Exploration of the Ability of Fungi for Decomposing Natural Resources Based on Multiple Regression Equation and Cellular Automata

Yufei Shi*, Yijiang Guan, Dingye Zhang
School of Electronic Information and School of Computer Science, Sichuan University, Sichuan, China

* Corresponding author email: yufeishi@scu.edu.cn

Abstract. With the development of society, carbon emissions are increasing. The key organisms to maintain the stability of the carbon cycle are fungi that can be easily seen and ignored. In this paper, we selected several fungi to establish the model of decomposition and reproduction so that we can understand the role they played. First of all, we studied several physiological indexes of fungi, and established the degradation model through multiple regression analysis, and multiple linear regression equation for the relationship between decomposition rate, growth rate, unit volume density of mycelium, temperature and humidity tolerance. Next, we established the competitive growth model based on logistic model, simulated the competitive growth process of strains with different growth rates, humidity tolerance, and the total decomposition rate. In order to be closer to the real situation, we set up the competitive growth model among four species. By arranging fungal communities randomly to simulate different biodiversity, we analyzed the effects on the decomposition rate in the case of that the environmental temperature and humidity changed by 10% respectively. After that, we established a growth prediction model based on ARIMA. By querying the climate data of five typical climates, we established the competitive growth model with 4 combinations, and we obtained a short-term model, a medium-term trend and a long-term forecast to describe growth, reproduction and decomposition rate. In order to refine the strains of the pressure of competition and the influence of the distance between the strains of competition, we have established improved competition evolution model based on the cellular automata theory of population. The model helped us comprehend the competition between species on a micro level. All these analyses showed us the significance of biodiversity and the great role decomposers play in Earth.

Keywords: multiple linear regression; lasso regression; logistic model; ARIMA model; Cellular Automata.

1. Introduction
Decomposers break down dead organisms so that the nutrients in them can be recycled back into the ecosystem. Fungi account for the majority of decomposers. Research shows that there is a certain
relationship between some characteristics of fungi, such as growth rate and humidity tolerance. Besides, different fungi in the same place may interact each other’s growth rate. So, we explored the factors affecting the reproduction and decomposition rate of fungi based on these materials.

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

| Abbreviation | Description                                |
|--------------|--------------------------------------------|
| $x_1$        | The growth rate of fungi                   |
| $x_2$        | Unit bulk density of hypha                 |
| $x_3$        | Ambient temperature                        |
| $x_4$        | Resistance to humidity                     |
| $y$          | The decomposition rate of woody fibers     |
| $\mu$        | Other related factors                      |
| $A$          | Represent a kind of fungus named Armilaria_galica_FP102531_C6D |
| $B$          | Represent a kind of fungus named Armilaria_galica_EL8_A6          |
| $C$          | Represent a kind of fungus named Armilaria_sinapina_PR9            |
| $D$          | Represent a kind of fungus named Phlebiopsis_flvidoalba_FP102185   |
| $c_1(t)$     | The number of fungal population A          |
| $c_2(t)$     | The number of fungal population B          |
| $r_x$        | Constant growth rate of fungal population X|
| $N_x$        | Environment capacity of fungal population X|
| $g_x$        | The growth rate of fungi                   |
| $t$          | Ambient temperature                        |
| $h$          | Ambient humidity                           |
| $N_{i, j-1}$ | The left adjacent cell                     |
| $N_{i, j+1}$ | The right adjacent cell                    |
| $N_{i-1, j}$ | The up adjacent cell                       |
| $N_{i+1, j}$ | The down adjacent cell                     |

3. Strain degradation model based on lasso regression and multiple linear regression

3.1. Modeling of fungal activity ability under different environmental conditions
At the beginning of the problem, through consulting a large number of existing literature and the limitations of the title[1-6], we found that the most direct measurement index of fungi in the degradation reaction is decomposition rate, growth rate, unit volume density of mycelium, temperature, and humidity. These four variables with mutual influence are the important factors affecting the decomposition rate, and there are also differences among these factors.

Therefore, the five most important parameters in the process of fungi decomposing lignocellulose were set as variables.

3.2. Establish multiple linear regression equation based on OLS
According to the analysis, the change of temperature, humidity resistance, growth rate and bulk density will inevitably lead to the change of decomposition rate. At the same time, for each kind of fungus, its decomposition rate is different at different temperatures. This is consistent with multiple observations of multiple linear regression. The linear fitting model is as follows.
Among them, the independent variables are growth rate, unit volume density of mycelium, temperature, humidity tolerance, and decomposition rate, which are unobservable and meet certain conditions. It means the change of $Y$ caused by changing one unit while controlling other independent variables.

3.3. Solution of multiple linear regression equation

For 38 species, the decomposition rates of each species were recorded at 10 °C, 16 °C, and 22 °C. We use the repeat command in Stata software to make multiple linear regression and get the following results.

1. Decomposition rate and unit volume density of mycelium

Firstly, the relationship between different unit volume densities of various fungi and their corresponding decomposition rate $y$ is studied. We use the tool of state multiple linear regression to calculate the results as the table 2.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + \mu$$  \hspace{1cm} (1)

It can be seen from table 2 that the decomposition rate is negatively correlated with the unit volume density of mycelium. The larger the unit volume of mycelium, the worse the ability of decomposing wood fiber. The accuracy of our model is also verified.

2. Decomposition rate and humidity resistance

According to the experience in (1), we can draw an analogy between humidity resistance and its corresponding decomposition rate $y$ as the table 3.

$$y = 14.2166 - 6.2236x_2$$  \hspace{1cm} (4)

And because $P < 0.05$, it can be determined that the regression coefficient meets the condition at 95% confidence level.

We use the tool of state multiple linear regression to calculate the results as the table 2.

Table 2. Linear equation parameters and p-value

|       | Coef.    | P>|t|   |
|-------|----------|------|
| $x_2$ | -6.223635| 0.023|
| cons  | 14.2166  | 0.000|

It can be seen from table 2 that the decomposition rate is negatively correlated with the unit volume density of mycelium. The larger the unit volume of mycelium, the worse the ability of decomposing wood fiber. The accuracy of our model is also verified.

It can be seen from table 3 that the decomposition rate is positively correlated with humidity tolerance, and the higher the humidity tolerance, the higher the decomposition rate.

Table 3. Find the linear equation parameters and p

|       | Coef.    | P>|t|   |
|-------|----------|------|
| $x_4$ | 12.21508 | 0.020|
| cons  | 18.65938 | 0.000|

It can be seen from Table 3 that the decomposition rate is positively correlated with humidity tolerance, and the higher the humidity tolerance, the higher the decomposition rate.
Similarly, $P < 0.05$, that is, at the 95% confidence level, the regression coefficient meets the conditions.

$$y = 18.65938 + 12.21508x_4$$

(3) Relationship model among growth rate, temperature and decomposition rate

According to the relationship between the temperature of growth rate and its decomposition rate $y$, a binary linear regression model was established. The result is as the table 4.

### Table 4. Find the linear regression equation parameters and $p$

| $y$ | Coef. | $P>|t|$ |
|-----|-------|--------|
| $x_1$ | 1.792675 | 0.000 |
| $x_3$ | 1.35902 | 0.000 |
| cons | -16.14538 | 0.000 |

According to table 4, the decomposition rate is positively correlated with temperature and growth rate. The higher the temperature and growth rate, the higher the decomposition rate. The results are as follows:

$$\begin{cases}
\beta_0 = -16.14538 \\
\beta_1 = 1.792675 \\
\beta_3 = 1.35902
\end{cases}$$

$$P < 0.05$$, which means that the regression coefficient meets the condition at 95% confidence level.

$$y = -16.14538 + 1.792675x_1 + 1.35902x_3$$

By calculating the decomposition rate under different temperature and different growth rate of fungi. We make the following Table 3.

![Three-dimensional diagram of growth rate, temperature and decomposition rate](image)

It can be seen from Fig. 1 that when the temperature is between 10 °C and 22 °C, the decomposition rate of bacteria increases with the growth rate and temperature, indicating that most bacteria are more suitable for higher temperature (between 10 °C and 22 °C)

(4) The decomposition rate was related to temperature, humidity, unit volume of mycelium and humidity tolerance

In the regression, when there are many variables, it may lead to multicollinearity problem, which makes the regression system not significant. If the influence is large, it may even lead to the failure of
OLS estimation. In order to avoid the influence of this factor, we directly use lasso regression in this multivariate regression, and compared with the ordinary OLS based regression model, the loss of lasso regression in OLS regression model is obvious. Different penalty terms are added to the loss function, and the added penalty term is composed of regression coefficient function, which can compress the regression coefficient of relatively unimportant variables to 0, and make the model more estimable. Even if the previous data does not meet the full rank, the parameters can be estimated.

The regression formula of Lasso is as follows

$$\hat{\beta} = \text{arg}\min \left( \sum_{i=1}^{n} (y_i - x_i \hat{\beta})^2 + \lambda \sum_{i=1}^{k} |\hat{\beta}_i| \right)$$

(9)

We use cvlasso instruction in Stata to do 10 fold cross validation, and we set random number seed to 520 to make the result repeatable. We calculate the influence of four factors on Y and get in Table 5.

Table 5. Gives the parameters of linear regression equation and P

| y     | Lasso     | P>|t| |
|-------|-----------|-----|
| $x_1$ | 1.742443  | 0.000 |
| $x_2$ | -0.2095803| 0.000 |
| $x_3$ | 1.100623  | 0.000 |
| $x_4$ | 1.100623  | 0.000 |
| _cons| -15.28673 | 0.000 |

It can be seen from Table 5 that the decomposition rate is positively correlated with the growth rate, temperature and humidity resistance. When these three factors increase, the decomposition rate increases. At the same time, the decomposition rate is negatively correlated with the unit volume density of mycelium. The higher the unit volume density is, the lower the decomposition rate is.

According to P < 0.05, the regression coefficient meets the condition at 95% confidence level.

$$y = -15.28673 + 1.742443x_1 - 0.2095803x_2 + 1.100623x_3 + 1.100623x_4$$

(10)

3.4. Establish a differential equation to describe the growth rate of competing fungi

Because the experimental environment is closed, it means that the increase and decrease of the number of fungi only depends on the reproduction and death of fungi, and each individual has the same reproduction and mortality. Moreover, due to the limited natural resources, we use logistic model to describe the evolution process of the number of fungi, which is

$$\frac{dc(t)}{dt} = r_1 \left( 1 - \frac{c}{N} \right)$$

, the number of fungi increases and decreases. Among them, R is the growth rate, n is the maximum number that can be contained in the environment, and C is the current number of the strain.

Assuming that the two fungi compete with each other, $c_1(t)$ and $c_2(t)$ are defined as the number of species A and B. $r_1$ and $r_2$ are defined as the inherent growth rate of the two species A and B. $N_1$
and \( N_2 \) are defined as the maximum capacity of species A and B, and \( c_1(0) \) and \( c_2(0) \) describe the initial number of species A and B.

If the effect of B is not considered, the quantity of strain a follows the logistic model \( \frac{dc_1(t)}{dt} = r_1c_1\left(1 - \frac{c_1}{N_1}\right) \), and the factor \( 1 - \frac{c_1}{N_1} \) reflects the blocking effect of strain A on its own growth due to the consumption of limited resources, and \( \frac{c_1}{N_1} \) can be expressed as the quantity of nutrition which \( N_1 \) give for A when the quantity is \( c_1 \) (assuming only the nutrition for strain A is considered, the total amount is 1).

If the competition of strain B is considered, the growth rate of a will be reduced because strain B consumes the same natural resources. Therefore, \( \frac{dc_1(t)}{dt} = r_1c_1\left(1 - \frac{c_1}{N_1} - \sigma_1 \frac{c_2}{N_2}\right) \) is set. \( \sigma_1 \) means that the nutrition consumed by the unit number of strain B is \( \sigma_1 \) times as much as that consumed by the unit number of strain A.

Similarly, for population B, \( \frac{dc_2(t)}{dt} = r_2c_2\left(1 - \frac{c_2}{N_2} - \sigma_2 \frac{c_1}{N_1}\right) \) is set. The coefficient \( \sigma_2 \) here means that the nutrition consumed by a strain per unit number is \( \sigma_2 \) times as much as that consumed by B strain per unit number.

So, we brought in Armillaria_gallica_FP102531_C6D (A), Armilaria_galica_EL8_A6f (b) two kinds of fungi data, through the ode45 function in MATLAB, calculate the number of two kinds of fungi and the total decomposition rate with time, as shown in the Table 6.

![Figure 2. The number of bacteria in populations A and B](image1)

![Figure 3. The total decomposition rate curve of population A](image2)
It can be seen from Fig. 2 that when strains A and B grow in the same environment, A is more affected by B, the growth rate of population number decreases, and population B is in a dominant position in the competition; at the same time, due to the Fig. 3, with the passage of time, the number of both populations tends to be stable, and the total decomposition rate becomes flat with the passage of time and the no longer increase of species number.

3.5. Long-term trend model of fungal interaction based on ARIMA
First of all, from the questions above, we can draw the decomposition rates of strains A, B, C and D at different temperatures and humidity. The Fig. 4 and Fig. 5 indicate the relationship between decomposition rate of strain A and temperature and humidity and the relationship between the decomposition rate of strain B and temperature and humidity respectively.

\[ g_1 = -0.001794 t^2 - 0.1791 h^2 + 0.06495 t + 0.6073 h - 0.531 \\
 g_2 = -0.01279 t^2 - 4.608 h^2 + 0.4879 t + 5.886 h - 4.56 \\
 g_3 = -0.01164 t^2 - 5.732 h^2 + 0.4212 t + 13.66 h - 8.126 \\
 g_4 = -0.003883 t^2 - 0.278 h^2 + 0.1428 t + 1.433 h - 1.128 \]

\( g_k \) stands for the growth rate; \( t \) stands for temperature; \( h \) stands for humidity.
ARIMA model is the differential integrated moving average autoregressive model, also known as the integrated moving average autoregressive model (movement can also be called sliding), which is one of the time series prediction analysis methods. MA means the "moving average", Q means the number of terms of the moving average, and D is the number of difference (order) made to make it a stationary sequence. The word "difference" does not appear in the English name of ARIMA, but it is the key step.

ARIMA (p, d, q) model:

\[ y_i = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{i-1} + \varepsilon_i + \sum_{i=1}^{q} \beta_i \varepsilon_{i-1} \]  

\[ y_i = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{i-1} + \varepsilon_i + \sum_{i=1}^{q} \beta_i \varepsilon_{i-1} \]  

\[ \sum_{i=1}^{p} \alpha_i L^i \text{ is AR(p), } (1 - L)^d \text{ is D order difference, } \sum_{i=1}^{q} \beta_i L^i \text{ is MA(q).} \]

Then we calculate the long-term humidity change and temperature change as the following Fig. 6 and Fig. 7.

![Figure 6. Time series diagram of long-term humidity (including prediction) [7]](image)

![Figure 7. Time series diagram of long-term temperature (including prediction) [8]](image)

To process the data, we use an expert modeler. Its principle is to find each dependent sequence model automatically. After the predictor variable of independent variable is specified for it, the mathematical model that has significant statistical relationship with the dependent sequence will be automatically selected in the contents of the ARIMA model. When the obtained model fits, the model variables are transformed using difference, square root, or natural logarithm transformations. By default, the expert modeler will consider both the exponential smoothing model and the ARIMA model.
Table 6. Obtain ARIMA model parameters

|           | estimate | Standard error | t     | significance |
|-----------|----------|----------------|-------|--------------|
| constant  | 18.963   | 0.219          | 86.505| 0.000        |
| delay1    | 0.0647   | 0.100          | 6.452 | 0.000        |

\[
\begin{align*}
\alpha_0 &= 18.963 \\
\sum_{i=1}^{r} \alpha_i &= 0.0647 \\
\sum_{i=1}^{q} \beta_i &= 0 
\end{align*}
\]  

(15)

Next, white noise residual test is performed on the obtained results as the Fig. 8.

Figure 8. ACF and PACF of residual diagram

Figure 9. Decomposition rates of the four fungi when the temperature fluctuates by 10%
Figure 10. Decomposition rates of four fungi when humidity fluctuates by 10%.

Table 7. $R^2$ and Q of residual diagram

|        | $R^2$ | Q(18) |
|--------|-------|-------|
|        | .401  | .191  |

As can be seen from Fig. 8, the autocorrelation coefficient and partial autocorrelation coefficient of the lag order have no significant difference from 0. According to Table 7, the significance $P$ obtained by the Q test of the residual is 0.191, which is greater than 0.05, so the null hypothesis cannot be rejected and the residual can be considered as a white noise sequence. Therefore, the ARIMI model can well represent the population performance of fungus AB at the temperature of Los Angeles. The decomposition rates of the four fungi when the temperature and humidity fluctuate by 10% in the Fig. 9 and Fig. 10 are credible.

4. Fungal competition model based on cellular automata

In front, the population competition evolution model based on logistic model considers the reproduction and competitive growth of the population from a macro perspective. In order to facilitate the calculation, we consider the competition among strains at all positions on the wood fiber to be the same. Although this method greatly simplifies the calculation, it also makes the model ignore the influence of strains at different positions. We don't need this factor in order to reduce the pressure of competition.

In order to eliminate this inappropriateness, we use the theory of cellular automata to build a discrete microscopic model to simulate the process of population competition and evolution[9].

In the sub model, we use 10000 cells to represent the lignocellulosic fibers. This wood fiber is the entire cellular space. The cells are divided into three types: forbidden cells (representing the boundary of lignocellulosic fibers), fungi occupied cells and empty cells. Each cell is equivalent to a small ecological system.

Our cell space is the 10000 cells in which the whole lignocellulosic fiber is located. Our defined rule is that in the same cell, the colonies are close enough to meet the population competitive evolution relationship of the previous logistic model. During the process of growth and reproduction, strains will migrate from high density cells to low density cells. Therefore, we can list two strains to establish based on cellular automata, the model of competition between the two is as follows.
The subscripts $i, j$ denotes the cells in the i-th row and j-th column of the grid on the two-dimensional plane, and $\sum (\sum_{i \pm 1, j \pm 1} - 4y_{i,j})$ is the diffusion of the i-th and j-th cells by adjacent strains. In this model, the cellular automata design adopts von Neumann type, that is, there are four cell bodies adjacent to the cell, namely $N_{i-1,j}$ (left cell), $N_{i+1,j}$ (right cell), $N_{i,j-1}$ (upper cell), $N_{i,j+1}$ (lower cell). $\mu$ is the diffusion coefficient, $N_X$ and $N_Y$ is the maximum capacity of strain A and strain B.

The competition between two species in a single cell is extended to space. We simulate the population dynamics of the two species in a 100*100 two-dimensional grid in time. In order to simulate the scene in real life more realistically, in the 100*100 area, there are 10 initial species A in 10 cells randomly, and another 10 initial species B in the other 10 cells randomly. It is assumed that in the von Neumann neighborhood, the cell with high density diffuses to the cell with low density, and the diffusion rate is 0.25 times of the density difference, that is, we use $\mu=0.25$.

According to the previous experimental data, we take growth $a=0.25$, growth $B=1.96$, Level of competitiveness $a=0.232, b=0.569$. The maximum environmental capacity in each cell is 100. Then we use MATLAB to simulate the growth and reproduction of population.

In Figure 11-14, yellow represents fungus B, and blue represents fungus A. The shade of the color represents the number of fungi A and B in the cell. The darker the color, the closer it is to the capacity of the environment. The lighter the color, the less the quantity. The color of each cell represents the dominant population in that cell. Figures 11-14 show the simulation of the spatial distribution of a colony in the 5th, 10th, 15th and 30th generations.

Figure 11. The spatial distribution of 5 generations of colonies
Figure 12. The spatial distribution of 10 generations of colonies

Figure 13. The spatial distribution of 15 generations of colonies

Figure 14. The spatial distribution of 30 generations of colonies
As can be seen from the figures 11-14, the distribution of the whole cellular space tends to be stable after about 30 generations. You can see that after 30 generations, the yellow fungus B takes up most of the entire cellular space. This is consistent with the natural growth rate of fungus B and the competitive experimental data, which verifies the accuracy of the model.

5. Conclusion
There are close relationships between the decomposition rate of fungi and temperature, humidity tolerance, growth rate, mycelial density per unit volume. Commonly colony which has more competitive ability or more resistance of low temperature are strong. With the gradual enrichment of fungal species, the overall decomposition rate has been greatly increased, and the resistance to climate change has also been significantly improved. It is important to protect the biological diversity.

References
[1] Lustenhouwer, N., Maynard, D. S., Bradford, M. A., Lindner, D. L., Oberle, B., Zanne, A. E., & Crowther, T. W. (2020). A trait-based understanding of wood decomposition by fungi. Proceedings of the National Academy of Sciences, 117(21), 11551-11558.
[2] Maynard, D. S., Bradford, M. A., Covey, K. R., Lindner, D., Glaeser, J., Talbert, D.A., ... & Crowther, T. W. (2019). Consistent trade-offs in fungal trait expression across broad spatial scales. Nature microbiology, 4(5), 846-853.117(21), 11551-11558.
[3] Maynard, D. S., Bradford, M. A., Lindner, D. L., van Diepen, L. T., Frey, S. D., Glaeser J. A., & Crowther, T. W. (2017). Diversity begets diversity in competition for space. Nature ecology & evolution, 1(6), 1-8.
[4] Maynard, D. S., Crowther, T. W., & Bradford, M. A. (2017). Fungal interactions reduce carbon use efficiency. Ecology letters, 20(8), 1034-1042.
[5] Maynard, D. S., Crowther, T. W., & Bradford, M. A. (2017). Competitive network determines the direction of the diversity–function relationship. Proceedings of the National Academy of Sciences, 114(43), 11464-11469.
[6] https://www.timeanddate.com/weather/usa/new-york/climate [Online; accessed 7-February-2021]
[7] https://www.timeanddate.com/weather/usa/los-angeles/climate [Online; accessed 7-February-2021].
[8] Niu Bo, Zhang Yuan, Qin Jianping, Lian Feiyu, Wang Xue. Journal of Henan University of Technology: Natural Science Edition, 2013, 34(2): 77-79.