Team Number:	apmcm2300871
Problem Chosen:	C

2023 APMCM summary sheet

Abstract

This article mainly studies the main factors of the development of new energy electric vehicles in China, predicts the development status of new energy electric vehicles in China, and studies the impact of the development of new energy electric vehicles, in order to promote further development of new energy electric vehicles.

For question one, firstly, this article collected relevant data and found that the development of new energy electric vehicles in China is mainly influenced by market demand, innovation capacity, infrastructure construction, and economic development (per capita GDP). Secondly, an AHP mathematical model was established to analyze the influencing factors, and the weights of these factors on the development of new energy electric vehicles in China were obtained, resulting in quantitative results.

For question two, firstly, this article collected industry development data on new energy electric vehicles in China. Secondly, ARIMA models, nonlinear regression models, and adaptive hybrid ARIMA nonlinear regression models were established and solved to successfully predict the development of new energy electric vehicles in China in the next decade. The conclusion was drawn that the development of new energy electric vehicles in China will steadily increase in the next five years and there will be a leap in development from the fifth to the tenth year.

For question three, firstly, this article collected sales and production data of new energy vehicles, as well as relevant data on traditional energy vehicles worldwide. Secondly, conducting the Shapiro Wilk test on the collected data revealed that one factor did not follow a normal distribution, while the other factor followed a normal distribution. Therefore, establishing a Spearman correlation analysis model for correlation analysis resulted in a conclusion that the correlation between automobile production and new energy vehicle production was not strong.

For question four, firstly, this article collects relevant data on boycott policies of various countries and China's new energy exports. Secondly, conducting the Shapiro Wilk test on the collected data resulted in the conclusion that the data does not follow a normal distribution. Therefore, a Kruskal Wallis H-test model was established to determine its correlation, and the conclusion was drawn that the boycott policy will to some extent affect China's export trend to that country. Finally, further analysis was conducted on the data.

For question five, firstly, this article collects relevant data and performs data preprocessing. Secondly, establish a regression analysis model for the relationship between new energy electric vehicles, ecological environment, and population, and use differential evolution algorithm to solve the model. Finally, with a fixed population, it can be concluded that the impact of new energy electric vehicles on the environment can be roughly divided into two stages

For question six, this article combines the conclusion of question five to write an open letter to citizens in order to promote the benefits of new energy electric vehicles and the contributions of the electric vehicle industry in various countries around the world.

Keywords: AHP, ARIMA model, Nonlinear Regression Model, Shapiro Wilk test, Spearman correlation analysis, Kruskal Wallis H-test, regression analysis model, Differential Evolution Algorithm.

Content

1.Introduction

1.1 Problem Background

With the increasingly severe environmental problems and the need for energy security, countries are increasing investment in the research and promotion of new energy vehicles, and new energy vehicles are gradually developing. Among them, new energy electric vehicles include electric buses and household electric vehicles with less than 7 seats. Due to their characteristics of low pollution, low energy consumption, and peak shaving power consumption, they have achieved rapid development in recent years. The Chinese government has also actively promoted the development of new energy electric vehicles and formulated a series of preferential policies, making them gradually become another symbol of China after the "China High Speed Railway". Therefore, studying the main factors of the development of new energy electric vehicles in China, predicting the development status of new energy electric vehicles in China, and studying the impact of the development of new energy electric vehicles have become important topics to promote further development of new energy electric vehicles.

1.2 Work

Based on the collected data, mathematical modeling is used in this article to solve the following problems:

(1) Analyze the main factors affecting the development of new energy electric vehicles in China, establish mathematical models, and describe the impact of these factors on the development of new energy electric vehicles in China.

(2) Collect industry development data on new energy electric vehicles in China, establish mathematical models, and describe and predict the development of new energy electric vehicles in China in the next 10 years.

(3) Collect data, establish mathematical models, and analyze the impact of new energy electric vehicles on the global traditional energy vehicle industry.

(4) Some countries have formulated a series of targeted policies to resist the development of new energy electric vehicles in China. Establish a mathematical model to analyze the impact of these policies on the development of new energy electric vehicles in China.

2.Problem analysis

2.1 Analysis of Question One

For question one, the question requires analyzing the main factors that affect the development of new energy electric vehicles in China, and establishing a mathematical model to describe the impact of these factors on the development of new energy electric vehicles in China. Firstly, collect relevant data, such as the historical sales volume of new energy vehicles, the number of patents related to new energy vehicles, the number of charging stations, etc., to identify the main factors affecting the development of new energy electric vehicles in China. Secondly, establish a suitable mathematical model and use a systematic evaluation method (Analytic Hierarchy Process) to analyze the influencing factors, in order to obtain the impact of these factors on the development of new energy electric vehicles in China.

2.2 Analysis of Question Two

For question two, the question requires collecting data on the development of China's new energy electric vehicle industry and establishing a mathematical model to describe and predict the development of China's new energy electric vehicles in the next 10 years. Firstly, the industry development status can be described based on the influencing factors considered in the first question. Secondly, since the historical data considered in this paper are mainly the annual data of 2011 and later, which leads to incomplete data volume, and the future development of China's new energy electric vehicles is affected by both past development and accidental events in the past, it is not suitable to directly use a single model to predict future data. So, it is possible to consider using hybrid models or integrated models (this article uses the Autoregressive Integrated Moving Average Model - Nonlinear Regression Model) to predict data for the next 10 years.

2.3 Analysis of Question Three

For question three, it requires collecting data, establishing mathematical models, and analyzing the impact of new energy electric vehicles on the global traditional energy vehicle industry. Firstly, collect relevant data. Secondly, to analyze the impact of new energy electric vehicles on the global traditional energy vehicle industry, the correlation between the sales of new energy vehicles and traditional energy vehicles can be explored. By conducting distribution tests on the data, corresponding correlation analysis models (Pearson correlation coefficient and Spearman correlation coefficient selected in this article) can be selected to discuss the mutual influence between the two.

2.4 Analysis of Question Four

For question four, it is required to establish a mathematical model and analyze the impact of policies formulated by some countries to resist the development of new energy electric vehicles in China on the development of new energy electric vehicles in China. Firstly, collect the boycott policies of various countries towards China's new energy vehicles and their implementation time, and compare the growth rate and quantity of China's exports of new energy vehicles to that country before and after the policy implementation. Secondly, to assess the impact of these policies on the development of new energy electric vehicles in China, Kruskal Wallis text can be applied to the classified data after conducting normality tests and homogeneity of variance tests

2.5 Analysis of Question Five

For question five, the question requires analyzing the impact of electrification of urban new energy electric vehicles (including electric buses) on the ecological environment, and assuming a population of 1 million cities, providing the calculation results of the model. Firstly, collect relevant data. Secondly, for the three factors of new energy electric vehicles, ecological environment, and population, a ternary nonlinear regression model can be established separately to solve the relationship between the three factors, and the parameters of the model can be solved based on differential evolution algorithm. Finally, after constructing a relationship model between the three factors, the relationship between new energy electric vehicles and the ecological environment can be obtained by fixing the urban population to 1 million people.

2.6 Analysis of Question Six

For question six, the title requires writing an open letter to citizens based on the conclusion of question five, promoting the benefits of new energy electric vehicles and the contributions of the electric vehicle industry in various countries around the world. Based on the conclusion of question five, combined with the advantages and disadvantages of the traditional energy vehicle industry, the benefits of the development of new energy electric vehicles for the ecological environment, and the convenience that the development of new energy electric vehicles brings to people's lives, a letter can be written to promote the benefits of new energy electric vehicles and the contributions of the electric vehicle industry in various countries around the world.

3. Symbol and Assumptions

3.1 Symbol Description

The symbol description is shown in Tab 1.

Tab 1 Symbol Description Table

Symbol	Implication

$de_{m,i}$	The decrease in trend caused by boycott policy m after the i-th year
yu_i	Normalized observations for year i
Inf_j	The impact of factor j on the development of new energy electric vehicles in China
$y_i = \{y_1, y_2, y_3 \Lambda\ , y_t\}$	Actual historical observations
$\hat{y}_{k,i} = \left\{\hat{y}_{k,1}, \hat{y}_{k,2}, \hat{y}_{k,3}, \Lambda\ , \hat{y}_{k,i}\right\}$	Predicted value of algorithm k
$wmape_k$	Prediction error of algorithm k
$\hat{y}_{k,i}$	The predicted value of algorithm k in the i-th period
$wmape_{k,j}$	The prediction error of algorithm k on factor j
y_t	Time series of historical observations
d	Order of difference
ε_t	Independent identically distributed white noise sequence with zero mean and constant variance
$\theta_0, \theta_1, \Lambda\ , \theta_0$	P+1 parameters to be estimated
y_{dt}	The multivariate linear functions of $y_{d(t-1)}, y_{d(t-2)}, \Lambda\ y_{d(t-p)}$ in the previous p-period and $\varepsilon_{d(t-1)}, \varepsilon_{d(t-2)}, \Lambda\ , \varepsilon_{d(t-p)}$ in the previous q-period,

3.2 Fundamental assumptions

(1) There will be no breakthrough progress in new energy vehicle technology during the forecast period;

(2) Other countries have not formulated covert policies to resist the development of new energy electric vehicles in China;

(3) The ecological environment will not experience significant fluctuations due to other factors;

(4) There will be no substitutes for new energy electric vehicles during the forecast period;

(5) The demand for cars in the market will not fluctuate significantly due to external factors.

4.Model building and Solutions

4.1 Problem One

4.1.1 Data collection and Preprocessing

In order to deeply analyze the main factors affecting the development of new energy electric vehicles in China, we conducted data collection from four key aspects: market demand, innovation capability, infrastructure construction, and economic development (per capita GDP). Specifically, we collected historical sales volume, number of related patents, number of charging stations, and per capita GDP of new energy vehicles from 2011 to 2022. And the data was preprocessed, and detailed data sources and statistical information can be found in the supporting materials.

4.1.2 Establishment of AHP model

Based on the data we have collected, we believe that the main factors affecting the development of new energy electric vehicles in China include sales volume, number of related patent applications, number of charging stations, and per capita GDP. To further investigate the impact of these factors on the development of new energy electric vehicles in China, we chose to use the Analytic Hierarchy Process (AHP) method.

The AHP method performs well in dealing with target systems with multi-level and cross evaluation indicators, especially in decision-making problems where the target values are difficult to quantify. AHP decomposes decision-making problems into different hierarchical structures, including overall objectives, sub objectives at each level, evaluation criteria, and specific alternative solutions. Then, by solving the eigenvectors of the judgment matrix, we obtain the priority weights of each element at each level relative to a certain element at the previous level. Finally, the final weight of each alternative solution to the overall goal is calculated step by step through the weighted sum method to determine the optimal solution.

- Establish a hierarchical structural model

To explore the impact of these factors on the development of new energy electric vehicles in China, only the target layer and criterion layer exist. The results of the AHP model constructed in this question are shown in the following figure (AHP framework diagram):

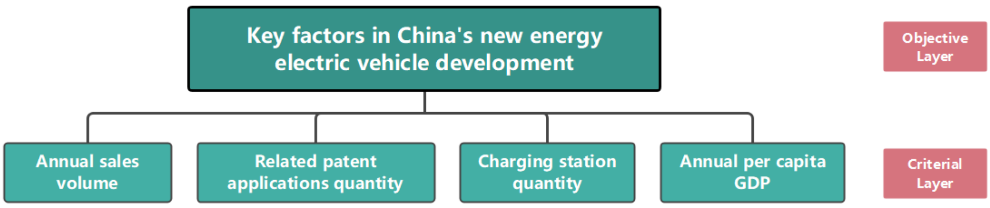

Figure 1 AHP Framework for Assessing Key Factors in China's NEV Development

As shown in the figure, this question constructs four factors to describe the development of new energy electric vehicles in China: sales volume, number of related patents, number of charging stations, and per capita GDP.

4.1.3 Solving of AHP model

- Construct a judgment matrix

When determining the weights between various levels and factors, if it is only a qualitative result, it is often not easily accepted by others. Therefore, Saaty et al. proposed the consistent matrix method, which does not compare all factors together, but compares them pairwise. At this time, relative scales are used to minimize the difficulty of comparing factors with different properties and improve accuracy. For a certain criterion, compare the various options under it pairwise and evaluate the level based on its importance. a_{ij} is the result of comparing the importance of element i and element j.

Tab 2 lists the four importance levels and their assigned values given by Saaty. The matrix formed by pairwise comparison results is called a judgment matrix. The judgment matrix has the following properties:

$$a_{ij} = \frac{1}{a_{ji}} \tag{1}$$

The scaling method for determining matrix elements is shown in Tab 2:

Tab 2 AHP 1-9 Degree Scale

Factor i / Factor j	Quantized Value
Equally important	1
Strongly important	5
extremely important	9
the median value of two adjacent judgments	2，4，6，8

- Hierarchical single sorting and consistency testing

The eigenvector corresponding to the maximum eigenvalue $\lambda_{\max}$ of the judgment matrix, after normalization (such that the sum of the elements in the vector is equal to 1), is denoted as W. The element W represents the ranking weight of the relative importance of a factor at the same level compared to a factor at the previous level. This process is called hierarchical single ranking. To confirm the hierarchical order, consistency testing is required, which refers to determining the allowable range of inconsistency for A. Among them, the unique non-zero eigenvalues of n-order uniform matrices are n;The maximum eigenvalue $\lambda \geq n$ of an n-order reciprocal matrix A is a consistent matrix if and only if it is $\lambda = n$.

Due to λ's continuous dependence on a_{ij}, the larger λ is compared to n, the more.The more severe the inconsistency of A, the consistency indicator is calculated using CI. The smaller the CI, the greater the consistency. Using the eigenvector corresponding to the maximum eigenvalue as the weight vector of the influence of the compared factor on a certain factor in the upper layer, the greater the degree of inconsistency, the greater the judgment error caused.Therefore, $\lambda - n$ can measure the degree of inconsistency of A by the magnitude of its numerical values. The consistency indicator is defined as:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{2}$$

CI=0, with complete consistency; CI is close to 0, with satisfactory consistency; The larger the CI, the more severe the inconsistency.

To measure the size of CI, the random consistency index RI is introduced:

$$RI = \frac{CI_1 + CI_2 + \Lambda + CI_n}{n} \tag{3}$$

Among them, the random consistency index RI is related to the order of the judgment matrix. Generally, the higher the order of the matrix, the greater the possibility of random consistency deviation. The corresponding relationship is shown in Table 2 (standard values of the average random consistency index RI):

Tab 3 RI Standard Values for Matrix Order

matrix	1	2	3	4	5	6	7	8	9	10

coefficient										
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Considering that the deviation of consistency may be caused by random reasons, when testing whether the judgment matrix has satisfactory consistency, it is also necessary to compare CI with the random consistency indicator RI to obtain the test coefficient CR, which is expressed as follows:

$$CR = \frac{CI}{RI} \tag{4}$$

Generally, if $CR < 0.1$, it is considered that the judgment matrix has passed the consistency test, otherwise it does not have satisfactory consistency.

- Hierarchical overall ranking and consistency testing

Calculating the relative importance of all factors at a certain level to the highest level (overall goal) is called hierarchical total ranking. This process is carried out sequentially from the highest level to the lowest level.

The input judgment matrix is shown in Tab 4 :

Tab 4 Analytic Hierarchy Process (AHP) Judgment Matrix for Key Factors in China's New Energy Electric Vehicle Development

	Annual sales volume	Related patent applications quantity	Charging station quantity	Annual per capita GDP
Annual sales volume	1.000	6.000	3.000	2.000
Related patent applications quantity	0.200	1.000	0.500	0.330
Charging station quantity	0.333	2.000	1.000	0.670
Annual per capita GDP	0.500	3.000	1.500	1.000

Calculation result:

The calculated consistency ratio CR=0.0180<0.1 indicates that the consistency of the judgment matrix is acceptable.

The degree of influence of each factor calculated by AHP is shown in Table 5:

Tab 5 Weighted Impact Factors for Key Factors in China's New Energy Electric Vehicle Development

Factors	Factor Influence Weights
Annual sales volume	0.4980

Related patent applications quantity	0.0869
Charging station quantity	0.1662
Annual per capita GDP	0.2490

From the above table, it can be seen that sales volume is the main factor affecting the development of new energy electric vehicles in China, with per capita GDP being the secondary influencing factor. The number of related patents and charging stations also have a certain impact on the development of new energy electric vehicles in China.

4.2 Problem Two

4.2.1 Data collection and preprocessing

For this issue, in order to describe and predict the development of new energy electric vehicles in China in the next 10 years, we collected industry development data on new energy electric vehicles in China and preprocessed the data. Please refer to the supporting materials for specific data.

4.2.2 Establishment of ARIMA model

The ARIMA model mainly consists of three parts, namely autoregressive model (AR), differential process (I), and moving average model (MA), denoted as ARIMA (p, q, d). The ARIMA model attempts to extract time series patterns hidden behind data through autocorrelation and differentiation, and then use these patterns to predict future data. Among them:

(1) The AR part is used to process the autoregressive part of the time series, which considers the impact of observations from past periods on the current value.

(2) Part I is used to make non-stationary time series stationary, eliminating trend and seasonal factors in the time series through first-order or second-order differential processing.

(3) The MA part is used to process the moving average part of the time series, taking into account the impact of past prediction errors on the current value.

Combining these three parts, the ARIMA model can not only capture the trend changes of data, but also handle data with temporary, sudden changes, or high noise. So, ARIMA models have good applications in many time series prediction problems.

The ARIMA model can be expressed as:

$$\varphi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \tag{5}$$

B is a lagging operator, and B satisfies the following expression:

$$B^n y_t = y_{t-n} \tag{6}$$

$$\varphi(B) = 1 - \varphi_1 B - \Lambda - \varphi_p B^p \tag{7}$$

$$\theta(B) = 1 - \theta_1 B - \Lambda - \theta_q B^q \tag{8}$$

The key to establishing an ARIMA (p, q, d) model is to select the three parameters (p, q, d). D is the order of the difference. For the selection of p and q, this paper adopts the Bayesian Information Criterion. The Bayesian Information Criterion can provide a simple approximate logarithmic model evidence, as shown below:

$$BIC = Accuracy(m) - \frac{p}{2} loN \tag{9}$$

Among them, p is the number of parameters and N is the number of data points.

4.2.3 Establishment of Nonlinear Regression Models

In order to reflect the trend of recent data changes to the greatest extent possible, a linear regression model is selected. This model can be expressed as:

$$y = f\left(x, \theta_0, \theta_1, \Lambda, \theta_p\right) + \varepsilon_i \tag{10}$$

Among them, x is the independent variable and y is the true value.

4.2.4 Establishment of Adaptive Hybrid ARIMA Nonlinear Regression Model

For each prediction algorithm, a partial sequence is used as the test set. The mixing idea of the hybrid algorithm designed in this article is mainly that the better the performance in previous predictions, the higher the weight in future predictions, and the higher the contribution to the predicted values.

For actual observations, they are recorded as: $y_i = \{y_1, y_2, y_3, \Lambda, y_t\}$.The predicted value of algorithm k is denoted as: $\hat{y}_{k,i} = \left\{\hat{y}_{k,1}, \hat{y}_{k,2}, \hat{y}_{k,3}, \Lambda, \hat{y}_{k,i}\right\}$ .Due to the limited amount of data studied in the article, only the last observation value was used to adjust its weight. As a result, the prediction error can be extremely small:

$$wmape_{k,id} = \frac{\sum \left| y_i - \hat{y}_{k,i} \right|}{\sum y_i} \tag{11}$$

The total error of algorithm k can be expressed as:

$$wmape_k = \sum_{id} wmape_{k,id} \tag{12}$$

In hybrid algorithms, if an algorithm performs better in the test set, the weight will be higher. The weight of algorithm k in hybrid algorithms can be denoted as:

$$\omega_k = \frac{1/mape_k}{\sum_k (1/wmape_k)} \tag{13}$$

When calculating mixed prediction values, it is necessary to combine the predicted values of ARIMA algorithm and nonlinear regression algorithm. The calculation formula can be expressed as:

$$\hat{y}_i = \sum_k \omega_k \hat{y}_{k,i} \tag{14}$$

4.2.5 Solution of ARIMA model

The accuracy of ARIMA model prediction lies in the selection of the three parameters (p, d, q). D is the order of difference that makes the time series stationary. In this model, the stability of the model is tested using the Augmented Dickey Fuller test (ADF). At a significance level of 5%, if the p-value is less than 5%, the sequence is considered stationary. The ARMA model can be used for fitting differential stationary sequences. Defined as a stationary sequence after differencing, the ARMA (p, q) model can be denoted as:

$$y_{dt} = \phi_0 + \phi_1 y_{d(t-1)} + \phi_2 y_{d(t-2)} + \Lambda + \phi_p y_{d(t-p)} - \theta_1 \varepsilon_{d(t-1)} - \theta_2 \varepsilon_{d(t-2)} - \Lambda - \theta_q \varepsilon_{d(t-p)} \tag{15}$$

This article uses BIC to select the parameter values of (p, q). At this point, after obtaining the three parameters (p, d, q), the ARIMA (p, d, q) model can be applied to predict the data. Using the established model, predict the sales volume, number of patent applications, number of charging stations, and annual per capita GDP for the next ten years. The results are as follows:

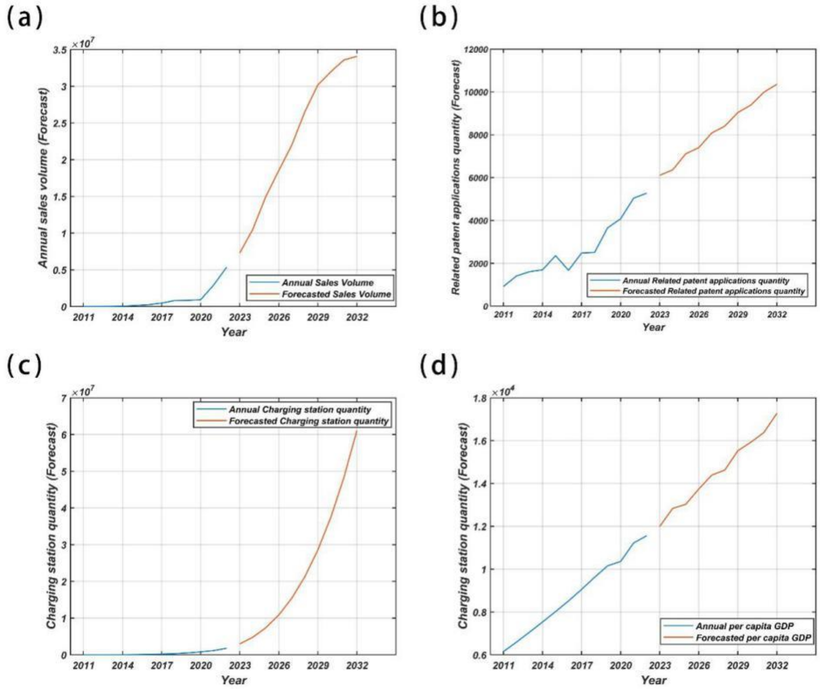

Figure 2 Forecast Projections Over a Decade Using ARIMA

a). Annual Sales Volume Forecast. b). Related Patent Applications Quantity Forecast.

c). Charging Station Quantity Forecast. d). Per Capita GDP Forecast.

From the above figure, it can be seen that the ARIMA model roughly describes the trend and some fluctuations of the shipment volume of new energy electric vehicles. However, there are still many fluctuations that have not been reflected in the predictions, and further improvement is needed.

4.2.6 Solving Nonlinear Regression Models

The established nonlinear regression model was used to predict the sales volume, number of patent applications, number of charging stations, and annual per capita GDP for the next ten years. The results are shown in the following figure:

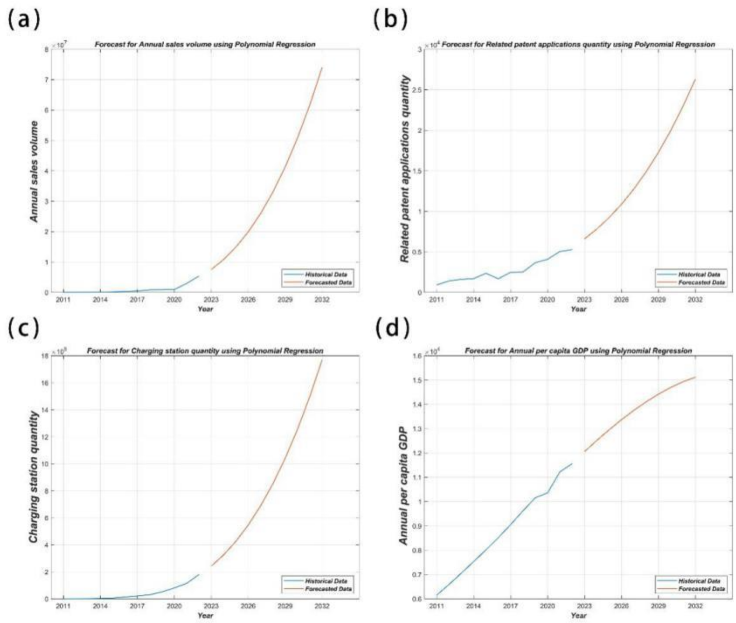

Figure 3 Forecast Projections Over a Decade Using Nonlinear Regression Models

a). Annual Sales Volume Forecast. b). Related Patent Applications Quantity Forecast. c). Charging Station Quantity Forecast. d). Per Capita GDP Forecast.

4.2.7 Solution of Adaptive Hybrid ARIMA Nonlinear Regression Model

Based on the calculation results in sections 2.5 and 2.6, the prediction errors of ARIMA model and nonlinear regression model can be obtained as shown in the table below.

Tab 6 The Prediction Error Table for ARIMA and Nonlinear Regression Models

Wmape	ARIMA	Nonlinear Regression
Annual sales volume	0.582810	0.121392
Related patent applications quantity	0.000146	0.355171
Charging station quantity	0.094675	0.233776
Annual per capita GDP	0.000341	0.042856

Tab 7 Tablex Normalized Results

	ω_1	ω_2
Annual sales volume	0.172382334000000	0.827617666000000
Related patent applications quantity	0.999588416000000	0.000411584000000000
Charging station quantity	0.711751571000000	0.288248429000000
Annual per capita GDP	0.992114484000000	0.00788551600000000

The predicted value is:

$$\hat{y}_i = \hat{y}_{ARIMA,i} \times \omega_1 + \hat{y}_{NR,i} \times \omega_2 \tag{16}$$

The predicted values are shown in the table below:

Tab 8 Adaptive Hybrid ARIMA-Nonlinear Regression Model Forecast Results

Year	Annual sales volume	Related patent applications quantity	Charging station quantity	Annual per capita GDP
2023	7488464	6107	2815788	11994
2024	10752025	6367	4338695	12835
2025	14910910	7112	6462587	13027

2026	19666293	7403	9297891	13739
2027	25235549	8085	12965892	14387
2028	31887525	8408	17595468	14622
2029	39366641	9043	23324071	15518
2030	47571410	9393	30297434	15916
2031	56865239	9994	38669658	16368
2032	67168342	10368	48603185	17259

From the data in the table, it can be seen that the sales volume, number of charging stations, and number of patent applications have steadily increased from 2023 to 2027, and there has been a significant increase every year from 2027 to 2032. It can be inferred that the development of new energy electric vehicles in China will steadily increase in the next five years, and there will be a leap in development from the fifth to the tenth year.

4.3 Problem Three

4.3.1 Data collection and processing

For this issue, this article collected sales and production data of new energy vehicles, as well as relevant data on traditional energy vehicles worldwide. Please refer to the supporting materials for specific data. Moreover, due to certain missing data, they are deleted in pairs during data preprocessing.

4.3.2 Distribution of test data

Before conducting correlation testing, it is necessary to first test the distribution of the data. Due to the small sample size, the Shapiro Wilk distribution was chosen to test small-scale samples.

Establishment of Shapiro Wilk distribution model：

Null Hypothesis H_0： Sample $X_1, X_2, \cdots, X_n$ comes from a normally distributed matrix；

Alternative Hypothesis H_1： The sample $X_1, X_2, \cdots, X_n$ does not come from a normally distributed matrix；

$\alpha = 0.05$.

The test statistic for this test is:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i-1}^{n}\left(x_i - \bar{x}\right)^2} \tag{17}$$

The constant a_i is calculated using the following formula:

$$\left(a_1, \cdots, a_n\right) = \frac{m^T V^{-1}}{\left(m^T V^{-1} V^{-1} m\right)^{1/2}} \tag{18}$$

If the P-value is less than 0.05, the null hypothesis can be rejected and the alternative hypothesis can be accepted, assuming that the data does not follow a normal distribution.

Distribution of test data:

The normality test results are shown in the table below:

Tab 9 examination table

	significance
New energy vehicles	0.000148
Automobile production	0.618789

If the P-value of the sales test for new energy vehicles is greater than 0.05, the null hypothesis can be accepted, and it is considered that the data follows a normal distribution. If the P-value of the automobile production test is less than 0.05, the null hypothesis can be rejected, and it can be assumed that the data does not follow a normal distribution.

4.3.3 Establishment of correlation model

Due to one factor not following a normal distribution and the other factor following a normal distribution, Spearman correlation analysis was chosen for analysis.

Establishment of Spearman correlation analysis model:

The Spearman correlation coefficient can be expressed as:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{19}$$

Among them, n represents the number of samples, and d represents the level difference between data x and y.

(1)For small sample situations (n ≤ 30), directly look up the critical value table

Null Hypothesis $H_0: r_s = 0$;

Alternative Hypothesis $H_1: r_s \neq 0$;

$\alpha = 0.05$.

Compare the obtained Spearman correlation coefficient r with the corresponding critical value.

(2) In the case of large samples, statistics

$$r_s\sqrt{n-1} \sim N(0,1) \tag{20}$$

Null Hypothesis $H_0: r_s = 0$;

Alternative Hypothesis $H_1: r_s \neq 0$;

$\alpha = 0.05$.

Calculate the test value and compare the corresponding p-value with 0.05.

4.3.4 Solving the correlation model

The results of the Spearman correlation coefficient test are shown in the table below:

Tab 10 Spearman test results

		Sales volume of new energy vehicles	Automobile production
New energy vehicles	correlation coefficient	1.000	0.360*
	Sig		0.024
	N	47	39
New energy vehicles	correlation coefficient	0.360*	1.000
	Sig	0.024	
	N	39	39

This indicates that there is a certain correlation between new energy vehicles and the automotive industry, but the correlation is not significant. The increase in automobile production includes the increase in the production of new energy vehicles. Furthermore, it can be inferred from the weak correlation between the two that when the sales of new energy vehicles increase, it is likely to lead to an increase in their own production, further leading to an increase in the total production of automobiles. However, the correlation between automobile production and new energy vehicle production is not strong, indicating that the growth of vehicle types is not significant except for new energy vehicles, and new markets are generally occupied by new energy vehicles.

4.4 Problem Four

4.4.1 Data collection and processing

In this section, an analysis of the impact of resistance policies on the development of new energy electric vehicles in China is conducted. This paper collects data related to resistance policies in various countries and China's new energy exports. Specific data can be found in the supporting materials.

Since policy data is often challenging to quantify, this paper transforms it into categorical data, specifically "whether resistance policies have been implemented." For the characteristics of new energy vehicle development, it is relatively easy to represent them using quantifiable data, which will be presented later in this paper.

4.4.2 Testing Data Distribution

Before conducting a correlation test, it is necessary to examine the distribution to which the collected data adheres. Due to the relatively small sample size, the Shapiro-Wilk distribution test is chosen for small-scale samples.

- Establishment of the Shapiro-Wilk Distribution Model

Null Hypothesis H_0: The sample X_1, X_2, Λ, X_n are drawn from a population that follows a normal distribution;

Alternative Hypothesis H_1: The sample X_1, X_2, Λ, X_n are not drawn from a population that follows a normal distribution; $\alpha = 0.05$

The test statistic for this test is:

$$W=\frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i-1}^{n}\left(x_i-\bar{x}\right)^2} \tag{21}$$

Where $x_{(i)}$ is the i_{th} order statistic,, representing the i_{th} smallest value in the sample, and is the mean of the sample, $\bar{x}$ and is the mean of the sample. The constant a_i is calculated using the following formula:

$$\left(a_1,\Lambda\ ,a_n\right)=\frac{m^T V^{-1}}{\left(m^T V^{-1} V^{-1} m\right)^{1/2}} \tag{22}$$

If the test p-value is less than 0.05, then we can reject the null hypothesis H_0, accept the alternative hypothesis H_1, and conclude that the data does not follow a normal distribution.

The distribution of the tested data

The results of the normality test are as follows:

Tab 11 Normality Test Results for Examined Data

Variable	Significance
Whether resistance policies exist	1.074E-13
Trend decrease in exports before and after policy implementation	1.1318E-10

Based on the results in the table, the p-values for both variables are less than 0.05. Therefore, we reject the null hypothesis and accept the alternative hypothesis, indicating that both sets of data do not follow a normal distribution.

4.4.3 Variance Homogeneity Test: Model Establishment and Solution

- Establishment of the Homogeneity of Variance Test Model

Null Hypothesis H_0 :The sample data follows homogeneity of variance, $\sigma_1^2=\sigma_2^2=\Lambda\ =\sigma_{\mathrm{r}}^2$

Alternative Hypothesis H_1: The sample data does not follow homogeneity of variance; $\alpha=0.10$.

When the number of repetitions of the experiment is equal for each level, i.e.,

$$m_1=m_2=\Lambda\ =m_r=m \tag{23}$$

m is the number of repetitions for each level, and it represents the ratio of the largest sample variance among r groups to the smallest sample variance when the variances are equal for each group. Under the condition of equal variances among groups, you can obtain critical values of the H distribution through random simulation experiments.

This distribution depends on the number of levels r and the degrees of freedom for the sample variances $f=\mathrm{m}\text{-}1$, and it is denoted as $H(r,\ f)$.

When H_0 is true, the value of H should be close to 1, When the value of H is relatively large, it indicates greater differences between the variances of the groups being compared. In this case, we reject the null hypothesis.

The rejection region for the given significance level is determined as follows:

$$W_1=\left\{H>H_{1-\alpha}\left(r,f\right)\right\} \tag{24}$$

Where $H_{1-\alpha}(r,f)$ is the $1-\alpha$ the quantile of the H distribution.

- Solution to the Homogeneity of Variance Test Model

Results of the Homogeneity of Variance Test:

Tab 12 Homogeneity of Variance Test Results for Examined Data

Variable	Significance
Whether resistance policies exist	5.2904E-280

According to the table, the significance level is greater than 0.1, indicating that we accept the null hypothesis, suggesting that the data satisfies homogeneity of variance.

4.4.4Kruskal-Wallis H Test: Model Establishment

In this study, we need to explore the relationship between categorical variables and continuous variables. Before conducting the Kruskal-Wallis H test, we have performed distribution and variance homogeneity tests, which show that the unordered variable data do not follow a normal distribution but satisfy homogeneity of variance. Therefore, we can use the Kruskal-Wallis H test to assess their correlation. The Kruskal-Wallis H test is a non-parametric method used to determine if there are significant differences in the distributions of two or more independent samples.

Assumptions:

Null Hypothesis H_0： $\mu_1=\mu_2=\Lambda=\mu_k$；

Alternative Hypothesis H_1： At least one pair $\mu_i\neq\mu_j$ exists;

$\alpha=0.001$

Steps for the test：

(1)Arrange all observations $N=\sum_{i=1}^{m}n_i$ from each sample in ascending order in a single column.

(2)For $R_i\left(i=1,\cdots,m\right)$ ranking of each group of data, calculate the sum of ranks n_i for each group of data, and the corresponding $X_{1,}\cdots,X_{ni}$ total rank sum;

(3)Test statistic

$$H=\frac{12}{N(N+1)}\sum_{i=1}^{m}\frac{R_i^2}{n_i}-3(N+1) \tag{25}$$

If there are r identical data points in each sample,， Assuming that $t_1\left(i=1,\cdots,r\right)$ represents the number of times the i_{th} common observation value appears in all N observed values for each sample,the corrected test statistic is calculated as follows:

$$H^{'}=\frac{N\left(N^2-1\right)}{\sum_{i=1}^{r}\left(t_i^3-t_i\right)}H \tag{26}$$

(4)For a given significance level α and degrees of freedom $v=m-1$,consult table or use software to obtain the results.

To characterize the impact of resistance policies on the development of China's new energy electric vehicles, this study employs the trend reduction quantity before and after the implementation of resistance policies.

$$de_{m,i} = (y_i - y_{i-1}) - (y_{i+1} - y_i) = 2y_i - y_{i-1} - y_{i+1} \tag{27}$$

4.4.5 The solution of the Kruskal-Wallis H test model

The Kruskal-Wallis H test yielded a p-value of 0.001577. The results indicate that the significance of the non-ordinal variable is less than 0.001, suggesting that the presence of resistance policies and the trend reduction quantity before and after their implementation are not entirely the same. In other words, there is an association between the presence of resistance policies and the change in export trends. Therefore, it can be observed that when certain countries implement resistance policies against China's new energy vehicles, it can have a certain impact on China's export trends to those countries.

Here, we present a selection of representative data for countries where resistance policies were either implemented or not implemented. The complete dataset can be found in the supporting materials (0 denotes no resistance policy implemented, 1 denotes the implementation of resistance policy):

Tab 13 Impact of Resistance Policies on China's New Energy Vehicle Exports

Whether Resistance Policies Exist	Trend Decrease in Exports Before and After Policy Implementation	Whether Resistance Policies Exist	Trend Decrease in Exports Before and After Policy Implementation
0	-1336	0	4219
1	999	0	-193
0	-14811	0	-252
0	5009	1	664
1	8	0	-2330
0	611	0	-3563
0	-6152	1	-6248

In the above table, higher numerical values indicate a greater degree of export trend decline, which corresponds to a more negative impact on the development of China's new energy vehicles. Observing the data in the table, it can be seen that the implementation of resistance policies can indeed lead to a deterioration in the development of China's new energy vehicles. However, there are still isolated cases where "resistance policies cannot stop the continuous expansion of new energy vehicles from China." This suggests that even when some countries implement resistance policies against China's new energy vehicles, they are unable to curb the growth of China's new energy vehicles.

4.5 Problem Five

4.5.1 Data collection and processing

In this question, the task is to analyze the impact of urban electrification of new energy electric vehicles (including electric buses) on the ecological environment. We have collected data on provincial-level green GDP estimation, comprehensive pollution index, annual population data for major cities, and the number of charging stations in various provinces nationwide. Furthermore, we have conducted data preprocessing. Please refer to the supporting materials for the relevant data.

4.5.2 Establish the model

To analyze the impact of urban electrification of new energy electric vehicles (including electric buses) on the ecological environment in conjunction with the urban population, we can establish a regression analysis model that relates new energy electric vehicles, the ecological environment, and population. Through this model, we can quantify the relationships between these three factors. When the population is given, we can determine the relationship between new energy electric vehicles and the ecological environment.

For easier data visualization based on the collected data, we can create three-dimensional scatter plot representing the number of charging stations, population, and comprehensive environmental pollution index. This plot would allow us to observe the relationships among these variables more conveniently.See fig.

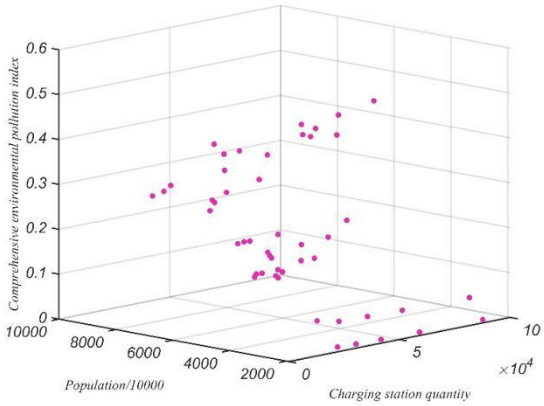

Figure 4 Table 1Three-Dimensional Scatter Plot: Relationship Analysis of New Energy Electric Vehicles, Population, and Ecological Impact

In this paper, we use the number of charging stations to characterize the development status of new energy electric vehicles. The relationship between them can be represented as:

$$y_i = f\left(x_1^i, x_2^i, \Lambda, x_j^i, \theta_1, \theta_2, \Lambda, \theta_p\right) + \sigma_i \varepsilon \left(i = 1,2,\Lambda, n\right) \tag{28}$$

In the equation provided，yrepresents the actual values，i indicates the i_{th} set of data，$f\left(x_1, x_2, \Lambda, x_j, \theta_1, \theta_2, \Lambda, \theta_p\right)$ is a multivariate nonlinear function， x_1, x_2, Λ, x_j are the independent variables， $\theta_1, \theta_2, \Lambda, \theta_p$ are the unknown parameters model for the multivariate nonlinear function， $\sigma_i \varepsilon$ are the random components， ε is a random variable following a N(0,1) distribution, σ_i is the random distribution standard deviation for the i_{th} set of data。

This paper has determined, through experimental methods, that the regression model suitable for this relationship is:

$$y = p_1 + p_2 x + p_3 x^2 + p_4 In(y) + p_5 In(y)^2 + p_6 In(y)^3 + p_7 4 \tag{29}$$

4.5.3 Model Solving Using Differential Evolution Algorithm

Based on the regression model determined in the previous sections, a total of 7 parameters need to be determined. To address the optimization and fitting problem with multiple parameters, commonly used methods include gradient descent and genetic algorithms. However, both of these methods have certain limitations as they tend to find local optima, leading to significant biases in the results.In comparison to genetic algorithms, the differential evolution algorithm retains a population-based global search strategy. It employs real-valued encoding, differential-based simple mutation operations, and "one-to-one" competitive survival strategies, reducing the complexity of evolutionary computations. Moreover, the unique memory capability of the differential evolution algorithm allows it to dynamically track the current search situation and adjust its search strategy. It exhibits strong global convergence and robustness without requiring specific feature information, making it suitable for solving complex optimization problems that are challenging or even impossible to solve using conventional mathematical programming methods. Therefore, in this paper, we have chosen to use the differential evolution algorithm for solving this problem. Below are the steps involved in solving with the differential evolution algorithm:

(1).Determine the control parameters and specific strategies to be used in the differential evolution algorithm. Control parameters include population size, mutation operator, crossover operator, maximum evolution generations, termination criteria, and so on.

(2).Randomly generate an initial population, set the evolution generation to $k=1$;

(3).Evaluate the initial population by calculating the objective function values for each individual in the population.

(4).Check if the termination criteria have been met or if the maximum evolution generation has been reached. If either condition is true, terminate the evolution process and output the best individual at that point as the solution. Otherwise, proceed to the next step.

(5).Perform mutation and crossover operations while handling boundary conditions, resulting in a temporary population.

(6).Evaluate the temporary population by calculating the objective function values for each individual.

(7).Perform "one-to-one" selection between individuals in the temporary population and their corresponding counterparts in the original population to obtain the new population.

(8).Increment the evolution generation $k=k+1$ and return to step (4).

Tab 14 Differential Evolution Algorithm Model Parameter Settings

Parameters	Values
Population Size	100
Crossover Rate	0.7
Mutation Rate	0.85
Convergence Tolerance	10^{-10}
Convergence Check Frequency	1000
Maximum Allowable Iterations	30000

Using the aforementioned differential evolution algorithm, we have obtained the optimized results for the seven parameters. Here are the optimized results for the seven parameters:

Tab 15 Coefficients of the Determined Regression Model

Parameters	Optimization Results
p_1	-8886.0736394066
p_2	1.97957878403193E-6
p_3	-2.12440068861636E-11
p_4	4272.58687746262

p_5	-769.476413873198
p_6	61.5188299361025
p_7	-1.84217149523183

Certainly, we will also calculate the error of the model.

(1) Root Mean Squared Error (RMSE): It is a statistical concept used in linear regression analysis. It measures the accuracy of estimates by calculating the square root of the average of squared differences between actual values and predicted values. The formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}\left(y_i - y_i^g\right)} \tag{30}$$

(2) SSE (Sum of Squared Errors): The estimated standard error is an indicator of the relative deviation between actual values and their estimated values. The formula for RMSE is as follows:

$$SSE = \sum_{i=1}^{n}\omega\left(y_y - \hat{y}_i\right)^2 \tag{31}$$

(3) Correlation Coefficient (R): Correlation is a measure of a non-deterministic relationship, and the correlation coefficient quantifies the degree of linear correlation between variables.

(4).R-Squared (R-squared): The closer R-squared is to 1, the better. A value greater than 0.8 is generally considered an acceptable result.

Before describing the determination coefficient, we need to introduce two other parameters, SSR and SST, as the determination coefficient depends on them:

1).SSR (Sum of Squares of the Regression): This is the sum of thesquared differences between the predicted data and the mean of the original data. The formula is as follows:

$$SSR = \sum_{i=1}^{n}\omega_i\left(\hat{y} - \overline{y_i}\right)^2 \tag{32}$$

2). SST(Total Sum of Squares):SST represents the sum of squares of the differences between the original data points and their mean value. The formula for calculating SST is as follows:

$$SST = \sum_{i=1}^{n}\omega_i\left(y_i - \overline{y_i}\right)^2 \tag{33}$$

Therefore, the formula for calculating R-Squared (R-squared) is as follows:

$$R^2 = \frac{SSR}{SST} \tag{34}$$

After applying the differential evolution algorithm, the errors for this quantitative model are as follows in the table:

Tab 16Model Evaluation Metrics and Errors After Differential Evolution Optimization

Indicator	Result
RMSE	0.110357865596187

SSE	0.596776066448361
R	0.727660625705266
R^2	0.529489986201779

4.5.4 The results of the solution

The fitting results obtained through the differential evolution algorithm are displayed in the following graph:

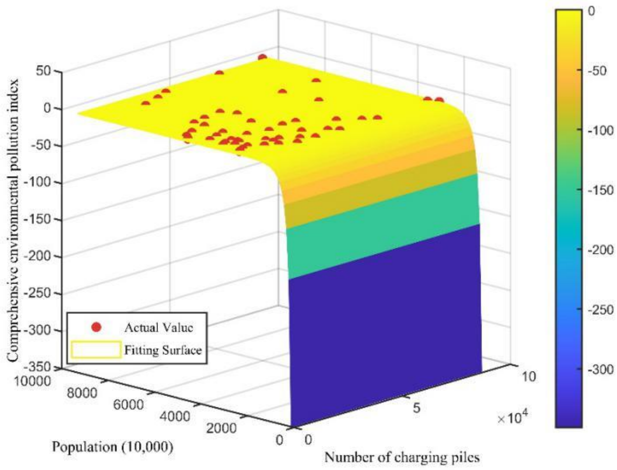

Figure 5 Differential Evolution Algorithm Fitting for Environmental Impact of EV Charging Infrastructure

When the urban population is held constant at 1 million people, we can observe the relationship between new energy electric vehicles and the ecological environment, as shown in the following graph:

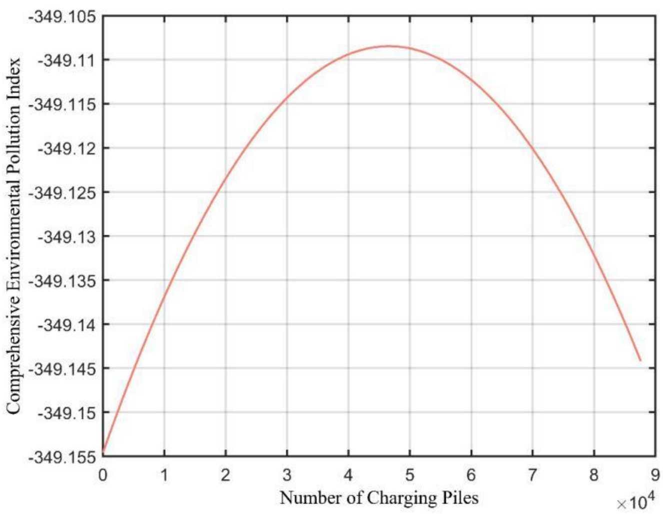

Figure 6 Environmental Impact of New Energy Electric Vehicles with Constant Urban Population

Based on the content in the figure, it can be observed that the impact of new energy electric vehicles on the environment can be broadly divided into two phases: Phase One: As the number of new energy vehicles gradually increases from zero, the environmental pollution index rises. During this phase, the introduction and initial adoption of new energy vehicles appear to lead to an increase in the environmental pollution index. This could be due to factors like increased production and charging infrastructure. Phase Two: As the market demand for automobiles approaches saturation, and new energy vehicles begin to gradually replace traditional vehicles, the environmental pollution index starts to decrease. The decline in the pollution index is likely a result of the environmentally friendly nature of new energy vehicles, which emit fewer pollutants. These two phases illustrate the overall trend of how the adoption of new energy electric vehicles can affect the environment, emphasizing the importance of considering the broader context and market dynamics when evaluating environmental impacts.

4.6 Problem Six

4.6.1 A letter to citizens

Dear fellow citizens,

We are writing this letter to raise awareness about the numerous benefits of new energy electric vehicles and to discuss the contributions of the electric vehicle industry to the environment and society worldwide.

As an environmentally friendly and sustainable mode of transportation, new energy electric vehicles offer many significant advantages. Firstly, they produce almost zero tailpipe emissions, making them highly effective in improving air quality and reducing air pollution. Compared to traditional gasoline-powered vehicles, they emit fewer harmful substances into the atmosphere, thus mitigating the adverse effects on global climate change.

Furthermore, new energy electric vehicles contribute to a reduction in noise pollution. Their electric engines run quieter than the older internal combustion engine vehicles, creating a more tranquil urban environment. This has a positive impact on residents' comfort, health, and overall quality of life.

In addition, the electric vehicle industry has made significant contributions to the economy and society. It has driven the development and innovation of green technologies, promoting economic growth and employment opportunities. The production and sale of electric vehicles require a substantial workforce, and advancements in battery technology have fueled growth in the battery manufacturing sector. This has injected new vitality into the economies of many countries and laid the foundation for sustainable development in the future.

Countries around the world have recognized the importance of electric vehicles and have taken action to promote their development. China, as one of the largest electric vehicle markets globally, has implemented a series of supportive policies, including vehicle purchase subsidies and the construction of charging infrastructure, to encourage more citizens to adopt electric vehicles. Similarly, European countries have established stringent emissions standards and provided funding and incentives for the growth of electric vehicles. Other nations are also actively pushing for the development of the electric vehicle industry and implementing regulations and policies to facilitate the transition.

Therefore, I sincerely invite everyone to join the ranks of electric vehicle users and support the development of new energy electric vehicles. Choosing to purchase and drive electric vehicles is not only for the environment but also for our future and the next generation. Each person's choice can have a positive impact on the environment and promote the development of sustainable transportation. Let us work together to create a cleaner and greener future!

5.Strengths and Weakness

5.1 Strengths

(1) Using the AHP model, a systematic analysis method was established, and a concise and practical decision-making method was obtained.

(2) The comprehensive model of ARIMA nonlinear regression is used to accurately describe and predict the development of new energy electric vehicles in China in the next 10 years.

(3) The Spearman correlation analysis model is used to analyze the impact of new energy electric vehicles on the global traditional energy vehicle industry, and its interpretability is strong.

(4) The Kruskal Wallis H test is used to test the significant differences between boycott policies and the development of new energy electric vehicles in China, and this test method can handle missing data and outliers because it is calculated based on rank rather than raw data. To illustrate the impact of these policies on the development of new energy electric vehicles in China.

(5) The differential evolution algorithm is used to solve the regression model of problem 5, which avoids the error of obtaining local optimal solutions and causing significant deviation in the results. It has strong global convergence ability and robustness, and does not require the use of problem feature information.

5.2 Weakness

(1) Using the AHP model cannot provide new solutions for decision-making, and there is limited quantitative data and more qualitative components. Not comprehensively analyzing the more factors that affect the development of new energy electric vehicles in China.

(2) The precise calculation of eigenvalues and eigenvectors using the AHP model is complex and often involves certain errors.

(3) Due to the lack of statistics on the global market, this article only takes China as an example to analyze the impact of new energy vehicles on the traditional energy vehicle industry.

(4) Due to the difficulty in quantifying policy data, converting it into categorical data for processing, i.e. whether a boycott policy has been implemented, is more absolute.

(5) Data collection inevitably has incompleteness, which can also lead to inaccuracies in model establishment and solution.

6.References

[1] Saaty, T. L. (1980). The analytic hierarchy process. McGraw-Hill.

[2] Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). Time series analysis: forecasting and control. John Wiley & Sons.

[3] Bates, D., & Watts, D. (2007). Nonlinear regression analysis and its applications. John Wiley & Sons.

[4] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591-611.

[5] Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15(1), 72-101.

[6] Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260), 583-621.

[7] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis. John Wiley & Sons.

[8] Storn, R., & Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization, 11(4), 341-359.

[9] Vaidya, O. S., & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. European Journal of Operational Research, 169(1), 1-29.

[10] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

[11] Faraway, J. J. (2006). Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. CRC Press.

7.Appendix

Code Appendices

Table of Contents

3. Question 5

3.1 Fig1

3.2 Fig2

3.3 Fig3

Codes:

1. Question 1

1.1 Analytic Hierarchy Process (AHP) for Weight Calculation and Consistency Check

```
clc
clear
A = xlsread('AHP Judgment Matrix.xlsx');

[n, n] = size(A);

[V, D] = eig(A); % Calculate all eigenvalues of matrix A, form diagonal matrix D, and calculate eigenvectors of A forming column vectors in V.
Max_eig = max(max(D));
[r, c] = find(D == Max_eig, 1);
disp('Results of Eigenvalue Method for Weight Calculation:');
w = V(:,c) ./ sum(V(:,c));
disp(w)
```

```
    % % % % % % % % % % % % % The following calculates the Consistency Ratio
CR % % % % % % % % % % % %
    CI = (Max_eig - n) / (n - 1);
    RI = [0 0.0001 0.52 0.89 1.12 1.26 1.36 1.41 1.46 1.49 1.52 1.54 1.56 1.58
1.59];  % Note that RI supports up to n = 15
    % When n = 2, the matrix is always consistent, so CI = 0. To avoid a
denominator of 0, we set the second element here to a very small positive
number.
    CR = CI / RI(n);
    disp('Consistency Index CI ='); disp(CI);
    disp('Consistency Ratio CR ='); disp(CR);
    if CR < 0.10
        disp('Since CR < 0.10, the consistency of judgment matrix A is
acceptable!');
    else
        disp('Attention: CR >= 0.10, so judgment matrix A needs to be
modified!');
    end
```

2. Question 2

2.1 ARIMA

```
    function [forData] = ARIMA(data, step, count)
    ytick_labels1 = {'Annual sales volume (Forecast)', 'Related patent
applications' ...
        ' quantity (Forecast)', 'Charging station quantity (Forecast)',
'Annual per capita GDP (Forecast)'};
    ddata = data;
    d = 2;
    while kpsstest(ddata) == 1
        ddata = diff(data);
        d = d + 1;
        if d > 3
            break
        end
    end
```

```
% Test data stationarity
adftest(data);
kpsstest(data);

% First-order differencing and stationarity test
ddata = diff(data);
d1_adf = adftest(ddata);
d1_kpss = kpsstest(ddata);

%% Calculate p and q values
pmax = 3;
qmax = 3;
[p, q] = findPQ(data, pmax, qmax, d);

%% Build the model
Mdl = arima(p, d, q);  % The second variable is set to 1 for first-order
differencing
EstMdl = estimate(Mdl, data);

%% Model prediction
[forData, YMSE] = forecast(EstMdl, step, 'Y0', data);

% Plot the figure
figure
grid on;
plot(1:length(data), data)
hold on
plot((length(data) + 1):(length(data) + step), forData)
hold on

% Modify y-axis label
hXLabel = xlabel('Year');
hYLabel = ylabel(ytick_labels1(count));

% Modify font size
set(hXLabel, 'FontSize', 12);
set(hYLabel, 'FontSize', 12);

% Modify x-axis range
xticks([1:length(data), (length(data) + 1):(length(data) + step)]);
xticklabels({'2011', '2012', '2013', '2014', '2015', '2016', '2017',
'2018', '2019', '2020', '2021', '2022', '2023', '2024', '2025', '2026',
'2027', '2028', '2029', '2030', '2031', '2032'});
```

```
% Set x-axis label positions
xtick_positions = [1, 4, 7, 10, 13, 16, 19, 22, 25];

% Set custom x-axis labels
xtick_labels = {'2011', '2014', '2017', '2020', '2023', '2026', '2029',
'2032'};

% Modify x-axis range
xticks(xtick_positions);
xticklabels(xtick_labels);
set(gcf, 'Color', [1 1 1]);

% Add legend
legend('Annual per capita GDP', 'Forecasted per capita GDP', 'Location',
'SouthEast');
grid on;
end
```

2.2 NR

```
clc
clear

% Load the data from the '4Tables_total.xlsx' Excel file
data = xlsread('4Tables_total.xlsx');

% Extract the relevant data columns (2nd, 3rd, 4th, and 5th)
columns_to_forecast = data(:, 2:5);

refer_date = size(data, 1); % Number of rows, assuming each row
corresponds to a year
step = 10; % Number of future periods to forecast

% Define different y-axis labels for each column
ytick_labels1 = {
    'Annual sales volume',
    'Related patent applications quantity',
    'Charging station quantity',
    'Annual per capita GDP'
};

% Initialize a cell array to store forecasted data for each column
forecasted_data = cell(1, size(columns_to_forecast, 2));
```

```
    % Loop through each column and perform forecasting
    for col = 1:size(columns_to_forecast, 2)
        % Get the data for the current column
        column_data = columns_to_forecast(:, col)';

        % Polynomial regression for forecasting
        coefficients = polyfit(1:size(column_data, 2), column_data, 3); % You
can adjust the polynomial degree as needed
        x = (refer_date + 1):(refer_date + step);
        for kk = 1:length(x)
            forData(kk) = polyval(coefficients, x(kk));
        end

        % Store the forecasted data in the cell array
        forecasted_data{col} = forData;

        % Plot the historical data and forecasted values for the current
column
        figure('Renderer', 'painters', 'Position', [100, 100, 800, 600],
'Color', 'white', 'InvertHardcopy', 'off', 'PaperPositionMode', 'auto',
'PaperUnits', 'points', 'PaperSize', [800, 600]);
        plot(1:length(column_data), column_data, 'LineWidth', 1)
        hold on
        plot((length(column_data) + 1):(length(column_data) + step),
forecasted_data{col}, 'LineWidth', 1)
        xlabel('Year', 'FontWeight', 'bold', 'FontAngle', 'italic')
        ylabel(ytick_labels1{col}, 'FontWeight', 'bold', 'FontAngle',
'italic') % Set the y-axis label for the current column
        title(sprintf('Forecast for %s using Polynomial Regression',
ytick_labels1{col}), 'FontWeight', 'bold', 'FontAngle', 'italic', 'FontSize',
1.1*get(gca,'FontSize'))
        legend('Historical Data', 'Forecasted Data', 'Location', 'southeast',
'FontWeight', 'bold', 'FontAngle', 'italic')
        grid on

        % Set custom x-axis labels
        xtick_positions = 1:3:(refer_date + step);
        xtick_labels = arrayfun(@(year) sprintf('%d', year), 2011:3:2032,
'UniformOutput', false);
        xticks(xtick_positions);
        xticklabels(xtick_labels);

        % Save the plot as a JPG file with specified settings
        save_path = 'E:\APCP\Question2\Result';
        file_name = sprintf('Forecast_%s.jpg', strrep(ytick_labels1{col}, ' ',
'_'));
```

```
    full_path = fullfile(save_path, file_name);
    print(full_path, '-djpeg', '-r600');
end
```

2.3 Q2_1_main

```
clc
clear
data = xlsread('4Tables_total.xlsx');
% Read data from an Excel file


% Extract data from columns 2 to 5 and perform ARIMA forecasts for each
for column = 5:5
    % Extract data column
    column_data = data(:, column);

    % Set the forecasting period
    step = 10;

    % Call the ARIMA function for forecasting
    [forData] = ARIMA1(column_data, step, column-1);
end
```

2.4 ARIMA_for_error

```
function [forData] = ARIMA3(data, step)
    ddata = data;
    d = 0;
    while kpsstest(ddata) == 1
        ddata = diff(data);
        d = d + 1;
        if d > 3
            break
        end
    end

    % Calculate the values for p and q
    pmax = 2;
    qmax = 2;
    [p, q] = findPQ(data, pmax, qmax, d);
```

```
% Build the model
Mdl = arima(p, d, q);
EstMdl = estimate(Mdl, data);

% Model prediction
[forData, ~] = forecast(EstMdl, step, 'Y0', data);

% Function to calculate p and q
function [p, q] = findPQ(data, pmax, qmax, d)
    data = reshape(data, length(data), 1);
    LOGL = zeros(pmax + 1, qmax + 1);
    PQ = zeros(pmax + 1, qmax + 1);
    for p = 0:pmax
        for q = 0:qmax
            model = arima(p, d, q);
            [fit, ~, logL] = estimate(model, data);
            LOGL(p + 1, q + 1) = logL;
            PQ(p + 1, q + 1) = p + q;
        end
    end
    LOGL = reshape(LOGL, (pmax + 1) * (qmax + 1), 1);
    PQ = reshape(PQ, (pmax + 1) * (qmax + 1), 1);
    m2 = length(data);
    [aic, bic] = aicbic(LOGL, PQ + 1, m2);
    aic0 = reshape(aic, (pmax + 1), (qmax + 1));
    bic0 = reshape(bic, (pmax + 1), (qmax + 1));

    aic1 = min(aic0(:));
    index = aic1 == aic0;
    [pp, qq] = meshgrid(0:pmax, 0:qmax);
    p0 = pp(index);
    q0 = qq(index);

    aic2 = min(bic0(:));
    index = aic2 == bic0;
    [pp, qq] = meshgrid(0:pmax, 0:qmax);
    p1 = pp(index);
    q1 = qq(index);

    if p0^2 + q0^2 > p1^2 + q1^2
        p = p1;
```

```
            q = q1;
        else
            p = p0;
            q = q0;
        end
    end
end
```

2.5 Q2_2_error_ARIMA

```
clc
clear
data = xlsread('4Tables_total.xlsx');
% Read data from an Excel file

% Set the forecasting period
step = 1;
num_columns_to_forecast=4;
% Initialize an array to store forecast results for each column
forecast_results = zeros(size(data, 1), num_columns_to_forecast);
i=3;
% Loop through columns 2 to 5
for column = 5:5
    % Extract data column
    column_data = data(:, column);

    % Split the data into training and testing sets
    training_data = column_data(1:end-i);
    testing_data = column_data(end-i+1);

    % Call the ARIMA1 function for forecasting
    [forData] = ARIMA_for_error(training_data, step);


end

% Display the forecast results
result=[forData,testing_data]
```

2.6 NR_for_error

```
clc
clear
ytick_labels1 = {
    'Annual sales volume',
    'Related patent applications quantity',
    'Charging station quantity',
    'Annual per capita GDP'
};
% Load the data from the '4Tables_total.xlsx' Excel file
data = xlsread('4Tables_total.xlsx');

% Extract the relevant data columns (2nd, 3rd, 4th, and 5th)
columns_to_forecast = data(:, 2:5);

refer_date = size(data, 1); % Number of rows, assuming each row
corresponds to a year

% Initialize a cell array to store forecasted data for each column
forecasted_data = cell(1, size(columns_to_forecast, 2));

% Define the range of values for i (from 1 to 3)
for i = 1:3
    % Loop through each column and perform forecasting
    for col = 1:size(columns_to_forecast, 2)
        % Get the data for the current column
        column_data = columns_to_forecast(:, col)';

        % Polynomial regression for forecasting
        coefficients = polyfit(1:(refer_date-i), column_data(1:end-i),
3); % Use data up to the (length-i)-th point
        x = refer_date - i + 1;
        forData = polyval(coefficients, x);

        % Store the forecasted data in the cell array
        forecasted_data{col} = [forecasted_data{col}, forData];
    end
end

% Display the forecasted data
for col = 1:size(columns_to_forecast, 2)
```

```
    fprintf('Forecasted data for %s:\n', ytick_labels1{col});
    disp(forecasted_data{col});
end
```

3. Question 5

3.1 Fig1

```
clc
clear
data = xlsread('Consolidated table.xlsx');

D1 = data;

% Enlarge the size of points by 1.5 times
pointSize = 8 * 2;

% Set the color to be used
pointColor = [223/255, 48/255, 175/255]; % RGB value of hexadecimal color
"#a91a3a"

scatter3(D1(:, 1), D1(:, 2), D1(:, 3), pointSize, pointColor, 'filled')
hold on

hXLabel = xlabel('Charging Station Quantity');
hYLabel = ylabel('Population/10,000');
hZLabel = zlabel('Comprehensive Environmental Pollution Index');
set(gcf, 'Color', [1 1 1])
set([hXLabel, hYLabel, hZLabel], 'FontName', 'Times New Roman')
set(gca, 'FontSize', 10)
set([hXLabel, hYLabel, hZLabel], 'FontSize', 9)
```

3.2 Fig2

```
clc;
clear;
point = xlsread('Consolidated table.xlsx');
color = [223/255, 53/255, 48/255;
         130/255, 176/255, 210/255;
         190/255, 184/255, 220/255;
         231/255, 218/255, 210/255;
         153/255, 153/255, 153/255];
```

```
figure(1);
scatter3(point(:,1), point(:,2), point(:,3), [], color(1,:), 'filled')
hold on
xlabel('Number of Charging Piles', 'Color', 'k', 'FontSize', 14, 'FontName', 'Times New Roman')
ylabel('Population (10,000)', 'Color', 'k', 'FontSize', 14, 'FontName', 'Times New Roman')
zlabel('Comprehensive Environmental Pollution Index', 'Color', 'k', 'FontSize', 14, 'FontName', 'Times New Roman')

x = [0:100:max(point(:,1))];
y = [0:100:max(point(:,2))];
[X, Y] = meshgrid(x, y);
p1 = -8886.0736394066;
p2 = 1.97957878403193E-6;
p3 = -2.12440068861636E-11;
p4 = 4272.58687746262;
p5 = -769.476413873198;
p6 = 61.5188299361025;
p7 = -1.84217149523183;
Z = p1 + p2*X + p3*X.^2 + p4*log(Y) + p5*(log(Y)).^2 + p6*(log(Y)).^3 + p7*(log(Y)).^4;
Z(1,1) = Z(1,2);

% Plot the surface and use the "parula" colormap
Fig = mesh(X, Y, Z);
colormap(parula) % Use the "parula" colormap
colorbar

% Adjust the legend size to 0.5x
hLegend = legend('Actual Value', 'Fitting Surface', 'FontSize', 8, 'FontName', 'Times New Roman');
set(hLegend, 'Units', 'normalized', 'Position', [0.8, 0.8, 0.1, 0.1]) % Adjust legend position and size

set(gcf, 'Color', [1 1 1])
set(gca, 'FontSize', 8.5) % Set the font size for tick labels
```

3.3 Fig3

```
clc;
clear;
point = xlsread('Consolidated table.xlsx');
```

```
color = [223/255, 53/255, 48/255;
         130/255, 176/255, 210/255;
         190/255, 184/255, 220/255;
         231/255, 218/255, 210/255;
         153/255, 153/255, 153/255];

figure(1);
scatter3(point(:,1), point(:,2), point(:,3), [], color(1,:), 'filled')
hold on
xlabel('Number of Charging Piles', 'Color', 'k', 'FontSize', 14, 'FontName', 'Times New Roman')
ylabel('Population (10,000)', 'Color', 'k', 'FontSize', 14, 'FontName', 'Times New Roman')
zlabel('Comprehensive Environmental Pollution Index', 'Color', 'k', 'FontSize', 14, 'FontName', 'Times New Roman')

x = [0:100:max(point(:,1))];
y = [0:100:max(point(:,2))];
[X, Y] = meshgrid(x, y);
p1 = -8886.0736394066;
p2 = 1.97957878403193E-6;
p3 = -2.12440068861636E-11;
p4 = 4272.58687746262;
p5 = -769.476413873198;
p6 = 61.5188299361025;
p7 = -1.84217149523183;
Z = p1 + p2*X + p3*X.^2 + p4*log(Y) + p5*(log(Y)).^2 + p6*(log(Y)).^3 + p7*(log(Y)).^4;
Z(1,1) = Z(1,2);

% Plot the surface and use the "parula" colormap
Fig = mesh(X, Y, Z);
colormap(parula) % Use the "parula" colormap
colorbar

% Adjust the legend size to 0.5x
hLegend = legend('Actual Value', 'Fitting Surface', 'FontSize', 8, 'FontName', 'Times New Roman');
set(hLegend, 'Units', 'normalized', 'Position', [0.8, 0.8, 0.1, 0.1]) % Adjust legend position and size

set(gcf, 'Color', [1 1 1])
set(gca, 'FontSize', 8.5) % Set the font size for tick labels
```