ISSN 2616-5775 Vol. 8, Issue 3: 103-108, DOI: 10.25236/AJCIS.2025.080314

Wildlife Population Prediction Based on Time Series Models

Leyi Huang*

School of Information Science and Technology, Jinan University, 511443, Guangzhou, China
**Corresponding author: huangleyi450@gmail.com*

Abstract: *Wildlife population prediction is critical for conservation efforts as illegal wildlife trade and habitat loss continue to threaten global biodiversity, particularly for species like African elephants. Current research faces challenges in long-term forecasting due to data limitations and uncertainty. This study aims to forecast the population trends of endangered species over the next five years and specifically predict changes in African elephant numbers to provide insights for targeted conservation strategies. In this study, collect data from international statistical websites firstly, then establish ARIMA model to predict the change of the number of endangered animal species in the next five years, using Grey forecast model to predict the number of African elephants. The data in this paper suggest that the number of endangered species is increasing by about 5%, while the remaining number of African elephants fluctuates around 250,000, showing a downward trend in next five years. This study demonstrates the importance of time series models in wildlife population prediction. The findings provide a scientific basis for formulating targeted conservation policies, especially for African elephants. Future research should incorporate spatial distribution and environmental factors of different animals to enhance predictive accuracy. We should try our best to support practical conservation projects.*

Keywords: *Time Series, ARIMA, Grey Forecast, Quantity Prediction*

1. Introduction

Predicting future population changes of wildlife is crucial for global biodiversity conservation, as illegal wildlife trade and habitat loss pose significant threats to ecosystems. This study focuses on particularly vulnerable wildlife populations, such as African elephants. Accurate population forecasts can aid policymakers in devising targeted strategies to protect endangered species. In recent years, time series models [1,2] have gained popularity in addressing prediction problems, providing quantitative insights into population trends and their implications for conservation planning.

Many studies have employed advanced models to explore wildlife population dynamics. Traditional methods, such as linear regression and logistic models, have been widely applied but often fail to account for nonlinear trends and seasonal variations. In recent years, ARIMA and Grey models have emerged as effective tools for transient forecasting, addressing some of the complexities in population changes. The ARIMA model, a time series autoregressive technique, calculates short-term forecasts by analyzing historical time series data. It has been applied to past epidemiological predictions and the current pandemic [3]. Meanwhile, The Grey model has been proven effective for predictions with small sample sizes, such as forecasting carbon emissions [4]. However, these models are rarely combined to address complex problems, such as simultaneously forecasting multiple species or specific populations like African elephants.

This study collects recent population data of African elephants and innovatively integrates the ARIMA model with the grey prediction model, constructing a composite forecasting framework suitable for wildlife populations to predict population trends over the next five years. This approach breaks through the limitations of traditional single models in scenarios where nonlinear trends and seasonal fluctuations coexist. The integration of these models not only accommodates data analysis across different scales but also captures periodic and seasonal variations in the data. This approach serves as an effective solution for complex forecasting problems and provides technical support for implementing wildlife conservation policies. Furthermore, based on this research, additional factors influencing animal survival could be incorporated into future studies. By adjusting model parameters, the accuracy of predictions can be further improved.

ISSN 2616-5775 Vol. 8, Issue 3: 103-108, DOI: 10.25236/AJCIS.2025.080314

2. Model Principles

2.1 ARIMA Model

The ARIMA model (Autoregressive Integrated Moving Average Model) [5]is a statistical technique used for analyzing and forecasting time series data.

It consists of three parts: Autoregressive (AR), Differential (I), Moving Average (MA):

1) Autoregressive model: Under the premise that there is a linear relationship between the current value and the value at the historical moment, the regression relationship is fitted by linear equations to predict future values:

B is used as the lag operator in the model, which means that the time series is moved to a certain point in time: $BX_t = X_{t-1} \quad B^2X_t = X_{t-2}\cdots$

$$\Phi(B) = 1 - \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p$$

$$\Phi(B)X_t = \epsilon_t$$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \tag{1}$$

In this formula X_i: The value of the time series at each historical point in time,

ϕ_i: Represents the coefficient of each part,

ϵ_t Error term (effect of random unobservable factors other than the linear variable part)

2) Differential part: The time series is smoothed out by multiple differences, eliminating seasonal data. In the model, statistics that do not have temporal variability, such as mean and variance, are used

$$\begin{aligned}(1-B)X_t &= \Delta X_t = X_t - X_{t-1}\\ (1-B)^2X_t &= \Delta^2 X_t = \Delta X_t - \Delta X_{t-1} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})\end{aligned} \tag{2}$$

$(1-B)^d X_t$: The formula representing the d difference, the specific value is determined according to the stationarity requirements of the time series

3) Moving average part: Considering the error term part of each moment in history, q is the order of the moving average, which means that the random fluctuation of the time series can be effectively handled in combination with the prediction error of the past q time

$$\begin{aligned}\Theta(B) &= 1 + \theta_1 \mathrm{B} + \theta_2 \mathrm{B}\text{^}2 + \cdots + \theta_q \mathrm{B}^q\\ \Theta(B)\cdot\epsilon_t &= \mu + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q} + \epsilon_t\end{aligned} \tag{3}$$

Combining the above three parts, it involves the complete formulation of the formula:

$$\Phi(B)\cdot(1-B)^d X_t = \Theta(B)\cdot\epsilon_t$$

ARIMA(p,d,q) : The equation has three parameters, which are the order of the autoregressive part, the number of the difference part, and the order of the moving average part. Such models can process data with seasonal and cyclical variations, improving the accuracy of forecasts

2.2 Grey Forecast Model

The grey prediction model [6]is a method for predicting uncertain and incomplete information, which is mainly suitable for the processing and analysis of small-scale sample data. Finally, differential equations are established to predict future trends.

At present, gray prediction models are divided into continuous, discrete, univariate, multivariate and other linear models, which are suitable for different types of problems. The processing method using additive operations is a special feature of this model.

The question in this article uses the GM (1,1) model, and the basic form is as follows

Raw data sequences:

$$X^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\right\}$$

Accumulate to generate sequences:

ISSN 2616-5775 Vol. 8, Issue 3: 103-108, DOI: 10.25236/AJCIS.2025.080314

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$$

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1,2,\dots,n.$$

To better consider the regularity of the data, a series of mean values can be used:

$$Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}$$

$$z^{(1)}(k) = \frac{1}{2}\left(x^{(1)}(k) + x^{(1)}(k-1)\right), k = 2,3,\dots,n. \tag{4}$$

Establish the basic form of the model:

$$x^{(0)}(k) + az^{(1)}(k) = b$$

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)\mathrm{e}^{-ak} + \frac{b}{a}, k = 1,2,\dots,n-1 \tag{5}$$

Or use differential equations:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{6}$$

where a is the development coefficient and b is the gray effect.

To simplify the calculation, it is now converted to matrix form:

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \Theta = \begin{bmatrix} a \\ b \end{bmatrix}$$

Parameters can be estimated by using least squares to minimize the cost function

$$\widehat{\Theta} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^{\top}B)^{-1}B^{\top}Y \tag{7}$$

Finally, the predicted value can be obtained by bringing it into the formula:

$$Y = B \cdot \Theta \tag{8}$$

3. Results

3.1 The change rule of endangered animal population based on time series analysis

First of all, before solving the model, the normality test was performed on the original data to ensure the accuracy of statistics.

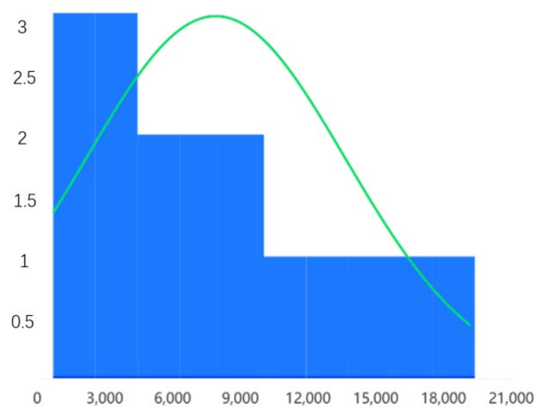

Figure 1. Histogram of normality tests

ISSN 2616-5775 Vol. 8, Issue 3: 103-108, DOI: 10.25236/AJCIS.2025.080314

Based on the above figure1 normality test histogram, the normal plot basically shows a bell shape (high in the center and low at the ends), which is basically acceptable as a normal distribution.

- Step1: Plot the Quantile-Quantile (figure2 Q-Q plot) to check whether it matches the law of normal distribution.

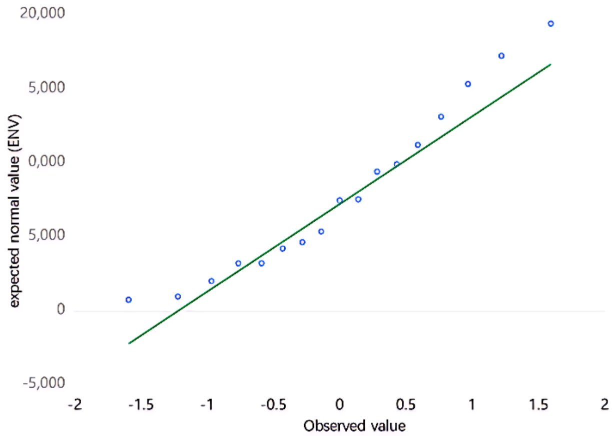

Figure 2. Q-Q chart

- Step2: Compute descriptive statistics of the data to get more information about the distribution of the data.
- Step3: Apply normality test statistics, Shapiro-Wilk test is used due to the number of samples N<5000.
- Step4: The significance p-value of 0.265 is obtained, the level does not present significance, therefore the data satisfies the normal distribution.

We chose ARIMA model to predict the number of endangered animals in the next five years.

The steps of model building and analysis are as follows:

- Step1: Conduct the smoothness test on the time series, the obtain d=1.
- Step2: The time series were also made for figure3 auto-correlation analysis plots and their p and q values were calculated:

$$p: Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \xi_t$$
$$q: Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \quad (9)$$

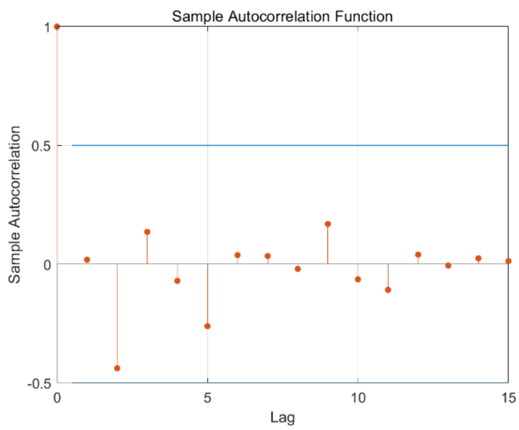

Figure 3. Auto-Correlation Function

- Step3: Analyze it in combination with the information criterion AIC and BIC values, the smaller the AIC and BIC values indicate that the better the model is; check the degree of fit of the model to the sequence R^2, the closer it is to 1 indicates that the model is more effective.

ISSN 2616-5775 Vol. 8, Issue 3: 103-108, DOI: 10.25236/AJCIS.2025.080314

$$AIC = -2 \cdot \ln(L) + 2 \cdot (p + q + 1)$$
$$BIC = -2 \cdot \ln(L) + \ln(n) \cdot (p + q + 1) \quad (10)$$

The model fitting and prediction plots are obtained as follows figure4:

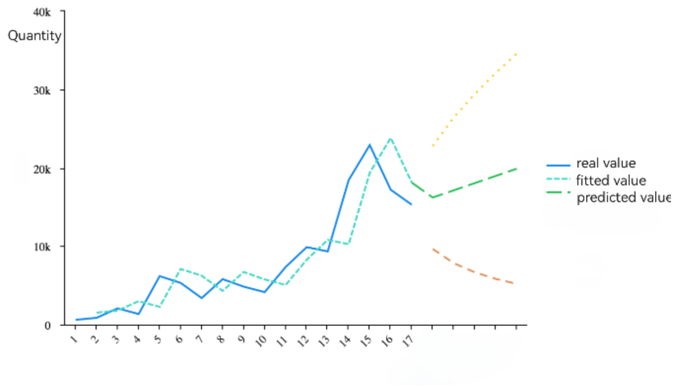

Figure 4. Time Trend Chart

Through the residual term LM test, we obtained a p-value of 0.04 for the endangered animal population statistic, implying the existence of serial correlation. Meanwhile, the prediction of endangered animal population for the next five years was obtained in Table 1:

Table.1. Forecasts of the number of endangered species

Years	Quantity
2023	16214
2024	17132
2025	18050
2026	18968
2027	19886

3.2 Taking the ivory trade as an example: Trends in quantity change based on gray prediction

We have chosen a typical example of illegal wildlife trade, the ivory trade, to reflect more specifically the survival of a certain type of wildlife, and thus to visualize the severity of illegal animal trade.

Modeling and Solving:

1) When using the GM (1, 1) model to predict the future number of African elephants, we need to first test the degree of fit of the GM (1, 1) model to the original data, first of all, to carry out the level deviation test:

Firstly, the grade ratio $\sigma(k)$of the original data is calculated from $x^{(0)}(k-1)$and $x^{(0)}(k)$:

$$\sigma(k) = \frac{x^{(0)}(k)}{x^{(0)}(k-1)}(k = 2,3,\cdots,n) \quad (11)$$

The corresponding grade ratio deviation and average grade ratio deviation are then calculated from the predicted development factor:

$$\eta(k) = \left|1 - \frac{1-0.5\hat{a}}{1+0.5\hat{a}}\frac{1}{\sigma(k)}\right|(\eta), \bar{\eta} = \frac{\sum_{k=2}^{n}\eta(k)}{n-1} \quad (12)$$

Calculations yielded a result of 0.064907 for the rank-order deviation test, which showed that the model fit the original data very well.

2) Derive the Grey Forecasting results in MATLAB, i.e., the predicted data of the African elephant population in the next five years:

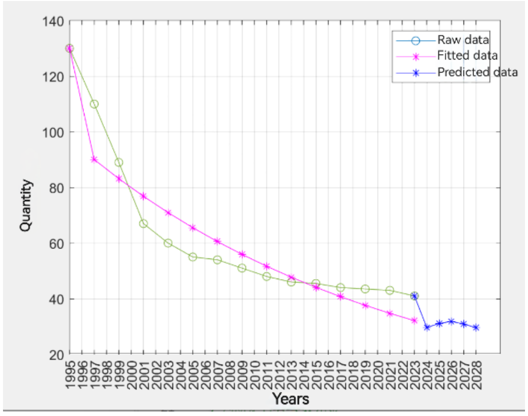

Figure 5. Gray prediction result chart

As can be seen from the Figure 5 above, African elephant populations globally are facing a continuing downward trend that shows no sign of slowing in the next five years.

4. Conclusions

In this paper, a time series model is established for the analysis and prediction of wild animal population. The results show that the number of endangered species is increasing, the remaining number of African elephants is decreasing.

However, there are still several aspects worthy of further research and exploration in the future. Firstly, changes in population size are influenced by a variety of factors, including environmental and socio-economic elements, which need to be incorporated to optimize the predictive capabilities of the model. Secondly, with advancements in technologies such as machine learning, the model's ability to capture non-linear relationships can be further enhanced. Lastly, research could also involve acquiring larger datasets and applying big data methodologies to address these issues, thereby improving the model's generalizability and practical utility. It is hoped that in the future, more precise data support can be provided for the conservation of wildlife.

References

[1] Zhang, G. Peter. "Time series forecasting using a hybrid ARIMA and neural network model" [J]. Neurocomputing, 2003, 50: 159-175.

[2] Lim, Bryan, et al. "Time-series forecasting with deep learning: a survey" [J]. Philosophical Transactions of the Royal Society A, 2021, 379(2194): 20200209.

[3] Ospina et al. "An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: Methodology and case study in Brazil" [J]. Mathematics, 2023, 11(14): 3069.

[4] Ye, Li, et al. "An enhanced multivariable dynamic time-delay discrete grey forecasting model for predicting China's carbon emissions" [J]. Energy, 2022, 249: 123681.

[5] Shumway, Robert H., et al. ARIMA models. Time series analysis and its applications: with R examples [M]. 2017: 75-163.

[6] Lee C W, Kong P. Forecasting Travel Sentiment under the Shock Effects of COVID-19 and Vaccination Using Grey Prediction Models[J]. Journal of Mathematics, 2023. DOI:10.1155/ 2023/ 7334544.