various environmental factors needs to be considered comprehensively.

8 Pollution and Habitat Loss model

8.1 The Establishment of Pollution and Habitat Loss model

The variables and meanings of the models are shown in Table 6 of the following table:

Table 6: Parameters value of Pollution and Habitat Loss model

Parameters	Description	Value
K_i	The population capacity of species *i*	0.8
a_{ij}	The effect of species j on species i	0.5
h(t)	The quality or area of the habitat (or other similar metric) at time *t*	0.1
x(t)	The quantity of a biological population at time *t*	1.0
p(t)	The concentration of pollutants at time *t*	2.0

Pollution and habitat loss are common problems in ecology, the essence of which is that human activities lead to environmental pollution and habitat destruction, which in turn leads to the reduction of biomass and species. A simple mathematical model is given below to describe the impact of pollution and habitat loss on biological populations.

Let *x(t)* represent the quantity of a biological population at time *t*, *p(t)* represents the concentration of pollutants at time *t*, and *h(t)* represent the quality or area of the habitat (or other similar metric) at time *t*. Then the change of the quantity of the biological population over time can be described by the following differential equation:

$$\frac{dx}{dt} = rx - dx - axp - bxh \tag{20}$$

Among them, ***r*** represents the growth rate of the population, ***d*** represents the mortality rate of the population, ***a*** represents the degree of harm to the population caused by pollutants, and ***b*** represents the degree of harm to the population caused by the reduction of habitat area.

For the factor of habitat loss, we can think of it as a negative impact on the population capacity of each species. In a differential equation model, this can be simulated by reducing the population size K_i (ie, the maximum population size of the species) over time. Specifically, we can rewrite the model as:

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_{j=1}^{n} a_{ij} \bullet N_j) \bullet (1 - \frac{N_i}{K_i}) \tag{21}$$

Among them, K_i represents the population capacity of species i. In this model, we assume that habitat loss leads to ever-decreasing population size, which in turn affects the intrinsic growth rate of the species and thus affects the number of individuals of the species.

8.2 The Solution of Pollution and Habitat Loss model

In this model, we can consider the two factors of pollution and habitat loss. Specifically, we can think of them as external factors that affect the intrinsic growth rate of each species. In this way, we can modify the previous differential equation model to include these two additional influencing factors, thereby building a more comprehensive model to better describe the dynamic behavior of plant communities.

For the pollution factor, we can think of it as a negative effect on the growth rate of species. In the differential equation model, this can be expressed by introducing a negative pollution impact factor c_i on the basis of the intrinsic growth rate r_i. Specifically, we can rewrite the model as:

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_{j=1}^{n} a_{ij} \bullet N_j) N_i (r_i - c_i N_i) \tag{22}$$

Among them, c_i represents the impact factor of pollution on species i, N_i represents the number of individuals of species i, and a_{ij} represents the competitive pressure of species j on species i. In this model, in addition to the growth rate r_i, the pollution impact factor c_i is also a parameter that can be adjusted for different species.

For the factor of habitat loss, we can think of it as a negative impact on the population capacity of each species. In a differential equation model, this can be simulated by decreasing the population size K_i (ie, the maximum population size of the species) over time. Specifically, we can rewrite the model as:

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_{j=1}^{n} a_{ij} \bullet N_j) \bullet (1 - \frac{N_i}{K_i}) \tag{23}$$

Among them, K_i represents the population capacity of species i. In this model, we assume that habitat loss leads to ever-decreasing population size, which in turn affects the intrinsic growth rate of the species and thus affects the number of individuals of the species.

It should be noted that these influencing factors are interactive, that is, pollution and habitat loss may affect each other, thereby affecting the dynamic behavior of species. Therefore, these complex interactions need to be considered when building models to more accurately describe the dynamic behavior of plant communities.

9 The plant communities and environment model

9.1 Advise on the impacts of plant communities and the wider environment

To solve this problem, our team decided to adopt a random forest model. An important feature of random forests is the ability to evaluate the importance of feature variables. The basic idea is that when noise is added to a feature variable, a significant reduction in the prediction accuracy of the RF indicates that the feature is more important.

(1) first test the performance of the generated random forest with OOB data to obtain an OOB accuracy rate

(2) then artificially add noise values to the OOB data, and test the performance of the random forest with the noisy OOB data to obtain a new OOB accuracy

(3) The difference between the original OOB accuracy (Gini value) and the noise added OOB accuracy is used as the importance measure of the corresponding feature variable *v*.

This feature of random forests can be used to rank the importance of the variables to ensure their authenticity and completeness.

We use the variable importance V_i score to represent the Gini index G_i, assuming that there are *J* features.

A decision tree $X_1, X_2, X_3, \ldots X_j$ with C categories, now calculate X_j the Gini index score for each feature $V_j^{G_i}$, That is, the average amount of node division impurity of the *j* th feature in all decision trees of RF.

The Gini index for node *q* of the *i* th tree is calculated as:

$$CI_q^i = \sum_{C=1}^{|c|} \sum_{C' \neq 1} P_{qC}^{(i)} P_{qC'}^{(i)} = 1 - \sum_{C=1}^{|C|} (P_{qc}^{(i)})^2 \tag{24}$$

C:(The C category) $P_{qc}^{(i)}$ (The Proportion of category c in q)

Intuitively, this is the probability that two samples are randomly drawn from the node *q*. The importance of feature X_j in the *i* tree node *q*, that is, the change of Gini index before and after branching of node *q* is:

$$V_{ijq}^{(G_i)} = CI_q^{(i)} - CI_l^{(i)} - CI_r^{(i)} \tag{25}$$

Where, $CI_l^{(i)}$ and $CI_r^{(i)}$ representing the Gini index of the two new nodes after the branch, if the node of the feature X_j appears in the decision tree i is the set Q, then the importance of the tree is:

$$V_j^{(G_i)} = \sum_{q \in Q} V_j^{(G_i)(i)} \tag{26}$$

Assuming I trees in RF, then

$$V_j^{(G_i)} = \sum_{i=1}^{I} V_j^{(G_i)(i)} \tag{27}$$

Finally, all the obtained importance scores are normalized。

To ensure the long-term viability of plant communities and the impact on the overall environment, the following measures should be taken:

1. Strengthen the protection of plant communities, Take effective protective measures, Regular repair and tailoring of the plant community, Trim the withered and yellow leaves, Clubranches reduce the unreasonable distribution of nutrients, Improve the efficiency of nutrition utilization; Regular insecticide-killing, Infestations can greatly weaken the viability of

plant communities, To reduces the number of plants and also cause abnormal performance, Insecticidal insects is essential to maintain the healthy growth of plant communities; Understand the living habits of the community, Try to place them in a favorite environment, Avoid or reduce the occurrence of bad return in; Note that the loose soil, Regular soil loosening, Keeping the soil water permeable, oxygen content, Ensure the survival of the root nature. Fertilization and watering, to provide sufficient nutrition for the long-term survival of the plant community, to ensure its survival, such as prohibited deforestation, prohibited fishing, etc.;

2. Strengthen the management of plant communities and take effective management measures, such as regularly checking the condition of the plant community and regularly cleaning up the weeds of the plant community;

3. Strengthen the restoration of plant communities and take effective restoration measures, such as the protection of planting plants and plant germplasm resources;

4. Strengthen the research of plant communities, and take effective research measures, such as studying species diversity, studying community structure, etc.

Impact: In the above measures, insecticidal fertilization may lead to the change of salinity of the soil, change the properties of the soil, and part of the chemical composition will flow into the groundwater, causing pollution to water resources.

10 Sensitivity Analysis

Assume that we randomly perturb the input parameters of the model, specifically, assuming that the growth rate r_i of each species ranges from [0.01, 0.05], each element $A_{i,j}$ of the interaction coefficient matrix A has a value range of [0, 1], the initial number of individuals $x_{0,i}$ has a value range of [50, 150], and the time step is $\Delta t = 1$, the length of time is $T=100$. We randomly generate a set of parameters and then calculate the simulation results.

Random perturbation can be achieved by adding a certain amount of random noise to the original parameters. A simple approach is to add a random value from a normal or uniform distribution to the original parameter value. For example, for an original parameter p_i, the following formula can be used for random perturbation:

$$p_i^{'} = p_i + \varepsilon_i \tag{28}$$

Among them, $p_i^{'}$ is the parameter value after perturbation, and ε_i is a random value that obeys a certain probability distribution. Different probability distributions and parameters can be chosen according to the specific situation.

The sensitivity analysis chart is shown in Figure 6 below:

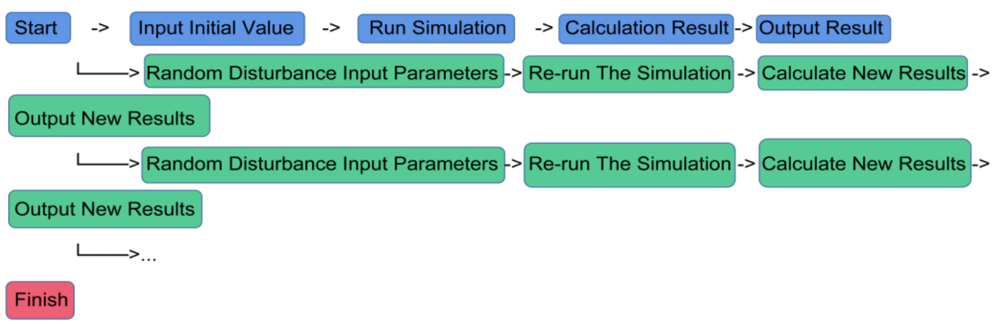

Figure 6: Sensitivity flowchart

In the flow chart above, we can see that the starting point of the whole process is to input the initial value, then perform simulation, calculate the result, and output the result. Then, we perform random perturbation on the basis of the original input parameters, re-simulate, calculate new results, and output new results. Through such a cyclic process, we can obtain a series of different results to evaluate the robustness and stability of the model in different situations.

Therefore, the calculation results are shown in Figure 7:

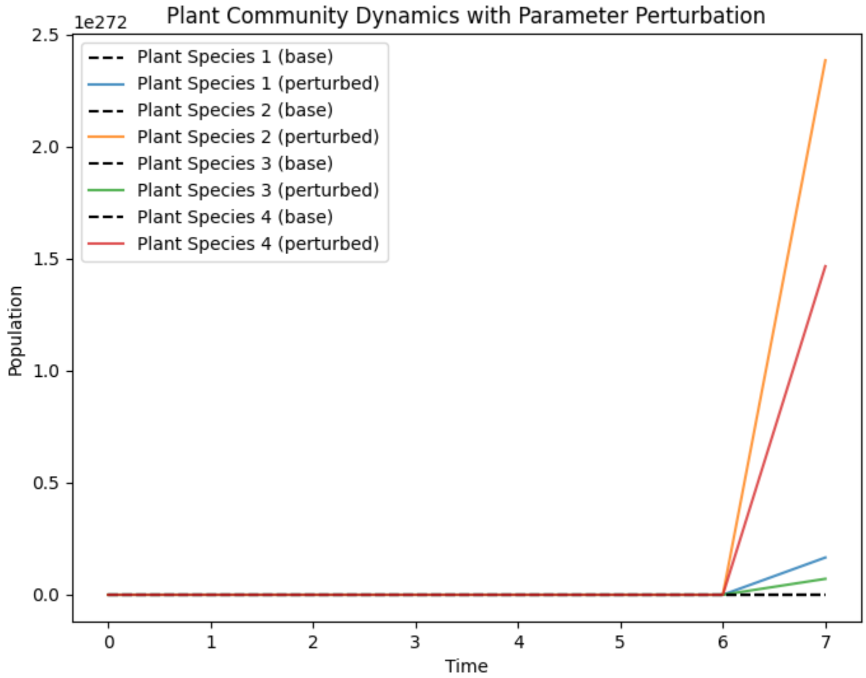

Figure 7: Plant Community Dynamics with Parameter Perturbation

From the results of random perturbation, the number of different plant populations in the community will change greatly over time. In some cases, certain plant populations may rapidly decline or disappear in a community, while in other cases they may increase rapidly and become dominant in the community. This shows that the number of plant populations in the community is interrelated, and the impact of the interaction on the number of plants is complex and subtle, and multiple factors need to be considered comprehensively for analysis and prediction.

11 Model Evaluation and Further Discussion

11.1 Strengths

Advantages of differential equation models:

1. Differential equation models provide a complete framework for describing the behavior of a system and can explain how the system changes between different points in time.

2. Differential equation models can relatively easily determine future states and provide reliable predictions to better control system behavior.

3. Theoretically, differential equation models can solve complex problems in many systems and provide a more general solution.

11.2 Weaknesses

Disadvantages of differential equation models:

1. Differential equation models are complex to set up and can require a lot of data and time to determine.

2. Differential equation models may not accurately describe real systems because there may be unpredictable factors in the real system that are often overlooked in differential equation models.

11.3 Further Discussion

Important aspects of differential equation model improvement include:

1. Improve the flexibility of the model: more complex models such as mixed differential equation model, dynamic model and multivariate differential equation model can be adopted to better describe the system behavior;

2. Improve data fitting performance: new data fitting techniques, such as matrix-based fitting methods, can better fit data;

3. Improved parameter estimation method: more accurate parameter estimation methods, such as parameter estimation based on fuzzy theory, can estimate model parameters more accurately;

4. Optimize the model structure: Using new model structure optimization methods, such as model structure mining methods, we can better explore the hidden laws in the system, so as to improve the model.

12 Conclusion

For this problem we use differential equation model, multiple linear regression equation model, LTI model and random forest model to solve the problem analytically.

For question one the Lotka-Volterra model was first used to describe the quantitative relationship between predator and prey, and the model was improved by means of multiple linear regression, which in turn estimated the number of individuals per species over time and predicted the response of species under different climatic conditions. The effects of different species populations on plant community adaptation and long-term survival can also be explored. Finally, the differential equations are solved and code is written in Python to predict changes in plant communities. From the images generated by the model, it is analyzed that the number of species with higher intrinsic growth rate has a significant growth trend than others when the aridity is several times the intrinsic growth rate of the plants.

For Problem 2-1, an LTI model was used to describe the population dynamics of each species by a set of differential equations. The LTI model provides a mathematical framework for understanding the long-term dynamics of ecological communities and the factors that affect their stability and resilience. The corresponding parameters were set and used to represent the population growth rate; assumptions were made to derive the growth rate and interaction matrix, and the interaction strength was represented using the matrix, and time was discretized to obtain the corresponding set of nonlinear equations

For problem 2-2, varying the interaction coefficient to reflect the effect of species type on the results, in order to verify the effect of species type on the results, we can try different interaction coefficient matrices and compare the differences in simulation results under different cases. Using the above interaction coefficient matrix, a mathematical model was constructed to solve the problem, and the model was simulated by computer to compare the differences in simulation results under different interaction coefficient matrices.

For problems 2-3, the differential equations are still solved by constructing a differential equation model and setting parameters to assist in the analysis to obtain the differential equations during drought. In the simulation, more frequent and extensive droughts can be simulated by increasing the probability and severity of drought events. Different parameter values can be tried to observe the simulation results.

For questions 2-4, two factors, pollution and habitat loss, are considered in this model, i.e., external factors that affect the intrinsic growth rate of each species. By taking both factors into account, the model can describe the dynamic behavior of plant communities more clearly. It is additionally important to note that these influencing factors interact with each other, i.e., pollution and habitat loss may influence each other and thus the dynamic behavior of the species.

For problems 2-5, a random forest model was used, and an important feature of random forests is the ability to assess the importance of characteristic variables, a property that can be used to rank the importance of variables to ensure their veracity and integrity. The corresponding solution policies to ensure the long-term viability of the plant community and the impact on the overall environment are also given.

References

[1] Vilagrosa A , Morcillo L , Muoz J C , et al. Impact of climate change disturbances affecting Mediterranean plant communities, ecohydrology dynamics and forest water balance. 2019.

[2] Rasmussen P U , Bennett A E , Tack A . The impact of elevated temperature and drought on the ecology and evolution of plant–soil microbe interactions[J]. Journal of Ecology, 2020, 108(1).

[3]Mariotte P , Vandenberghe C , Hagedorn F , et al. Subordinate plant species enhance community insurance to drought in semi-natural grasslands. 2012.