Predicting Nature's Demand for Bees based on Time Series Models

Chengjie Lu [1], Yixue Wu [2], Yufan Xu [3], Zixuan Wang [3]

[1] Shanghai Wenqi Middle School, Shanghai, 200000, China

[2] Chengdu Jiaxiang Foreign Language School, Chengdu, 610000, China

[3] Shanghai World Foreign Language Academy, Shanghai, 200000, China

Abstract. Bees play a very important role in the entire ecosystem, as their pollination ensures the genetic diversity of flowering plants. Therefore, we need to study the factors influencing bee population size and its pollination. In this paper, we build a mathematical model and solve it to obtain the influence of different factors on the number of bee populations, and calculate the number of bees required for pollination in a certain space. Finally, we analyze the advantages and disadvantages of the model, as well as some more discussion.

Keywords: Bee Populations; Principal Component Analysis; Time Series Forecasting; Grey Relation Analysis.

1. Introduction

1.1 Background

Bees are the most famous species of hymenoptera. There are nine bee species in the world, in addition to 30,000 wild bee species. They collect pollen from up to 1,000 flowers a day. Nectar bees collect pollen from about 300 flowers a day. Bees carry pollen to other plants, promoting plant reproduction. 80% of all flowering plants are pollinated by insects. According to a United Nations biodiversity survey, up to 40% of the world's insect species are facing extinction. Without these small pollinators, the stability of ecosystems would be difficult to guarantee, many plants in nature would face survival problems, and humans would have to do their jobs by hand. Therefore, we need to build a model to comprehensively consider the influence of various natural factors, external factors, and human factors on the bee population to accurately predict the number of bee populations. At the same time, we need to build a model to study the pollination of bee populations, which can effectively observe the impact of bee population on natural pollination.

1.2 Literature Review

The characteristics of population growth type generally have two basic types, namely J-type growth type and S-type growth type. The J-type growth model, also known as the Malthus exponential growth model, is the basic form of population growth in infinite space. The S- shaped growth type is modeled by the Logistic equation. The Logistic equation was proposed by the Belgian Verhulst and later discovered by Pearl-Reed, hence the Verhulst-Pearl equation. The logistic equation is a classical ecological model that points out the basic law of population growth in limited space, which has been widely used in the field of population ecology to describe the spatiotemporal dynamic relationship of biological population numbers. In recent years, the research on the logistic model has mainly focused on the correction of equations, parameters (population endowment growth rate r, environmental capacity k, integral constant a) estimation, data fitting and optimal capture strategy.

1.3 Restatement of the Problem

In problem A, the first question requires the establishment of a model to determine the change in the number of bee colonies over time, so it is necessary to investigate the relevant parameters, according to the conditions in the question, due to the heavy workload in summer, the life span of most bees will be shortened, so it is necessary to investigate the bee spawning cycle, egg production

rate, fertilized egg/unfertilized egg ratio, and other factors, and also need to consider the environmental capacity. The second question requires an analysis of the model to determine the factors that have the greatest influence on the size of the bee colony. The impact of changes in lifespan, spawning rate, fertilized/unfertilized egg ratio and other factons the total colony needs to be analysed. The third question is the need to build models and predict the number of hives that support pollination on a 20-acre plot of land. It should be taken into account that bees can travel 20 kilometers, but usually stay within 6 kilometers of the hive; A bee can visit about 2,000 flowers or more in a day. The fourth question requires a non-technical one-page blog or infographic that requires us to analyze the results obtained in combination with the first three questions.

2. Assumptions and Justification

To simplify the problem and make it convenient for us to simulate real-life conditions, we make the following basic assumptions, each of which is properly justified. Assume that the living space of the bee population is limited. Assume no natural enemies and population competition. Assume no effects of epidemics and infectious diseases on populations. Assume that there is no impact of emergencies in the natural state on the bee colony.

3. Model Overview

In this section, we will outline the model we designed. For the first part, by reviewing relevant information, we added some data necessary to solve this problem, such as macro data on the population changes in bee populations in recent years, bee fertility and mortality, etc.

Secondly, we use our basic population size model as the basis, comprehensively considering nine factors such as activity level, pollen consumption, protein abundance, birth rate, mortality, lifespan, spawning rate, fertilization rate, and unfertilized ratio, to optimize the basic model into a population model that comprehensively considers multiple factors.

In the second part, we took the bee population as the dependent variable and each influencing factor as the independent variable, conducted principal component analysis, obtained relevant statistical results, compared the primary and secondary relationship between each influencing factor, and divided it into four degrees of influencing factors. Therefore, as shown in figure 1, we first collected the relevant data and preprocessed the data.

Secondly, we designed mathematical models that could simulate bee populations. Thirdly, we use relevant data for principal component analysis and grey association analysis, and calculate relevant results to answer relevant questions.

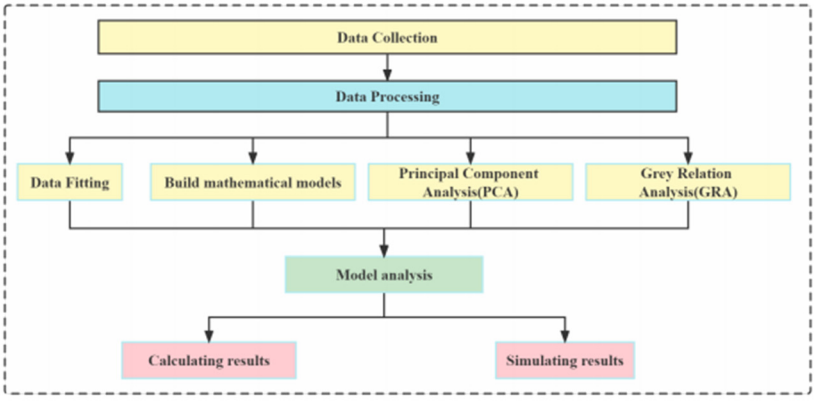


Figure 1. Modeling process

4. Model Description

4.1 Necessary Data Supplementation

In recent years, the total number of bee colonies in the world has shown a slow growth trend, with the number of global bee colonies increasing to 92.265 million in 2017 and 92.265 million in 2018. According to the trend of world bee colonies, it is estimated that the total number of global bee colonies in 2019 will be about 93. 145 million.

Changes in colony growth: In the spring season, as colonies grow, their growth slows down. The number of larvae assigned to each feeding bee decreased, while the number of bees not involved in feeding gradually increased. At this time, there are more bees in the colony that have emerged from the room, which has exceeded the needs of the colony to feed the larvae. When the group reaches 15-16 spleen, the growth rate is very slow, almost close to zero. At this time, the queen bee lays a gradual decline in egg production, and the colony develops a bee sharing mood.

Loss rate of bee colonies during the honey flow period: The growth rate of bee colonies depends largely on the number of bees dying, and bee deaths depend on the flow of honey in nature. When there is a shortage of honey sources, bee deaths decrease, and when honey sources are abundant, bee deaths increase. The average life expectancy of bees in the early stage of honey flow is 29.4 days, the average life expectancy of large honey flow is 25.7 days, and the average life expectancy of honey flow is 32.8 days in the late stage of honey flow. The large honey period shortens the life span by 16.6% compared with the early stage of the honey flow and 20% shorter than the late stage of the honey. Both feeding and gathering work can affect the lifespan of bees, and bees fly nervously and shorten their lifespan than in-nest work.

Bee nesting related data: leaf cutting time: 7~1002s, average 197.2±213.63s, 65% of the time of leaf cutting between 40~220s. Residence time in the nest with leaves: 11~363s, with an average of 101±91.92s, and the leaf cutting time is between 25~125s in 61% of cases. Collection time: 165~2263s, average 848.9±511.84s, 71% of the time of leaf cutting is between 250~1050s. Residence time in the nest with pink honey: 43~292s, with an average of 99.6±43.62s. In 71% of cases, the leaf cutting time is between 40~220s

5. Model Solution

In this section, we collected data on 9 possible influencing factors to analyze their impact on bee populations, the meaning of the 9 main influencing factors is explained below:

Table 1. Influencing factors

Variable	Influencing factors
X1	Activity Level
X2	Pollen Consumption
X3	Protein Abundance
X4	Birthrate
X5	Mortality
X6	Life Span
X7	Spawning Rate
X8	Fertilization Ratio
X9	Unfertilized Rate

The software took the bee population as the dependent variable and each influencing factor as the independent variable (Xi) for relevant analysis. Taking the monthly population of bee as the dependent variable and each influencing factor as the independent variable (Xi),the principal

component analysis and factor analysis method were used to analyze the principal component and the criterion for selecting the principal component was that the principal component accounted for more than 90% of the total information.

Next, the grey correlation coefficient between bee population and each influencing factor was analyzed by software. The margin method was used to analyze the amount of insects at the flowering stage.

According to the data, after multiple regression analysis, the mathematical model and test results of bee population prediction are as follows:Y= 11.8715 + 0.0381X1 + 5.7108X2−10. 1094X3−0.3779X4 + 2. 1882X5 + 0.0311X6−0.9127X7 + 15.5782X8−7.5512X9,After testing, we can get r = 0.7686, F = 5.05 > F0.01 (m, n − m − 1), P < 0.01.From the above results, it can be seen that there is a significant linear relationship between the population of bees and natural factors, and the significance level is P < 0.01. Therefore, the use of this formula to predict the number of bees has good reliability.

6. Model Solution

6.1 Time Series Forecasting Models

where t is the current period, h is the predicted pre-previous number, also known as the forecast step, xt is the estimated level of the t period, lt is the estimated level of the moment t, bt is the predicted trend of the moment t, a is the horizontal smoothing parameter, p is the smoothing parameter of the trend, where a, p is the main estimate of the model.The time series decomposition model is established with the help of SPSS software, and the trend values obtained are as follows:

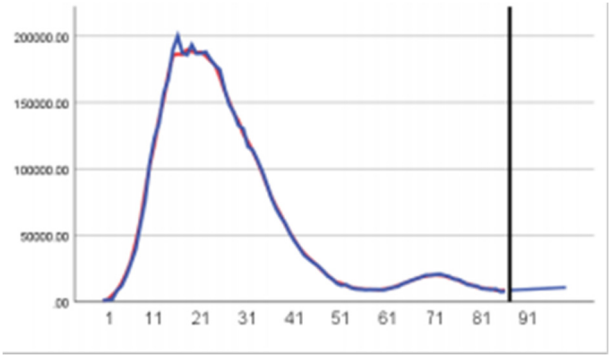


Figure 2. Time series forecasting models

The residual statistics for this model are Estimation solution of a data:

Table 2. Estimation solution of a data

	Estimate	Standard Error	t	Saliency cumulative
α	0.999	0.054	18.647	0.000

Therefore, the above model can answer question 1 well, and can simulate and predict thepopulation of bees.

6.2 Principal Component Analysis(PCA)

Taking the monthly bee species population as the dependent variable and each influencing factor as the independent variable, the factors affecting the growth and decline of the bee species population in nature were analyzed by principal component analysis. The results show that the cumulative contribution rate of the first principal component to the fourth principal component reaches 93. 2%, the variance contribution of other principal components gradually decreases. It shows that 90% of the

first four principal components have reflected the factors affecting the population change of bee species, which can be analyzed as a comprehensive index for the selection of influencing factors.

It can be seen from Table 3 that the coefficients of the four factors of activity level X1 , pollen consumption X2, protein richness X3 and lifespan X6 in the first principal component are large, which are the main influencing factors affecting the change of population, so the first principal component value is regard the activity level. The coefficient so monthly spawning rate X7 and unfertilized ratio X9 in the second principal component are large, indicating that fertilization is an indicator of the value of the second principal component. The third principal component is the largest factor with a mortality factor of X5, so the third principal component.

Table 3. Variance Proportion and Cumlative Variance Proportion

	Eigenvalue	Std.deviation	Proportion	Proportion cumulative(%)
P1	5.043	3.325	0.561	56.1
P2	1.723	0.858	0.192	75.3
P3	0.857	0.121	0.097	84.9
P4	0.744	0.317	0.084	93.2
P5	0.426	0.296	0.045	97.4
P6	0.133	0.072	0.016	99.2
P7	0.061	0.057	0.006	100.0
P8	0.003	0.003	0.000	100.0
P9	0.000	0.000	0.000	100.0

Table 4. Eigenvectors of correlation matrix of main factors

	X1	X2	X3	X4	X5	X6	X7	X8	X9
P1	0.433	0.425	0.432	0.299	0.208	-0.424	-0.209	0.295	0.082
P2	0 108	0 098	0 103	-0 182	-0 076	-0 096	0 645	-0 087	0 704
P3	-0.094	-0. 106	-0. 106	-0.332	0.859	0. 117	-0.016	0.305	0. 125
P4	0. 112	0. 165	0.063	-0.572	-0.412	0. 165	-0.003	0.645	-0. 141
P5	-0.221	-0.271	-0.221	0.592	-0. 127	0.081	0.052	0.611	0.282
P6	0.014	0.096	-0.003	0. 191	0. 155	-0.067	0.734	0. 125	-0.613
P7	0.227	0.285	0. 196	0.237	0.077	0.871	0.021	-0.081	0.056
P8	0.072	-0.679	0.714	-0.033	-0.012	0.062	0.051	0.027	-0.061
P9	-0.822	0.363	0.437	0.003	0.004	-0.008	-0.015	0.014	0.025

value is considered mortality. The coefficients of birth rate X4 and fertilization ratio X8 in the fourth principal component are large, and the birth rate and fertilization ratio are closely related, indicating that birth is an indicator of the fourth principal component value.

The factor loads of each influencing factor were obtained by factor analysis, and it can be seen from Table 3 and Table 4 that the first principal component represents the activity level and the role of the environment, and the activity level X1 and protein richness X3 have the largest load. Therefore, activity level and protein abundance are the main influencing factors; The load for life X6 is -0.950, indicating that lifespan is the main limiting factor in bee population growth, and the impact on bee population dynamics is up to 56.0%.

The second principal component has a maximum load of 0.925 at the unfertilized ratio X9; Therefore, the unfertilized rate is the main influencing factor.

The third principal component is determined by mortality X5, which has a mortality load of 0.798. The fourth principal component fertilization ratio X8 has a load of 0.556, the load of birth rate X4 is -0.492, indicating that birth conditions are also a limiting factor for bee population growth.

Table 5. Weight Values

	Factor 1	Factor 2	Factor 3	Factor 4
X1	0.969	0. 141	-0.088	0.095
X2	0 956	0 130	-0 099	0 143
X3	0.972	0. 134	-0. 100	0.056
X4	0.669	-0.242	-0.307	-0.492
X5	0.465	-0. 103	0.798	-0.355
X6	-0.950	-0. 124	0. 110	0. 141
X7	-0.463	0.845	-0.016	-0.002
X8	0.659	-0. 115	0.285	0.556
X9	0. 181	0.925	0. 115	-0. 121

6.3 Grey Relation Analysis(GRA)

6.3.1 Solve the Model Using GRA

Further grey correlation analysis was carried out on the monthly population of bees and the influencing factor Xi, and the results were as follows:

Table 6. Correlation coefficient

X1	X2	X3	X4	X5	X6	X7	X8	X9
0.5920	0.5901	0.5936	0.5920	0.5866	0.5852	0.5855	0.5932	0.5906

The order of association that can be obtained is:

$$X3 > X8 > X1 > X4 > X9 > X2 > X5 > X7 > X6$$

The abundance of protein has the greatest impact on the number of bee populations, followed by the fertilization rate(%), followed by activity level(%), birthrate(%)and unfertilized rate(%), and the least affected is lifespan. Comprehensive comparison found that the material conditions and birth conditions of the bee population were the main factors affecting the dynamics of the bee population. Part of the explanation given in the title: Due to the large workload in summer, resulting in shortened lifespan, a bee's activity level, pollen consumption and protein abundance affect its lifespan. The description of the topic coincides with the results of the data analysis.

6.3.2 Solving the Bee Pollination Problem

Problem 3 in this section. Simulate and predict how many bee hives will be needed to support pollination on a 20-acre (81,000 m2) plot of land containing crops that benefit from pollination. That is, solve for equation (13) in 5.3. First, the Holling model is applied to estimate the parameters using the least squares method, which can be obtained a = 0.8501, Th = 2.102×10^{-2}.

7. Sensitivity Analysis

Sensitivity analysis is a method of studying and analyzing the sensitivity of a system (or model) to changes in the state or output of a system (or model) to changes in system parameters or surrounding conditions. Sensitivity analysis is often used in optimization methods to study the

stability of the optimal solution when the original data is inaccurate or changes. Sensitivity analysis also allows you to determine which parameters have a greater influence on the system or model.

In the variance interpretation table, when the principal component is 4, the eigenroot of the variable interpretation is 1, and the contribution rate of the variable interpretation reaches 0.868, So there are 4 main factors, which is the same as the analysis results of the main influencing factors obtained in the previous article.

The gravel chart is a graph drawn according to the degree of interpretation of the data variation of each principal component, and its action is based on the slope of the decline of the characteristic value to confirm the number of factor principal components that need to be selected, combined with the variance interpretation table can be used to confirm or adjust the number of principal component factors, each principal component is a point, through the position of "slope tends to flat" to judge the number of extracted principal components, by observing the gravel map, you can get the same results above, so the model works well and passes sensitivity detection. In this Blog, we build a mathematical model and solve it to obtain the influence of different factors on the number of bee populations, and calculate the number of bees required for pollination in a certain space.

Table 7. Sensitivity Analysis

Principal Component	Eigenvalue	Percentage of variance	cumulation
1	5.885	42.035%	42.035%
2	2.214	15 814%	57 848%
3	1.243	8.880%	80.302%
4	0.731	5.221%	86.829%
5	0.256	1.829%	96.859%
6	0.091	0.648%	99.327%
7	0.063	0 451%	99 778%
8	0.023	0. 164%	99.942%
9	0.008	0.058%	100.0%

8. Conclusion

1. For problem1, we build a time series model to simulate the number of bee populations, which can simulate and predict the relationship between bee populations over time.

2. For problem2. we built a mathematical model that can well represent the impact of each influencing factor on population size. Then we used the methods of principal component analysis and grey association analysis to analyze the impact of various factors on population size, and the results of the analysis were: protein abundance had the greatest impact on bee population size, followed by fertilization rate (%), again activity level, birthrate and unfertilized rate, and the smallest impact was lifespan. Comprehensive comparison found that the material conditions and birth conditions of the bee population were the main factors affecting the dynamics of the bee population.

3. For problem3, we built a mathematical model of bee pollination, which can simulate the pollination process of bees. Under this model, we can calculate that approximately 87,500 bees would be needed for pollination on the 20 acres (81,000 square meters) required by the title.

4. For problem4, we give our thoughts and solutions in the attached page.

References

[1] Yang, XS. (2012). Flower Pollination Algorithm for Global Optimization. In: Durand-Lose, J., Jonoska, N. (eds) Unconventional Computation and Natural Computation. UCNC 2012. Lecture Notes in Computer Science, vol 7445. Springer, Berlin, Heidel- berg. https://doi.org/10. 1007/978-3-642-32894-7-27.

[2] Eickwort, G.C., Matthews, R.W., & Carpenter, J.E. (1981). Observations on the NestingBehavior of Megachile rubi and M. texana with a Discussion of the Significance of Soil Nesting in the Evolution of Megachilid Bees (Hymenoptera: Megachilidae). Journal of the Kansas Entomological Society, 54, 557-570.

[3] Bascompte, J., Jordano, P., & Olesen, J. M. (2006). Asymmetric coevolutionary net- works facilitate biodiversity maintenance. Science (New York, N.Y.), 312(5772), 431-433. https://doi. org/10. 1126/ science. 1123412.

[4] Kaiser-Bunbury, C.N., Muff, S., Memmott, J., Müller, C.B., & Caflisch, A. (2010). The robustness of pollination networks to the loss of species and interactions: a quantitative approach incorporating pollinator behaviour. Ecology letters, 13 4, 442-52 .

[5] Vázquez, D.P., Blüthgen, N., Cagnolo, L., & Chacoff, N. (2009). Uniting pattern and process in plant-animal mutualistic networks: a review. Annals of botany, 103 9, 1445-57.

[6] Jeger, M. J., Madden, L. V., & van den Bosch, F. (2009). The effect of transmission route on plant virus epidemic development and disease control. Journal of theoretical biology, 258(2), 198-207. https:// doi. org/ 10. 1016/j.jtbi.2009.01.012.

[7] Huang, Haomin and Mingxin Wang. The reaction-diffusion system for an SIR epidemic model with a free boundary. Discrete and Continuous Dynamical Systems-series B 20 (2015): 2039-2050.

[8] Ngonghala, C.N., Teboh-Ewungkem, M.I., & Ngwa, G.A. (2016). Observance of period-doubling bifurcation and chaos in an autonomous ODE model for malaria with vector demography. Theoretical Ecology, 9, 337-351.

[9] Dalal, N., Greenhalgh, D., & Mao, X. (2008). A stochastic model for internal HIV dynamics. Journal of Mathematical Analysis and Applications, 341, 1084- 1101.

[10] Dafni, A. (1984). Mimicry and Deception in Pollination. Annual Review of Ecology, Evolution, and Systematics, 15, 259-27.