Prediction of the Number of Pure Electric Vehicles Based on the Extended GM(1,1) Model

Xuyang Song¹, Yulu Wang*¹
¹ Business School of Shanghai Dianji University, Shanghai, China
*Corresponding Author: email: wangyl@sdju.edu.cn

Abstract: The gray system theory is used to study the changes in the number of pure electric vehicles nationwide. According to the data on the number of pure electric vehicles in the country from 2016 to 2020, the original time series prediction model is established based on the GM(1,1) model. Using MATLAB to establish the original GM(1,1), new information GM(1,1) and metabolic GM(1,1), and compare them, and select the metabolic GM(1,1) with the smallest sum of squared errors for model performance inspection and error analysis showed that the degree of fit was up to standard, and the predicted results were obtained. The prediction results show that the number of pure electric vehicles in my country will increase year by year from 2021 to 2023, and the number of pure electric vehicles in my country will exceed 10 million by the end of 2023.

1. Introduction
In recent years, in order to reduce energy consumption and ease the pressure of environmental pollution caused by energy use, the country is vigorously promoting the new energy automobile industry. According to the "New Energy Vehicle Industry Development Plan (2021-2035)"¹, pure electric vehicles account for the largest proportion of new energy vehicles. Compared with hybrid vehicles and fuel cell vehicles, pure electric vehicles can be said to be almost "zero pollution", with almost negligible noise generated. Moreover, the cost of using electric vehicles is low, only about one fifth of that of gasoline vehicles, and it has high energy conversion efficiency and easy maintenance. Some super large cities such as Guangzhou and Shanghai also have supporting policies in terms of licensing policies. In order to achieve the sustainable and high-quality development of my country's pure electric vehicles and related industries, it is necessary to further improve the infrastructure system, including the layout of the corresponding supporting facilities for pure electric vehicles, such as charging piles, fast charging stations, and substations. The construction of these infrastructure systems requires a more accurate prediction of the future possession of pure electric vehicles in order to make corresponding development plans, save costs, and even promote the high-quality sustainable development of the entire pure electric vehicle industry chain. Therefore, based on the GM(1,1) extended model, this paper realizes the prediction of the national pure electric vehicle ownership in 2020-2023 in MATLAB.

2. Overview of gray theory and its applications
Grey theory was founded by Chinese scholar Professor Deng Julong in the 1980s. It is a method specially used to study the uncertain system problem of "part of the information is known, part of the information is unknown"¹. New contributions in the field of research. After more than 30 years of development, a complete system has been basically established. The modeling system is based on the gray system predictive modeling method.
In terms of prediction, many scholars currently use methods such as BP neural network prediction model \cite{2}, Logistic growth retardation model \cite{3}, regression analysis prediction model \cite{4} and other methods to predict the development trend of various automobiles. Compared with traditional algorithms, the BP neural network prediction model requires a lot of data and is more complex. The regression analysis prediction model has the disadvantage of a large amount of data and a huge amount of calculation. Different from the above three, the gray prediction model only needs a small amount of data to make predictions, and when predicting sample data with obvious trends, the accuracy of the prediction results is very high.

The grey predictive modeling method is based on the GM(1,1) model, which is the most active, fruitful and widely used research branch in grey theory. It takes the "small sample, poor information" uncertainty system with "part of the information is known, and the part of the information is unknown" as the specific research object. Using information coverage, sequence generation can seek realistic laws from irregular data. Its characteristics are "Less data modeling"\cite{5} has been successfully applied to many industries in the national economy and people’s livelihood, such as population\cite{6}, resources\cite{7}, environment\cite{8}, economic forecasting\cite{9}, food production\cite{10} and gradually developed into a mainstream forecasting modeling method.

3. Grey forecasting GM (1, 1) model construction

3.1. Modeling principle
The grey system theory includes four types of models: GM(1,1), DGM(1,1), GM(1,N), and Verhulst. The most widely used model is the GM(1,1) model, which represents a For a first-order predictive model containing only one variable, the modeling steps are as follows:

3.1.1 Set the original sequence
Set sequence $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)\}$, and $X^{(0)}(k) >= 0, k = 1, 2, ..., k$.

3.1.2 Accumulate to generate a new sequence
Accumulate $X^{(0)}$ once, $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), i = 1, 2, ..., n$ and get the new sequence $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n)\}$

Let $Z^{(1)}$ be the mean value generation sequence of $X^{(1)}$,

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

$Z^{(1)} = \delta x^{(1)}(k) + (1 - \delta) x^{(1)}(k - 1), k = 2, 3, ... n$ and $\delta = 0.5$

3.1.3 Build differential equations
Establish the first-order linear differential equation of $X^{(1)}(k)$ for $X^{(1)}(k)$, that is using the differential equation to approximately describe the new sequence:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

In the formula: $a$, $u$ are the discrimination parameters, The effective interval of $a$ is (-2, 2), which is obtained by the least square method:

$$\hat{a} = \left(\frac{a}{u}\right) = (B^T B)^{-1} B^T Y,$$

and $B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$
3.1.4 Solve the equation
From the above, we can solve for \( \frac{dx^{(1)}}{dt} + ax^{(1)} = u \) to get:
\[
\hat{x}^{(1)}(k + 1) = \left( x^{(0)}(1) - \frac{u}{a} \right) e^{-ak} + \frac{u}{a}, \quad k = 1, 2, \ldots, n \]

3.1.5 Discrete and restore by difference
Discrete the function expressions \( \hat{x}^{(1)}(k + 1) \) and \( \hat{x}^{(1)}(k) \), then use the difference to restore the original sequence to obtain \( \hat{x}^{(0)}(k + 1) \) as follows:
\[
\hat{x}^{(0)}(k + 1) = \hat{x}^{(1)}(k + 1) - \hat{x}^{(1)}(k) = Ae^{-ak}, \quad k = 1, 2, \ldots, n, \quad A = (1 - e^{-a}) \left( x^{(0)}(1) - \frac{u}{a} \right) \]

3.2. Model test and error analysis
When using this model, you need to first check the fit of the GM(1,1) model to the original data. Here, two test methods are mainly used for comparison:

3.2.1. Residual test
Absolute residual: \( \varepsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), \quad k = 2, 3, \ldots, n \)
Relative residual: \( \varepsilon_r(k) = \frac{|x^{(0)}(k) - \hat{x}^{(0)}(k)|}{x^{(0)}(k)} \times 100\%, \quad k = 2, 3, \ldots, n \)
Average relative residual:
\[
\bar{\varepsilon}_r = \frac{1}{n-1} \sum_{k=2}^{n} |\varepsilon_r(k)| \]
If \( \bar{\varepsilon}_r < 20\% \), it can be considered that the model's fitting to the original data has reached the general requirements.
If \( \bar{\varepsilon}_r < 10\% \), it can be considered that the model fits the original data very well.

3.2.2. Step ratio deviation test
First calculate the level ratio \( \sigma(k) \) of the original data from \( x^{(0)}(k - 1) \) and \( x^{(0)}(k) \):
\[
\sigma(k) = \frac{x^{(0)}(k)}{x^{(0)}(k - 1)}, \quad k = 2, 3, \ldots, n \]
Then calculate the level ratio deviation and the average level ratio deviation from \( a \):
\[
\eta(k) = \left| 1 - \frac{0.5a}{1 + 0.5a \sigma(k)} \right|, \quad \bar{\eta} = \frac{1}{n} \sum_{k=2}^{n} \eta(k) / (n - 1) \]
If \( \bar{\eta} < 20\% \), it can be considered that the model's fitting to the original data has reached the general requirements.
If \( \bar{\eta} < 10\% \), it can be considered that the model fits the original data very well.

3.3. GM(1,1) extended model
The new information model and the metabolism model are models based on GM(1,1). The new information model is mainly to add the newly obtained data in the prediction process to the original data, and use this as a new data group to predict the next data. This process does not remove the first element of the original data.

The metabolic model is that the data predicted during the prediction process will be added to the original data, and at the same time, the first element of the original data will be removed.

The basic principles and modeling methods of the new information model and the metabolic model are similar to GM(1,1). The main difference is in the array. The original array sequence will change with the prediction. The specific manifestation is slightly different in the MATLAB code.
4. Case study

4.1. Model building

Based on the 2016-2020 national pure electric vehicle population statistics (Table 1), the national pure electric vehicle population in the next three years is predicted.

| Years | 2016 | 2017 | 2018 | 2019 | 2020 |
|-------|------|------|------|------|------|
| Quantity | 73   | 125  | 211  | 310  | 400  |

Table 1: my country’s possession of pure electric vehicles in the past five years

From the data in Table 1, establish the original data sequence:

\[ X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\} = \{73, 125, 211, 310, 400\} \]

Perform an accumulation to get a new data sequence:

\[ X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\} = \{73, 198, 409, 719, 1019\} \]

Using MATLAB software to perform subsequent operations can get the original GM(1,1) parameters:

\[ \hat{a} = \begin{pmatrix} a \\ u \end{pmatrix} = (B^T B)^{-1} B^T Y = \begin{pmatrix} -0.34557 \\ 95.4553 \end{pmatrix} \]

\[ \frac{u}{a} = -276.2257 \]

Then the prediction model of GM(1,1) can be obtained as:

\[ \hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{u}{a}\right) e^{-ak} + \frac{u}{a} = 349.2257 e^{0.34557k} - 276.2257 \]

4.2. The comparison with extended model

Based on the original GM(1,1) function in MATLAB, you can get new information GM(1,1) and metabolism GM(1,1) by making relevant changes, and compare the error square sum SSE of the three, and metabolism GM(1,1) The sum of squared errors of the model is the smallest, and the data predicted by the simulation is closest to the real data, so the model is selected for later prediction.
4.3. Performance test and error analysis

Figure 2 Relative residual error and grade ratio deviation

The average relative residual is:

$$\bar{\varepsilon} = \frac{1}{n-1} \sum_{k=2}^{n} |\varepsilon_k| = 0.069031$$

The results of the residual test show that: $\bar{\varepsilon} < 10\%$, the model fits original data very well.

The average grade ratio deviation is:

$$\bar{\eta} = \sum_{k=2}^{n} \eta_k / (n-1) = 0.11648$$

The results of the level ratio deviation test show that: $\bar{\eta} < 20\%$, the fit of model to the original data meets the general requirements.

4.4. Forecast result

Fitting to the original data and prediction results:

Table 2  Fitting of pure electric vehicles and forecasted possession

| Years | 2016 | 2017       | 2018        | 2019        | 2020        | 2021        | 2022        | 2023        |
|-------|------|------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Quantity | 73   | 144.158    | 203.665     | 287.7359    | 406.5104    | 574.3139    | 772.8736    | 1038.2417   |

Figure 3 Graph of fitted data and predicted data
4.5. Result analysis
From Table 2 and Figure 2 above, it can be found that the total number of pure electric vehicles in the country will continue to rise in the next three years. The growth rate of the total number of pure electric vehicles in the country in 2021 will be 43.6%, in 2022 it will be 34.6%, and in 2023, it was 34.2%.

However, according to the November 2020 operating data released by the China Electric Charging Infrastructure Promotion Alliance, from January to November 2020, the increase in charging infrastructure will be 0.32 million units. As of November 2020, the cumulative number of charging infrastructure (public + private) across the country was 1.539 million units, an increase of 31.1% year-on-year. At present, many places have certain deficiencies in electric vehicle charging facilities. In the future, my country’s pure electric vehicle charging infrastructure still needs to maintain a corresponding high-speed growth.

5. Summary and outlook
Based on the total number of pure electric vehicles in our country from 2016 to 2020, this paper uses gray system theory to establish the original GM(1,1) model, and then uses MATLAB software to expand new information GM(1,1) and metabolism GM(1,1) on the basis of the original model. Comparing with them, take the metabolism GM(1,1) with the smallest error square and SSE to perform model performance test and error analysis on the data from 2016 to 2020, and obtain $\xi$ and $\eta$ all indicate that the model fits the original data up to the standard, and finally the simulation prediction is performed to obtain the prediction result. The forecast results show that the number of pure electric vehicles in my country from 2021 to 2023 will show an upward trend with the growth of the years. This data forecast provides a scientific reference for the development of pure electric vehicles and related industrial chains in China. By the end of 2023, the number of pure electric vehicles in my country will exceed 10 million, while the current national charging infrastructure is only about 1.5 million, which is far from meeting the charging needs of pure electric vehicles. Therefore, the speed of my country's charging infrastructure layout also needs accelerate.

References
[1] Deng, J.L. (2005) Basic method of grey system: Chinese-English comparison [M]. Huazhong University of Science and Technology Press, Wuhan
[2] Shi, C.C. Liu, Y.H. (2020) Research on pure electric vehicle sales forecasting fusion of Markov and BP neural network[J]. Software Guide, 19(11): 50-53.
[3] Dai, X.Z. Wang, Yan. Peng, Z.P. Cheng, H.B. (2019) Car ownership forecast method based on double Logistic curve model[J]. Journal of Chongqing Jiaotong University (Natural Science Edition), 38(11): 21-26.
[4] Yang, B.R. (2017) Passenger car market prediction model based on multiple linear regression and BP neural network [D]. Huazhong University of Science and Technology.
[5] Deng, J.L. (1983) Review of Grey System[J]. World Science, (07):1-5
[6] Yang, L. Zhang, B.Q. Zhang, Z.X. (2020) Research on the population prediction of Hebei Province based on Logistic and GM (1,1) model[J]. Forum on Industry and Technology, 19(06): 37-38.
[7] Zhao, G.H. Chen, J.Y. Ye, N. (2020) Forecast of the development trend of forest resources in Fujian Province based on the GM (1,1) model[J]. East China Forest Manager, 34(02): 72-76.
[8] Zhang, E.L. Wang, Y.L. (2020) GM (1,1) model prediction of water environment quality of Dongfeng Canal in Zhengzhou City[J]. Journal of Yi bin University, 20(06): 95-98+103.
[9] Sun, M. Sun, W. (2021) Forecast of economic high-quality development prosperity based on GM (1,1) model[J]. Modern Marketing (Late Period), (02): 21-23.
[10] Zhang, Y.Y. Xu, W.K. (2019) Forecast of Grain Production in Heilongjiang Province Based on Grey GM(1,1) Model[J]. Natural Science Journal of Harbin Normal University, 35(03): 41-45.