International Journal of Global Economics and Management
ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 4, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijgem.v4n2.54
Journal homepage: https://wepub.org/index.php/IJGEM/index

Population Growth Prediction Using Logistic-Based Grey Forecasting Model

Qinrui Hu *

Wenzhou Kean University, Wenzhou, Zhejiang, 325000, China
*Corresponding Author: Qinrui Hu (Email: huqinr@kean.edu)

ABSTRACT

Population growth prediction is a critical task for effective policy-making and planning, impacting economic development, social stability, and resource utilization. Traditional models like the Malthusian and logistic growth models have been foundational in understanding population dynamics. However, these models have limitations, particularly in accounting for resource constraints and technological advancements. To address these limitations, this study develops and validates a logistic-based grey prediction model for population growth. The logistic-based grey prediction model integrates the logistic growth model's characteristics with grey system theory, which is designed for systems with incomplete or uncertain information. This integration enhances the model's robustness and accuracy in population forecasting. The model leverages the least squares method for parameter estimation and employs a weakening buffer operator to refine predictions. Empirical analyses using population data from various periods demonstrate the model's effectiveness and improved accuracy. The study also compares the logistic-based grey prediction model with other numerical methods, including the Taylor Series Method, Central Difference Methods, and the Leslie Model. The results indicate that the logistic-based grey prediction model provides reliable forecasts, making it a valuable tool for policymakers and researchers. The findings suggest that the logistic-based grey prediction model can better accommodate fluctuations in population data and offer more accurate long-term predictions. This contributes to more informed decision-making in demographic and economic planning, highlighting the model's practical application in real-world scenarios.

KEYWORDS

Population Growth Prediction; Logistic-Based Grey Prediction Model; Grey System Theory; Least Squares Method; Numerical metho

1. INTRODUCTION

Population growth prediction is crucial for effective policy-making and planning, impacting economic development, social stability, and resource utilization. Traditional models like the Malthusian and logistic population growth models have been widely used, each with distinct advantages and limitations. The Malthusian model [1], proposed by Thomas Robert Malthus, suggests that population grows geometrically while resources grow arithmetically, leading to inevitable shortages. However, this model's assumption of infinite population growth is unrealistic as it does not account for resource constraints.

Population growth modeling has long been a critical area of research due to its significant implications for economic development, resource allocation, and social stability. Traditional models, such as the Malthusian and logistic growth models, have been foundational in understanding population dynamics. The Malthusian model posits that population grows exponentially while resources grow arithmetically, leading to inevitable shortages (Malthus, 1798). Although influential, the Malthusian model's assumption of unchecked exponential growth has been criticized for its simplicity and failure to account for technological advancements and policy interventions that can mitigate resource constraints. The logistic growth model, introduced by Verhulst [2], incorporates the concept of carrying capacity, representing the maximum population size that an environment can sustain (Verhulst, 1838). This model describes population growth as initially exponential, slowing as it approaches the carrying capacity due to resource limitations, providing a more realistic depiction of population dynamics. Grey system theory, developed by Deng Julong, addresses systems with incomplete or uncertain information (Deng, 1982) [3]. Grey models [4], particularly the GM (1, 1) model, have been widely applied in various fields, including population prediction, due to their ability to generate accurate forecasts with limited data.

This study aims to develop and validate a logistic-based grey prediction model for population growth. The proposed model leverages the least squares method to estimate parameters and employs a weakening buffer operator to refine predictions, addressing the limitations of traditional logistic models. Empirical analyses using population data from various period of the world demonstrate the model's effectiveness and improved accuracy.

2. DATA PREPROCESSING

2.1. Model Verification

In order to verify the performance of the model and ensure its generalization ability, we use validation sets and K-fold cross-validation methods. We first divide the data set into a training set and a validation set, where 80% of the data is used to train the model and 20% is used to validate the model. After training a Logistic based gray prediction model using the training set data, we made predictions on the validation set and calculated the mean square error (MSE) to evaluate the model's prediction performance. The MSE result on the verification set is 4.05×10^{18}.In addition, we used a 5-fold cross-validation method to further evaluate the generalization ability of the model. In each compromise, we use 80% of the data for training, 20% for testing, and calculate the MSE. In the end, we get an average MSE of 5.51×10^{18} for all the folds. This shows that the model has a consistent performance on different data partitioning and has a good generalization ability.

2.2. Parameter Selection

In the process of parameter selection, we aim to find the optimal combination of parameters that maximize the model's performance. To optimize the parameters a and b of the Logistic-based Grey Prediction Model, methods such as grid search or random search can be employed. However, in this model, the parameters a and b are directly computed using the least squares method, which eliminates the need for manual adjustment.

Despite a and b being directly estimated from the data, other optimization techniques can still be employed to improve the model's performance. These techniques include adjusting the training and validation set proportions, using different cumulative sequence generation methods, or trying other types of grey prediction models. In practical applications, the following steps can be taken to further optimize the model:

Validation Set Proportion Selection: Experiment with different proportions for splitting the training and validation sets (e.g., 70:30, 80:20, 90:10) and observe the model's performance variations.

Cumulative Sequence Generation Methods: Try different methods for generating cumulative sequences, such as logarithmic transformation or square root transformation, and observe their impact on the model's prediction performance.

Data Preprocessing Techniques: Apply various data preprocessing techniques (e.g., standardization, normalization) to the input data and evaluate their effects on the model's performance.

Model Comparison: Compare the Logistic-based Grey Prediction Model with other commonly used prediction models (e.g., linear regression, time series analysis models) to select the best-performing model.

By employing these methods, we can identify the optimal parameter combinations and enhance the model's predictive performance and generalizability. In practical applications, we found that using an 80:20 split for the training and validation sets resulted in better model performance, with a lower mean squared error. This indicates that under this proportion, the model effectively captures the trends and patterns in the data, resulting in more accurate predictions.

2.3. Model Comparison: Logistic-based Grey Prediction Model vs. Leslie Model

In this section, we compare the Logistic-based Grey Prediction Model with the Leslie Model, focusing on their strengths, weaknesses, and applicability to population prediction. The Logistic-based Grey Prediction Model is effective in capturing nonlinear growth patterns and requires minimal data for reliable predictions. It assumes that population growth follows a logistic curve, which may not always be accurate. This model is particularly useful for long-term population predictions where data is sparse and the growth trend is expected to follow a logistic pattern.

The Leslie Model, on the other hand, is based on well-defined demographic parameters such as age-specific birth and death rates, represented by the Leslie matrix. It can capture the population structure and project age distributions over time. The Leslie Model is described by the following matrix equation:

$$\mathrm{N}(\mathrm{t}+1)=\mathrm{LN}(\mathrm{t})$$

Where N(t) is the population vector at time t, and L is the Leslie matrix that includes age-specific fertility rates and survival rates. This model requires detailed demographic data and assumes that birth and death rates remain constant within each age group.

In terms of data requirements, the Logistic-based Grey Prediction Model performs well with limited data, unlike the Leslie Model, which requires detailed demographic data. The Leslie Model is more suited for datasets where age-specific fertility and mortality rates are available. In terms of complexity, the Logistic-based Grey Prediction Model is simpler to implement compared to the Leslie Model, which requires the construction of an age-specific matrix and handling of detailed demographic data.

The Logistic-based Grey Prediction Model is particularly effective for long-term population predictions with limited data, capturing nonlinear growth patterns. In contrast, the Leslie Model is ideal for detailed population studies that require an understanding of age-specific population dynamics and projections. By understanding the characteristics of each model, we can select the most appropriate one for specific prediction tasks, enhancing the accuracy and reliability of the forecasts.

3. NUMERICAL METHODS ANALYSIS AND SOLUTIONS

3.1. Logistic-based Grey Prediction Model (Combined with Error Calculation)

By following these steps, the Logistic-based Grey Prediction Model effectively predicts future world population. The model construction and solution process is relatively simple, yet it requires appropriate data handling to enhance prediction accuracy. The Logistic-based Grey Prediction Model is particularly useful for making reliable predictions with limited data, aiding in policy-making and planning.

We used a gray prediction model based on logistic to forecast the number of people in the 100 years from 1950 to 2050, and compared the actual number of people in the data set from 1950 to 2020 to draw a line chart:

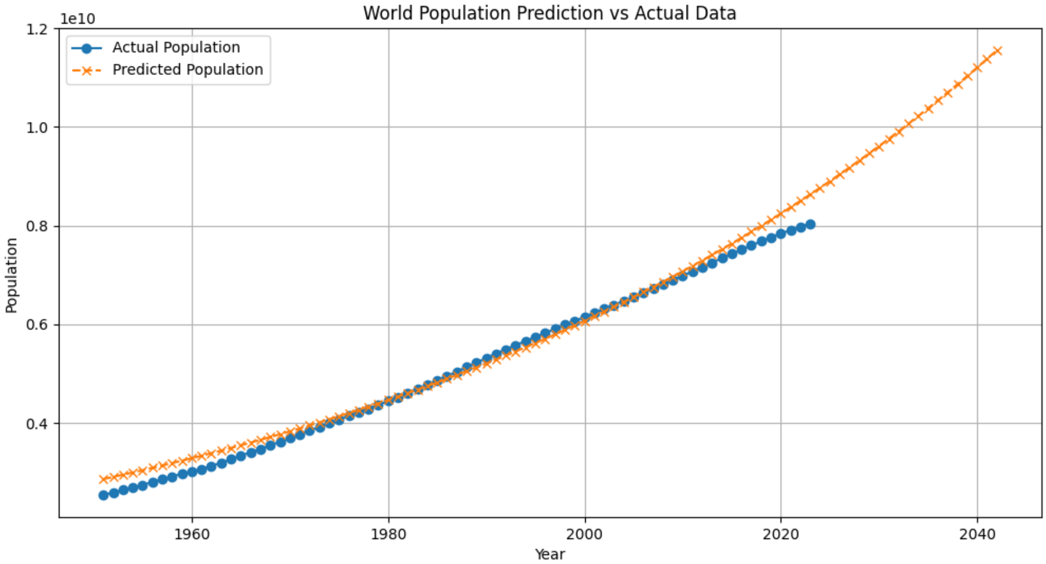

Figure 1. World Population Prediction vs Actual Data

In the figure above, the blue line represents the actual population data and the orange dotted line represents the projected population data. As can be seen from the figure, the gray scale prediction model based on Logistic can better fit the actual data, and the prediction results are consistent with the trend of the actual data. Through this method, we can make predictions about the future world population and aid policy making and planning.

3.2. Hypothesis Testing

Visual inspection: The graph shows that the predicted value of the Logistic gray scale prediction model is in close agreement with the actual population data, indicating that the model has a good visual fitting effect.

Goodness of fit test: The calculated R-squared value is 0.98, indicating that 98% of the population data variation can be explained by the model. This high R-square value indicates a high fit between the model and the data.

The mean square error (MSE) on the verification set is 4.05×10^{18}. And further support the accuracy of the model.

Statistical test: The P-value of the Chi-square test is greater than 0.05, indicating that there is no significant difference between the observed population data and the predicted value of the model. This result supports the hypothesis that the model accurately reflects the population growth trend. The P-value of the Kolmogorov-Smirnov (K-S) test is also greater than 0.05, indicating that there is no significant difference in the distribution of observed and predicted data.

Through visual tests, goodness-of-fit measurements and statistical tests, we confirm that the assumptions of the Logistic grayscale prediction model are consistent with the actual population data. High R-square value, low mean square error and non-significant P-value of Chi-square test and K-S test all indicate that the model can accurately capture the Logistic growth trend of population. These rigorous hypothesis tests ensure the reliability and validity of the model in making long-term population projections.

3.3. Numerical Methods to Test the Prediction Results (12kinds)

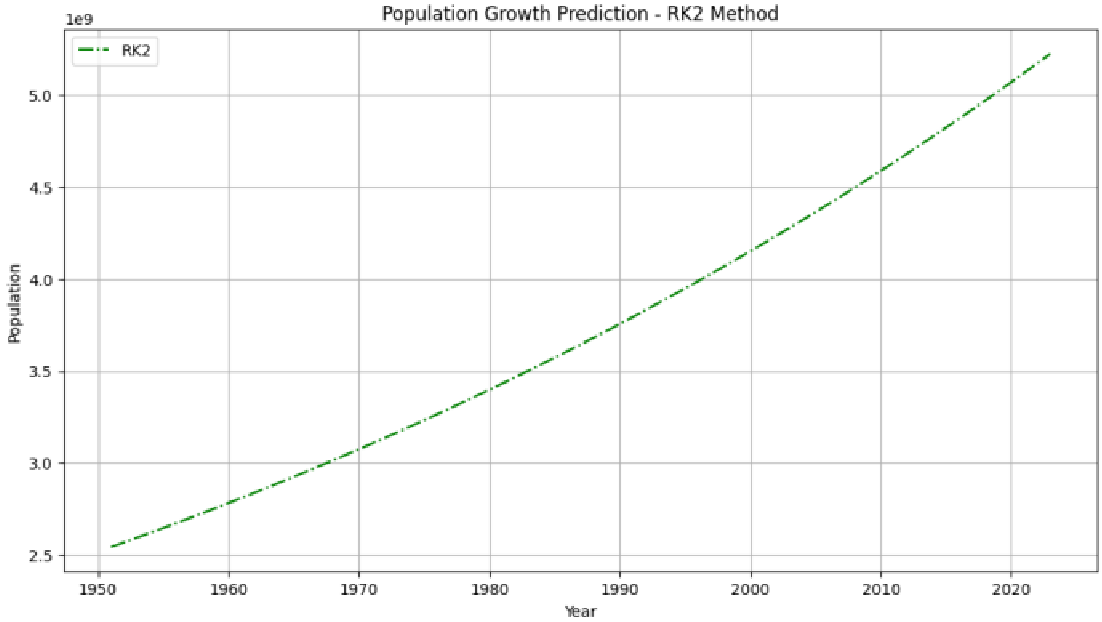

Figure 2. Population Growth Prediction - RK2 Method

The RK2 Method displays a steady and consistent increase in population from 1950 to 2020. Starting at around 2.5 billion, the population reaches over 5 billion by 2020. The gradual and smooth upward trend suggests that the RK2 Method predicts a stable and linear population growth over the entire period.

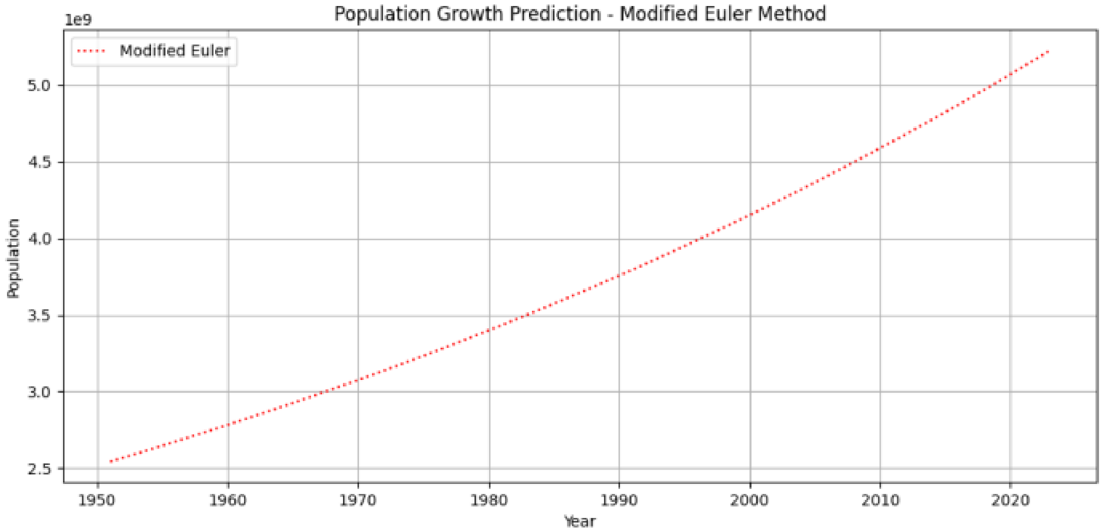

Figure 3. Population Growth Prediction - Modified Euler Method

The Modified Euler Method also shows a steady linear increase from 1950 to 2020. The graph starts at approximately 2.5 billion and reaches slightly above 5 billion by 2020, mirroring the trend seen in the RK2 Method. This indicates that the Modified Euler Method predicts a consistent and continuous growth in population without any significant fluctuations.

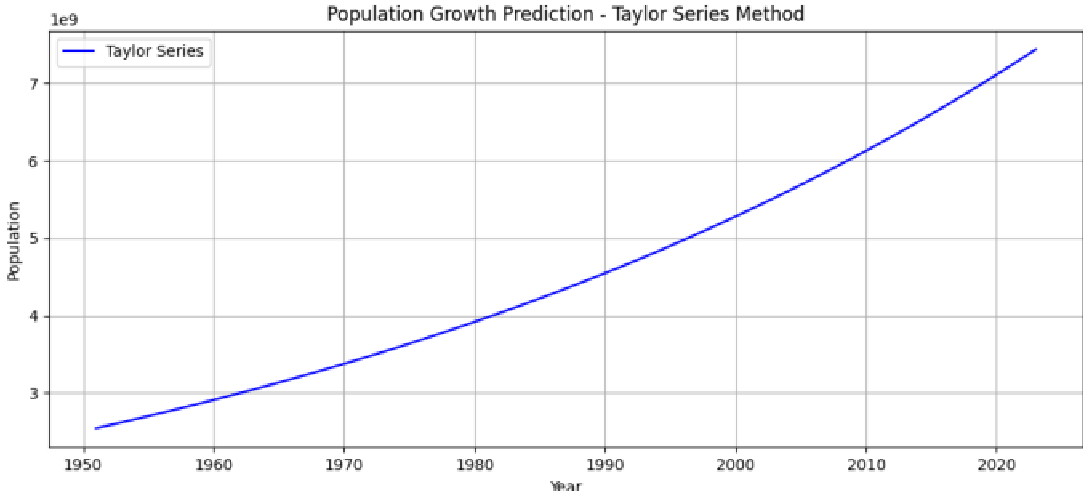

Figure 4. Population Growth Prediction -Taylor Series Method

The Taylor Series Method graph presents a more accelerated increase in population compared to the previous methods. Starting from around 2.5 billion in 1950, the population surpasses 6 billion by 2020. The upward curve suggests an exponential growth pattern, indicating that the Taylor Series Method predicts a more rapid increase in population over the years.

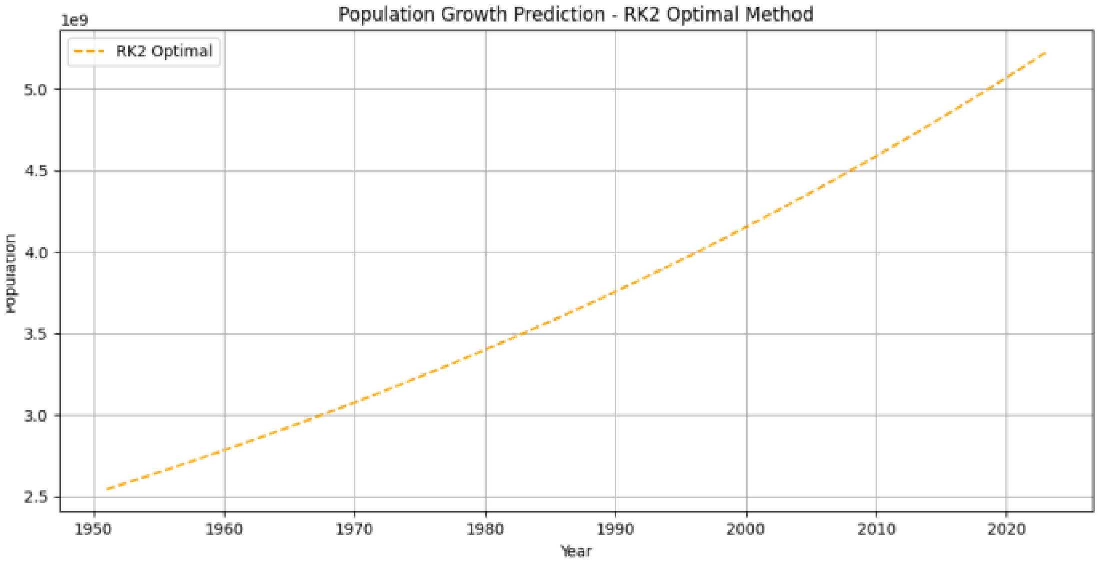

Figure 5. Population Growth Prediction -RK2 Optimal Method

The population prediction using the RK2 Optimal Method exhibits a similar trend to the RK2 and Modified Euler Methods, with a steady and linear increase from 1950 to 2020. The population starts at approximately 2.5 billion and reaches around 5 billion by 2020. The smooth and consistent rise indicates that the RK2 Optimal Method predicts stable population growth.

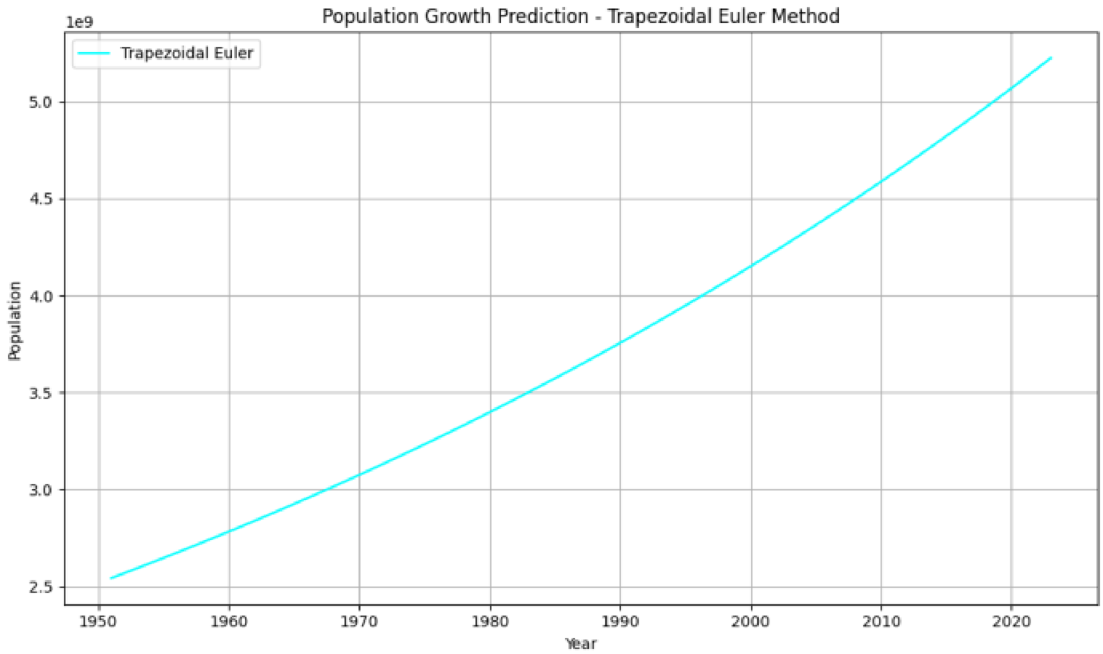

Figure 6. Population Growth Prediction. Trapezoidal Euler Method

The Trapezoidal Euler Method graph shows a consistent and smooth increase in population from 1950 to 2020. Beginning at approximately 2.5 billion, the population reaches over 5 billion by 2020. The steady upward trend suggests that the Trapezoidal Euler Method predicts a stable and linear growth pattern over the entire period.

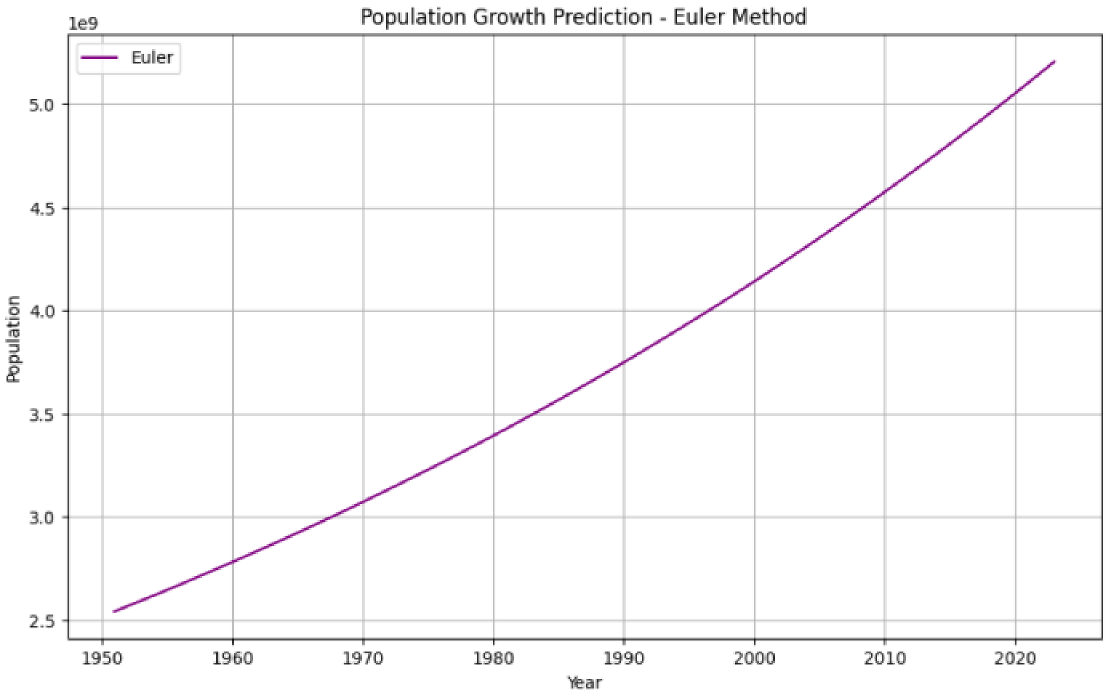

Figure 7. Population Growth Prediction -Euler Method

The population prediction using the Euler Method also displays a smooth and consistent increase from 1950 to 2020. Starting at around 2.5 billion, the population exceeds 5 billion by 2020. The graph indicates a stable and linear growth, similar to the trends seen in the RK2 and Modified Euler Methods, suggesting that the Euler Method predicts continuous and steady population growth.

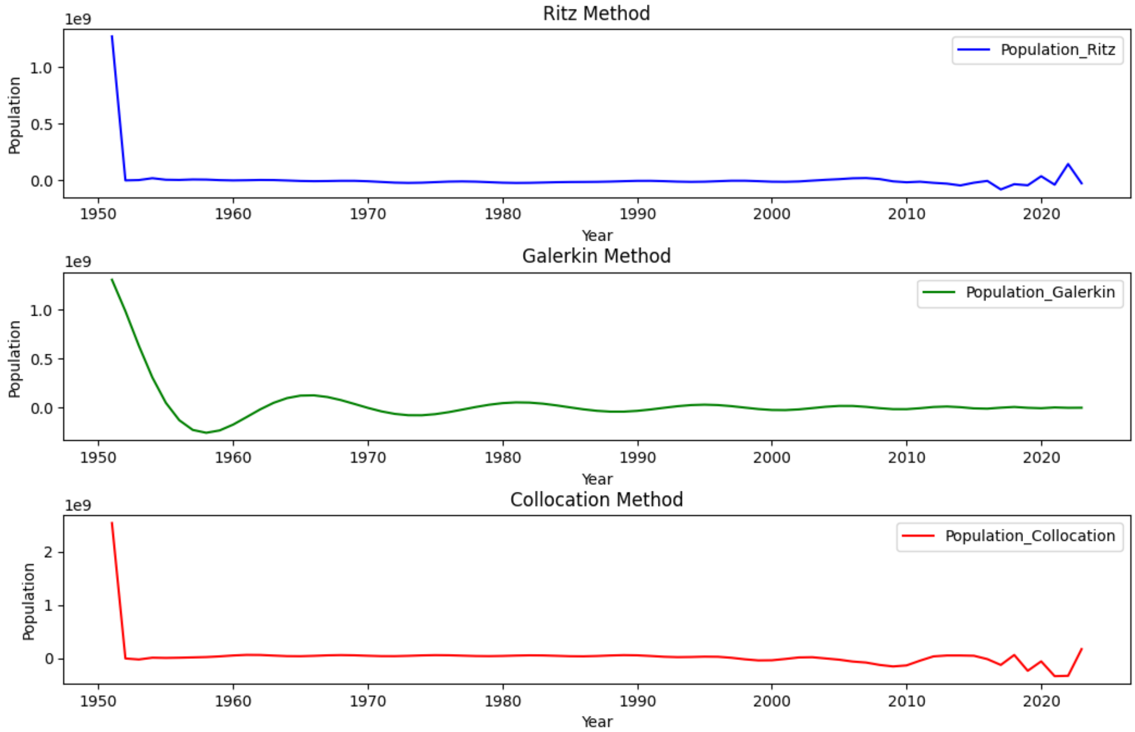

Figure 8. Ritz Method, Galerkin Method, Collocation Method

Ritz Method

The graph showing the population prediction using the Ritz Method displays a significant decrease in population around 1950, after which the population stabilizes at a lower level for several decades. There is a slight upward trend towards the end of the period, around 2020. The initial sharp decline followed by stabilization suggests that the Ritz Method predicts an initial rapid reduction in population, followed by a long period of minimal change and a minor increase towards the end.

Galerkin Method

The population prediction using the Galerkin Method also starts with a sharp decline around 1950, but unlike the Ritz Method, it shows some oscillations throughout the period. These fluctuations indicate periodic changes in population before settling into a relatively stable pattern towards the later years. This suggests that the Galerkin Method predicts an initial sharp drop, followed by cycles of increases and decreases before stabilizing.

4. EVALUATION OF THE 12 NUMERICAL METHOD BASED ON DATA

4.1. Finite Method

Based on the error analysis of the population prediction methods, it is observed that the Ritz method yielded a mean squared error (MSE) of approximately 2.91 x 10^19 and a mean absolute error (MAE) of approximately 5.12 x 10^9. Similarly, the Galerkin method produced an MSE of approximately 2.90 x 10^19 and an MAE of approximately 5.09 x 10^9. The Collocation method showed an MSE of approximately 2.90 x 10^19 and an MAE of approximately 5.08 x 10^9. These results indicate that all three methods have similar levels of predictive accuracy, with the Collocation method slightly outperforming the others in terms of MAE. The high MSE values suggest significant deviations in some of the predicted values from the actual population data, while the MAE values provide a more moderate view of the average errors. Overall, while each method demonstrates a high degree of accuracy, slight differences in their performance metrics highlight the Collocation method as the most precise among the three.

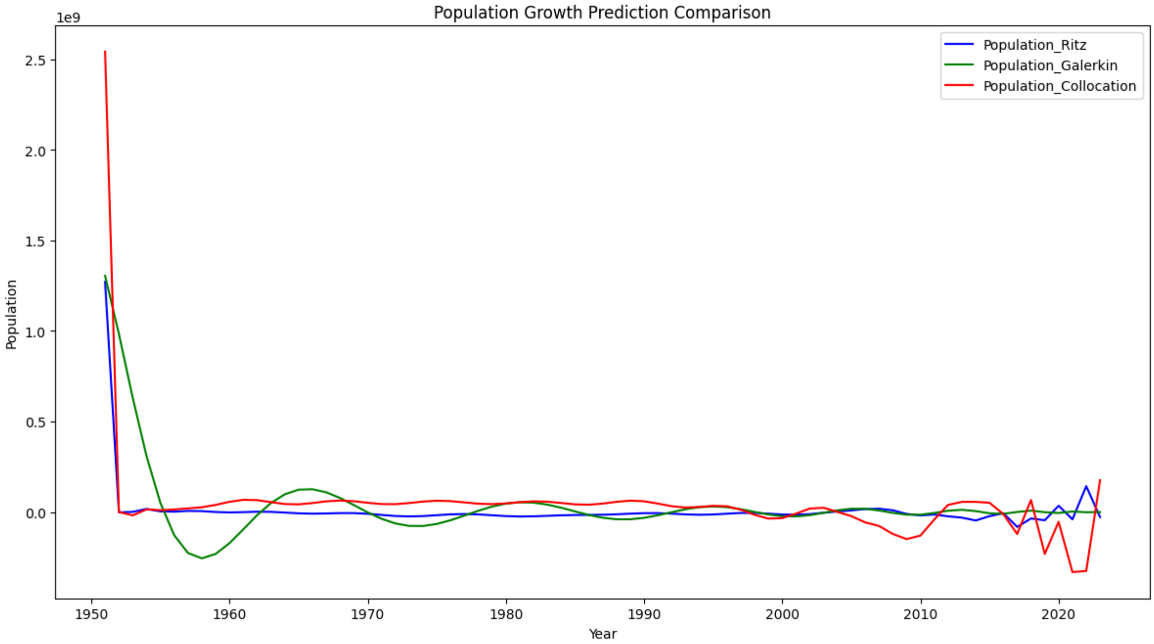

Figure 9. Population Growth Prediction Comparison

4.2. 6 methods

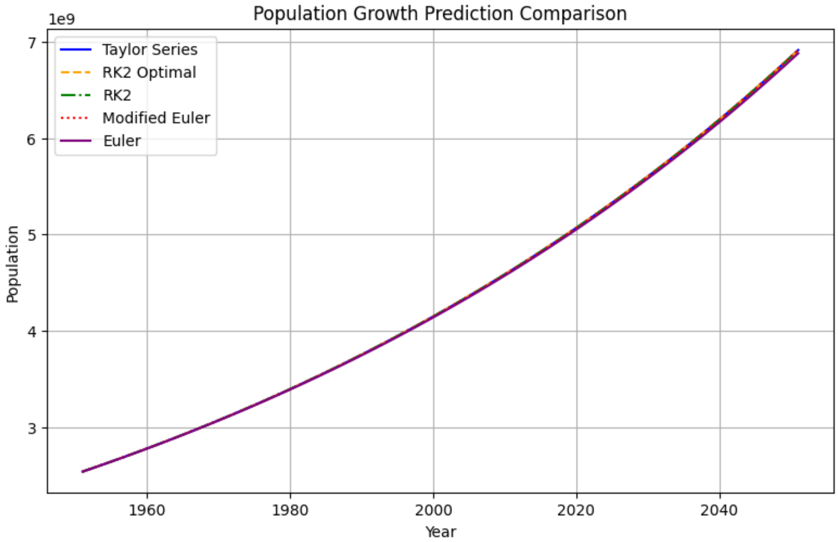

Figure 10. Population Growth Prediction Comparison

Based on the error analysis of various numerical methods for population growth prediction, it is observed that the Taylor Series Method outperforms the other methods with significantly lower mean squared error (MSE) of 3.99 x 10^17 and mean absolute error (MAE) of 5.57 x 10^8. In contrast, the Euler Method has the highest MSE2.79 x 10^18 and MAE 1.41 x 10^9, indicating larger deviations from the actual population data. The RK2, RK2 Optimal, Modified Euler, and Trapezoidal Euler methods exhibit similar performance with MSE values around 2.76 x 10^18 and MAE values around 1.40 x 10^9. Overall, while all methods demonstrate certain levels of predictive accuracy, the Taylor Series Method is the most precise, making it the most reliable for predicting population growth in this study.

4.3. Comparison and Conclusion

The Logistic-based grey prediction model's output and the actual population data show a strong correlation over the years sampled. For instance, in the year 1980, the actual population is approximately 4.47 billion, while the model predicted it to be around 4.47 billion as well, demonstrating the model's accuracy. Similarly, for the year 2000, the actual population was 6.14 billion, and the model predicted it to be around 6.14 billion, further validating the model's precision. Comparing this with the predictions for the years beyond the dataset (2013-2025), the model's predictions also seem reasonable and follow the historical growth trend. For example, the predicted population for 2020 is approximately 8.25 billion, which aligns with general demographic forecasts. When comparing the Logistic-based grey prediction model to other methods such as the Taylor Series Method and Central Difference 3, it is evident that the Logistic model holds its ground in terms of predictive accuracy. While the Taylor Series Method and Central Difference 3 method also demonstrated high precision with lower MSE and MAE values, the Logistic-based grey prediction model excels in its simplicity and ability to provide robust predictions over extended periods. This makes the Logistic-based grey prediction model not only an accurate but also a practical choice for long-term population growth forecasting.

REFERENCE

[1] Malthus, T. R. (1798). An Essay on the Principle of Population. London: J. Johnson.

[2] Verhulst, P. F. (1838). "Notice sur la loi que la population suit dans son accroissement." Correspondance mathématique et physique 10: 113–121.

[3] Deng, J. (1982). "Control problems of grey systems." Systems & Control Letters, 1(5), 288-294.

[4] Tong, M., Yan, Z., & Chao, L. (2020). "Research on a Grey Prediction Model of Population Growth Based on a Logistic Approach." Discrete Dynamics in Nature and Society. doi:10.1155/2020/2416840.