Scientific and Social Research, 2025, Volume 7, Issue 2
https://ojs.bbwpublisher.com/index.php/SSR
Online ISSN: 2981-9946
Print ISSN: 2661-4332

Integrating Functional Principal Component Analysis with Generalized Regression Neural Networks for Enhanced Forecasting of New Energy Vehicle Market Trends

Long Chen*, Yong Wang, Qianhao Cheng, Shuxian Shen, Ying Wang

Anqing Normal University, Anqing 246133, Anhui Province, China

Corresponding author: Long Chen, 3464842084@qq.com

Abstract: With the rapid development of the new energy vehicle (NEV) industry, it has become a critical component of modern productivity, offering immense potential for economic growth and environmental benefits. However, accurately predicting NEV trends, such as sales volume and the annual installed capacity of power batteries, remains challenging due to the high dimensionality of data and the limitations of existing predictive methods. In view of these challenges, this paper proposes a novel prediction method that combines Functional Principal Component Analysis (FPCA) and Generalized Regression Neural Network (GRNN) to forecast NEV sales and power battery installation trends. By employing FPCA to reduce the dimensionality of nine key variables and using GRNN to incorporate seven influencing factors, such as ownership and total energy consumption, the study constructs a robust predictive model. The GRNN model is further optimized using cross-validation, achieving a high level of accuracy. It is hoped that the proposed FPCA-GRNN method, which has demonstrated superior performance compared to traditional approaches such as BP neural networks and multiple linear regression, will serve as a valuable tool for predicting NEV development trends and provide guidance for industry growth and policy-making in the NEV sector.

Keywords: Functional principal component analysis; Generalized regression neural network; Legendre polynomials; Multiple linear regression prediction

Online publication: March 10, 2025

1. Introduction

In the context of globalization, new energy vehicles (NEVs) have emerged as a promising solution to address energy and environmental challenges. By reducing reliance on fossil fuels and fostering environmental awareness, NEVs play a pivotal role in enhancing energy security, driving industrial upgrades, and promoting technological innovation. As a cornerstone of sustainable economic and social development, NEVs have garnered significant attention globally. To support their growth, governments worldwide have implemented a range of policies, such as setting ambitious sales targets, offering purchase subsidies, restricting traditional fuel vehicles, and expanding charging infrastructure. These

initiatives have effectively lowered costs, improved accessibility, and spurred innovation across the NEV industry chain.

In China, the world's largest NEV market, industrial progress is particularly vital to the reform and sustainable development of the global automotive industry. Recent data highlights the remarkable growth of China's NEV sector, driven by technological advancements, increasing market demand, and improvements in the industry chain. Research focused on forecasting the development trends of NEVs is essential for optimizing industry strategies and guiding policymaking. Accurate predictions of sales volume and battery installation capacity are key to promoting advanced productivity and achieving high-quality economic growth [1].

Despite the significant strides made in NEV development, challenges remain in forecasting future trends due to the high dimensionality of relevant data and the limitations of existing predictive models. This study aims to address these gaps by proposing a novel prediction method that combines Functional Principal Component Analysis (FPCA) and Generalized Regression Neural Network (GRNN). By leveraging these advanced methodologies, this research provides a comprehensive framework for accurately forecasting NEV sales and battery installations, offering valuable insights to support industrial and policy decision-making.

2. Functional principal component analysis

2.1. FPCA principal component analysis

Functional principal component analysis (FPCA) is a statistical method for analyzing functional data, where functions are treated as the basic units of data rather than traditional individual data points [2]. FPCA aims to identify major patterns of variability within the data in function space, maximizing the interpretation of data variability by linearly combining the original data. Similar to principal component analysis (PCA), FPCA identifies principal components that are orthogonal to each other, which allows for dimensionality reduction. FPCA works by calculating the covariance matrix of the data and its eigenvalues and eigenvectors. The principal components that explain the most variability are then selected for effective data fitting and dimensionality reduction. In addition, correlation analysis using SPSS was conducted to examine the relationships between key factors and market indicators, providing valuable insights into the factors influencing market trends. Use selected principal components to reconstruct raw functional type data. Select the most important principal components to explain most of the variance in the data. This is usually achieved by retaining principal components that explain more than a certain threshold of variance. Functional principal components normalized data are shown in **Figure 1** and principal components and target variables in **Figure 2**.

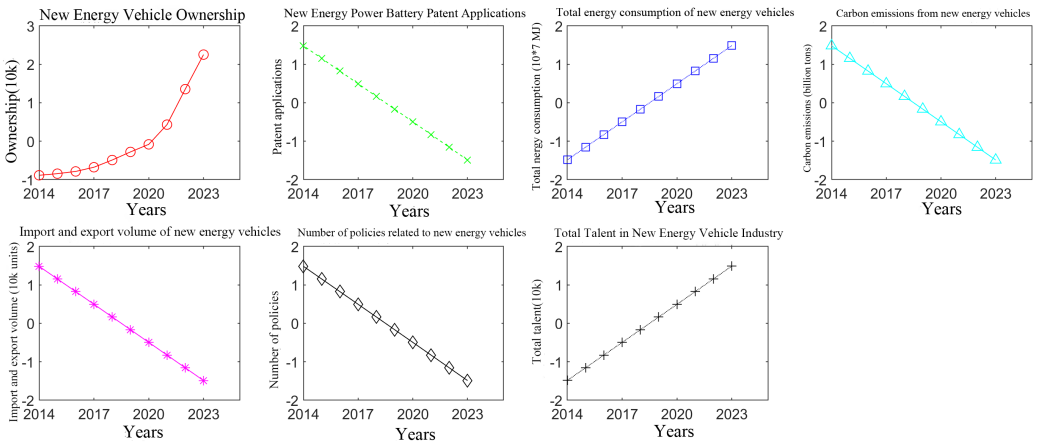

Figure 1. Function principal component normalized data image

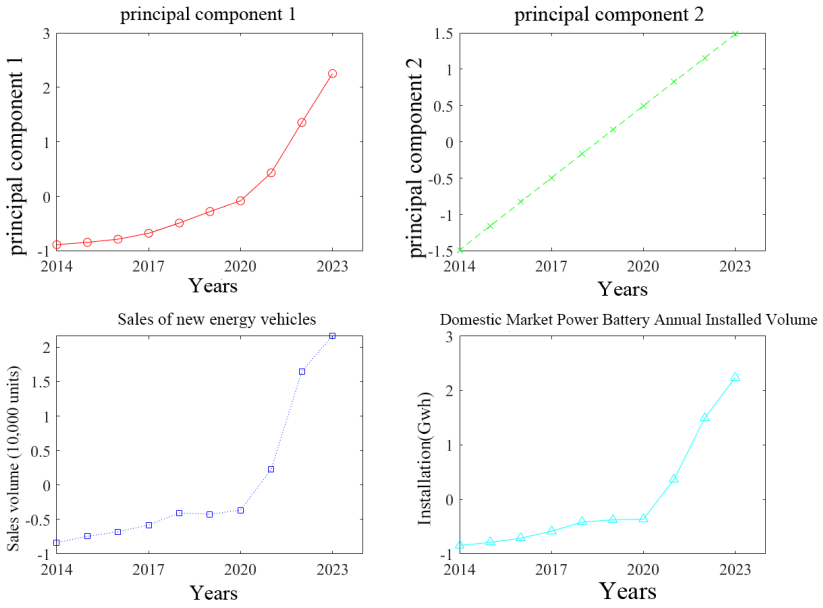

Figure 2. Trends in principal components and target variables

2.2. Calculating Legendre polynomials

This paper uses the Legendre function expansion method, where the Legendre functions are orthogonal polynomials, resulting in a symmetric matrix$W$. The resulting symmetric matrix$W$ is a unit matrix [3]. If other non-orthogonal basis functions are used, it is unavoidable to calculate each elementW_{ij} in the matrix$W$. If the order increases, the computation becomes more complicated. Each element W_{ij} of the matrix$W$ is inevitably computed, and if the computational effort is large as the order increases, it is fitted as shown in *Figure 3*.

$$f(x) = 0.0072771 - 44.0098x_1 + 88719.8687x_2 \\ -59617172.2584x_3 - 2.7022 \times 10^{-8}x_4 + 3.7497 \times 10^{-11}x_5$$

The Legendre multinomial equation is obtained as follows.

$$P_l(x) = \sum_{j=0}^{\left[\frac{l}{2}\right]} (-1)^j \frac{(2l-2j)}{2^l j!\,(l-j)!\,(j-2j)!} x^{l-2j}$$

$$\left[\frac{l}{2}\right] = \begin{cases} \dfrac{l}{2}, l = 2n \\ \dfrac{l-1}{2}, l = 2n+1 \end{cases} \quad (n = 0,1,2,\cdots)$$

$$P_0(x) = 1, P_1(x) = x, P_2(x) = \frac{1}{2}(3x^2-1), P_3(x) = \frac{1}{2}(5x^3-3x), P_4(x) \\ = \frac{1}{8}(35x^4 - 30x^2 + 3)$$

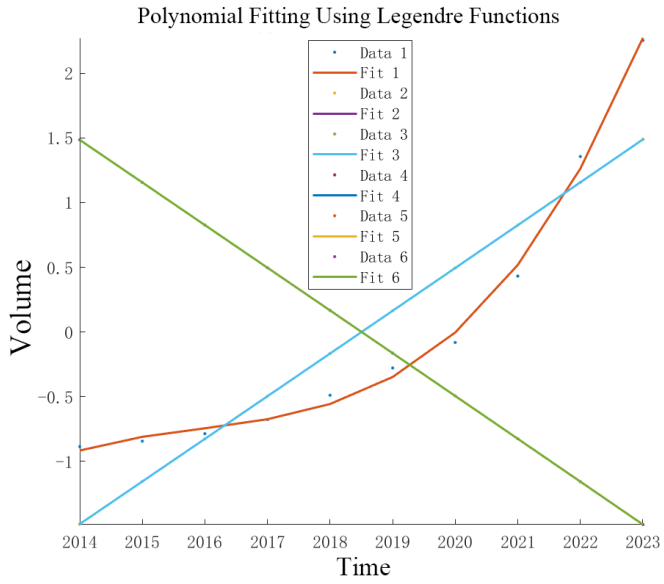

Figure 3. Legendre polynomial function fitted image

2.3. FPCA analysis results

FPCA retains as much information as possible from the original variables by extracting a small number of uncorrelated principal components. It achieves this by calculating the correlation or covariance matrix of the sample data and selecting the eigenvectors corresponding to the largest eigenvalues for dimensionality reduction. In this study, SPSS was used to conduct a correlation analysis of the seven factors influencing the annual installed capacity of power batteries and the sales volume of new energy vehicles in the domestic market. The results of the correlation analysis are summarized in the correlation coefficient matrix presented in **Table 1**.

Table 1. Matrix of correlation coefficients of observed variables

	Retention (10,000 units)	**Battery patent applications (pieces)**	**Total energy consumption (10*7 MJ)**	**Total carbon emissions (billion tonnes)**	**Import and export volume (10,000 units)**	**Number of relevant policies (number)**	**Total number of talents (10,000)**
Retention (10,000 units)	1.000						
Battery Patent applications (pieces)	0.916	1.000					
Total energy consumption (10*7 MJ)	0.930	0.941	1.000				
Total carbon emissions (billion tonnes)	0.729	0.854	0.874	1.000			
Import and export volume (10,000 units)	0.981	0.878	0.865	0.672	1.000		
Number of relevant policies (number)	0.461	0.526	0.504	0.737	0.526	1.000	
Total number of talents (10,000)	0.853	0.963	0.954	0.886	0.799	0.539	1.000

There is a strong correlation between several influencing factors. For instance, the correlation coefficient between import and export volume and retention volume is 0.9810. Additionally, the correlation coefficients between total talent volume and patent volume and between total talent volume and energy consumption are 0.9633 and 0.9538, respectively. To avoid the complexity and overfitting caused by these high correlations, FPCA is used to reduce the dimensionality of the new energy market development factors. This process converts the high-dimensional data into

a low-dimensional principal component dataset. The principal components retain most of the original information and have low correlation, simplifying the neural network inputs and improving both the training process and model generalization ability. Thus, FPCA effectively addresses the problem of variable correlation in new energy market prediction. The principal component analysis reveals that the eigenvalues of the first three components decrease sharply, dominating the explanatory power. In contrast, the eigenvalues of the subsequent three components decrease more gradually and remain small. Therefore, to simplify the model while retaining the main information, it is more appropriate to extract the first two factors as principal components for analysis. These two principal components can then be used to construct the calculation expression by treating their eigen contributions as coefficients, with the corresponding influencing factors as independent variables.

$$\begin{aligned} 0.127F_1 =& -0.2269x_1 - 0.0332x_2 + 0.4253x_3 + 0.4593x_4 + 0.5391x_5 \\ & + 0.5145x_6 + 0.0157x_7 \\ 0.974F_2 =& \ 0.5552x_1 - 0.7567x_2 - 0.2281x_3 + 0.1589x_4 + 0.0673x_5 \\ & + 0.1694x_6 + 0.0931x_7 \\ -0.003F_3 =& -0.2727x_1 - 0.1035x_2 - 0.0222x_3 - 0.0893x_4 - 0.0314x_5 \\ & - 0.0249x_6 + 0.9512x_7 \end{aligned}$$

Where: F_1, F_2, and F_3 are principal component eigenvectors; x_1, x_2, x_3, x_4, x_5, x_6, and x_7 are standardized data.

2.4. Summary of the chapter

In this chapter, FPCA is applied to the weight analysis and correlation modeling of nine variables related to new energy vehicles. The principal components are extracted through dimensionality reduction, providing key data such as sales volume and vehicle loading volume. Next, the data are standardized, and a standardization matrix is established to analyze the cumulative contribution and correlation coefficients. The correlation coefficient graph and feature vector are then generated. The trends of the principal components are visualized, and Legendre polynomials are used for computational fitting, which effectively addresses the orthogonality and recurrence relationships in FPCA. Finally, using the two extracted principal component eigenvectors and the standardized data, principal component expressions are constructed to derive the functional relationships of FPCA.

3. Prediction by multiple linear regression

3.1. Establishing a functional equation for predicting the target variable from the raw data

According to the respective new energy vehicle ownership (10,000 units), annual patent applications for new energy power battery is as follows.

(10*7 MJ), total energy consumption of new energy vehicles (10*7 MJ), total carbon emissions of new energy vehicles (billion tons), import and export volume of new energy vehicles (10,000 units), number of relevant policies promulgated for new energy vehicles, total number of talents in the new energy vehicle industry, etc., respectively. Import and export volume of new energy vehicles (10,000 units), number of relevant policies enacted for new energy vehicles, total number of talents in the new energy vehicle industry (10,000)（The seven data on new energy vehicles as independent variables are analyzed and forecasted, and the annual installed capacity of power battery (Gwh) in the domestic market is analyzed and forecasted. The annual installed battery volume (Gwh) and the sales volume of new energy vehicles (10,000 units) are used as dependent variables, and the MLR is used to build a model for new energy vehicles [4].

The prediction model for new energy vehicles is established by adopting MLR, and the functional equations of the

prediction model are as follows.

Predictive function equation for the first dependent variable:

$$Y_{i+1} = -1.615 \times 10^{-14} - 2.5345 \times 10^{-18} \times X_1 - 2.4432 \times 10^{-14} \times X_2 \\ -5.839 \times 10^{-17} \times X_3 - 4.3043 \times 10^{-16} \times X_4 + 2.9383 \times 10^{-14} \times X_5 \\ -7.7578 \times 10^{-16} \times X_6 - 3.6123 \times 10^{-16} \times X_7$$

Predictive function equation for the second dependent variable

$$Y_{i+1} = 8.9877 \times 10^{-15} - 4.0552 \times 10^{-17} \times X_1 + 2.0523 \times 10^{-13} \times X_2 \\ -8.6084 \times 10^{-15} \times X_3 + 1.7659 \times 10^{-15} \times X_4 - 7.1382 \times 10^{-13} \times X_5 \\ +2.2609 \times 10^{-14} \times X_6 + 2.2458 \times 10^{-14} \times X_7$$

3.2. Establishing the principal component data predictive target variable function equation

According to our FPCA on the existing data of new energy vehicles, the study extracted the two dependent variables that have more influence on the annual installed capacity of power battery (Gwh) and the sales volume of new energy vehicles (10,000) in the domestic market. The study extracted two independent variables that have a greater influence on our domestic market, namely, the annual installed capacity of power batteries (Gwh) and the sales volume of new energy vehicles (10,000 units).

The MLR method is used again to establish a prediction model for new energy vehicles. The prediction function equations are as follows.

The prediction function equation for the first dependent variable:

$$Y_i + 1 = -1.6009 \times 10^{-14} - 2.5345 \times 10^{-18} \times X_1 - 2.4432 \times 10^{-14} \times X_2$$

Predictive function equation for the second dependent variable:

$$Y_i + 1 = 8.9877 \times 10^{-15} - 4.0552 \times 10^{-17} \times X_1 - 2.2966 \times 10^{-13} \times X_2$$

The first dependent variable is sales volume, and the second dependent variable is the production of electric batteries according to the (raw data). Since the confidence interval set in conducting the linear regression is 95%, the two independent variables are more significant below the significance level of 0.05.

3.3. Analysis of results

According to data from China's new energy vehicle market between 2010 and 2020, several key factors are considered: new energy vehicle ownership (in 10,000 vehicles), the annual installed capacity of power batteries (in GWh), the number of patent applications for new energy power batteries (in pieces), the total energy consumption of new energy vehicles (in 10^7 MJ), and the sales volume of new energy vehicles (in 10,000 units). Other factors include the carbon emissions of new energy vehicles (in 100 million tonnes), import and export volumes (in 10,000 units), the number of relevant policies, the total number of talents in the new energy vehicle industry (in 10,000), and the number of new energy vehicle industry organizations (in 10^7 MJ). Using these variables, the established multiple linear regression (MLR) prediction equation indicates that the target variables—sales volume and electric battery production—are expected to show an upward trend.

4. Analysis of new energy vehicle market forecasting results based on BP neural network and generalized regression neural network

To address local minima and slow convergence in BP neural networks (BPNN), the study proposes a prediction model based on the generalized regression neural network (GRNN). Using data from China's new energy vehicle market (2014-2023), the GRNN and chaotic BP models predict market trends, aiding policy and strategy development. Traditional forecasting methods like regression, gray modeling, and Markov forecasting often lack accuracy and exhibit significant lag, requiring qualitative corrections [5].

While BPNN struggles with small samples and noisy data, GRNN offers better approximation, faster learning, and greater accuracy, especially with limited data. Thus, GRNN is used to build a prediction model for new energy vehicle trends based on historical data.

4.1. BP neural network prediction

The fitness value of the individual in this case is represented by the BPNN prediction value, making the accuracy of BP prediction crucial for determining the optimal location. The optimization function is nonlinear, with two input parameters and one output parameter. Thus, the construction of an effective BP network is essential for identifying the optimal position [6].

For this optimization function, the BP network structure is designed as 2-5-1. It uses 10 sets of input-output data from the nonlinear function, randomly selecting 9 sets for training and reserving 1 set for testing. This approach allows for the evaluation of the BP network's fitting performance. The BP training process is shown in **Figure 4**.

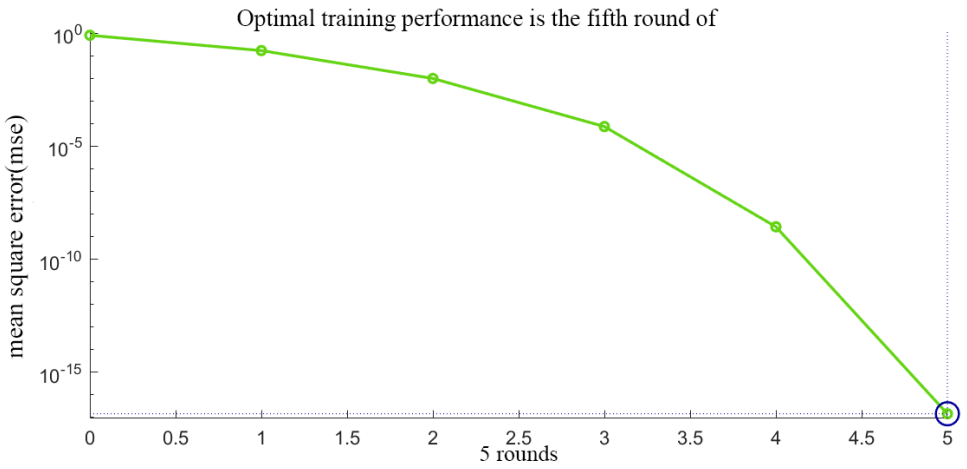

Figure 4. BP network performance graph

To eliminate scale and order of magnitude differences, the data were first standardized using MATLAB. The input set is the processed data, and the output set is the new energy vehicle data for prediction.

The input set is the processed data, and the output set is the new energy vehicle data for prediction. Generally, 80% of the data is used to train the network, and 20% is used for testing (**Figure 5**).

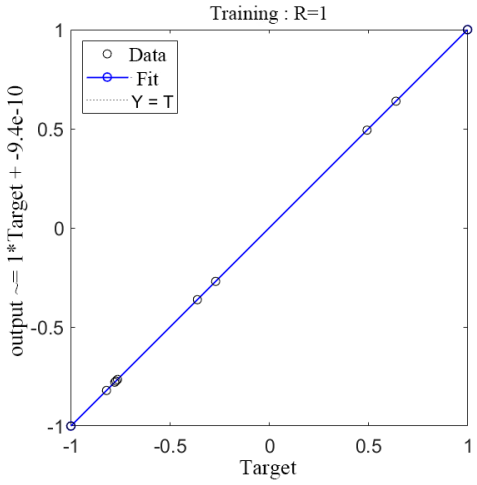

Figure 5. BP network predictive regression results plot

In this example, the number of neurons in the input layer is 3, so the number of neurons in the hidden layer is set to 7, from which a 3:7:1 network structure is formed. The network structure of 3:7:1 is formed. The training process of BPNN (**Figure 6**) shows that the BP network reaches the optimal training effect after 5 times of training, and its mean square error is $8.2322e^{-25}$, $1.4497e^{-17}$.

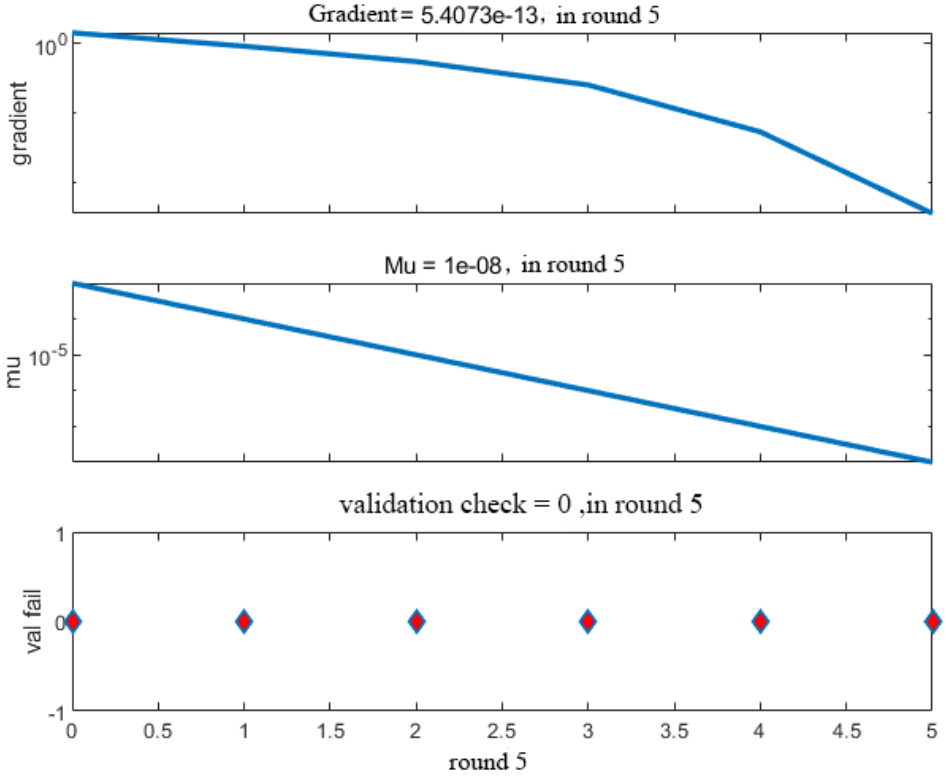

Figure 6. BP network training fit state diagram

4.2. GRNN projections

Under the Matlab2021a environment, Matlab language was used to write the algorithm calculation program, and the

Matlab neural network toolbox was applied to construct two kinds of prediction models: the GRNN-based new energy vehicle data prediction model (GRNN model) and the general chaotic BP prediction model (BP model), and to carry out prediction comparison experiments. The sample data were processed into a normalized time series with mean 0 and amplitude 1 according to the formula.

$$y_i = \left[x_i - \frac{1}{n} \sum_{j=1}^{m+1} x_i \right] / [max(x_i) - min(x_i)]$$

Given the limited sample size, the GRNN prediction model employs cross-validation for training and iterative validation to determine the optimal spread parameter [7]. The BP prediction model is structured with an 8-17-1 architecture, with training parameters set to 100 iterations, a target error of 0.0001, and a learning rate of 0.001.

To analyze the development of the new energy vehicle market, nine key indicators are selected as inputs for the GRNN model: ownership, battery patents, energy consumption, sales volume, carbon emissions, import and export volumes, number of policies, and total number of talents. The installed capacity and sales volume of power batteries serve as outputs. The model is trained using data from 2014 to 2022, while data from 2023 is used for testing. The prediction results are presented in **Table 2**.

4.3. Analysis of results

The MLR prediction model uses data from the previous year to estimate the following year's values, while the GRNN model applies cross-validation, using the first nine years of data to predict the final year. The GRNN model was trained nine times, generating nine prediction errors. A comparison of the errors in **Tables 2** and **3** shows that the GRNN model was used to predict the number of battery-loaded vehicles based on the original dataset.

The GRNN model achieved an average relative error of 2.23% in predicting battery loading, outperforming both the BP model and the multivariate linear regression (MLR) model. Its simple structure and single spread parameter contribute to fast and accurate predictions, making it well-suited for forecasting new energy vehicle data.

A comparison between the original data and the predicted data after FPCA dimensionality reduction indicates that predictions using the original data are more accurate than those after dimensionality reduction. As shown in **Table 3**, the GRNN model demonstrates higher prediction accuracy than both the BP and MLR models. Given the small sample size, this result is acceptable, as GRNN exhibits lower error when handling limited data. For the GRNN model, the choice of the spread parameter significantly affects performance: a smaller spread value enhances approximation to the sample, whereas a larger spread value smooths the approximation. In this experiment, setting the spread parameter to 1.4 produced the best prediction results, as illustrated in **Figure 7**.

Table 2. Comparison of prediction errors for battery loadings by three prediction methods

	Year/number of times	2015	2016	2017	2018	2019	2020	2021	2022	2023	Average value
Raw data	MLR	9.42	13.28	10.43	15.75	4.47	10.85	8.66	10.23	10.09	10.38
	BP	15.14	8.03	43.40	24.07	77.31	13.81	25.65	4.77	19.23	25.71
	GRNN	7.01	7.01	7.14	7.06	7.00	16.80	7.01	11.43	7.14	8.62
Downscaled data	MLR	158.96	188.56	166.22	156.35	136.68	231.63	320.80	179.29	140.48	186.55
	BP	155.17	191.60	153.69	168.05	150.68	226.63	142.03	156.58	143.35	167.14
	GRNN	134.93	139.68	220.61	144.60	139.68	150.89	155.35	156.86	134.93	153.06

Table 3. Comparison of forecasting errors of sales volume by three forecasting methods

	Year/number of times	2015	2016	2017	2018	2019	2020	2021	2022	2023	Average Value
Raw data	MLR	25.28	35.37	27.54	38.81	38.11	25.82	27.09	38.00	25.89	30.21
	BP	32.35	47.89	45.05	31.38	63.72	30.21	18.87	45.28	12.23	33.30
	GRNN	25.63	25.63	25.86	25.74	25.63	33.22	25.64	35.54	25.86	27.64
Downscaled data	MLR	288.26	290.63	380.65	290.54	305.51	290.96	350.21	310.25	280.52	309.72
	BP	328.58	387.71	277.65	336.69	288.63	337.10	273.63	286.98	327.30	316.03
	GRNN	273.58	294.60	364.90	279.44	294.60	291.64	328.02	303.11	273.58	300.39

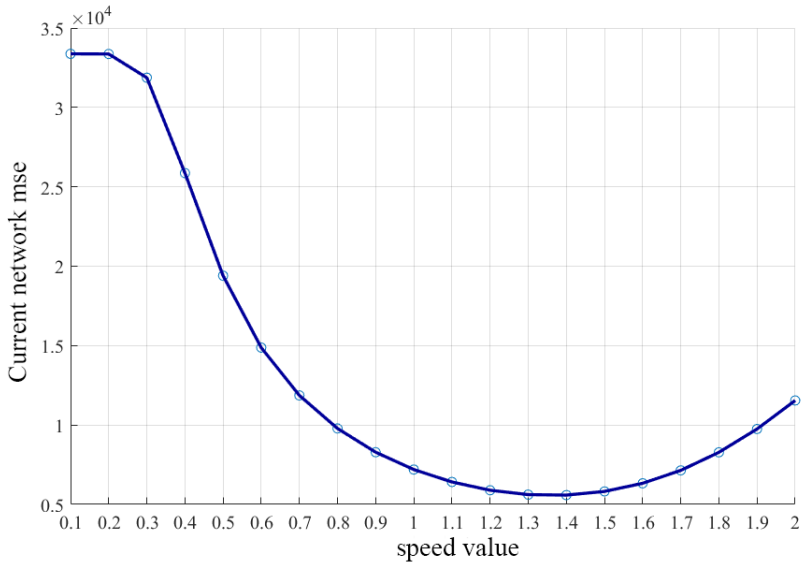

Figure 7. Variation curves of MSE value for networks with different spread values

4.4. Summary of the chapter

To address the issues of local minima and slow convergence in BPNN prediction, a prediction model based on GRNN is proposed. This model is applied to forecasting new energy vehicle data and compared with the BPNN model. Computational results demonstrate that the GRNN model exhibits good convergence, high accuracy, and strong practicality, making it a reliable tool for predicting new energy vehicle trends.

Compared with the BPNN model, the GRNN model offers several advantages: a simpler structure, the need to adjust only a single spread parameter, faster prediction speed, and the elimination of complex mathematical computations. These characteristics enhance its applicability and make it a promising approach for new energy vehicle prediction.

5. Summary and prospects

This study presents a comprehensive predictive analysis of new energy vehicle (NEV) market trends by integrating functional principal component analysis (FPCA) with multiple linear regression (MLR), back-propagation neural network (BPNN), and generalized regression neural network (GRNN). CiteSpace software was used to identify key terms and variables relevant to the study, while path analysis diagrams helped determine loading volume and sales

volume as primary research variables. During data analysis, FPCA was applied to reduce the dimensionality of high-dimensional data, effectively extracting the dataset's main features.

Building on this foundation, the MLR model was used for year-by-year prediction of the target variable. Additionally, BPNN and GRNN were employed to fit data from the first nine years, using the final year's data as a test set to forecast future market trends. By comparing the prediction errors of the three models, the one with the smallest error was selected as the final predictive tool [8]. This approach enables accurate forecasting of NEV market developments, providing a scientific basis for decision-making by policymakers, enterprises, and investors.

Future research can be expanded in several directions: (1) regularly updating data and iterating models to maintain prediction accuracy, (2) identifying and incorporating additional influencing factors within the NEV market, (3) exploring model fusion techniques such as ensemble learning to enhance predictive robustness, (4) conducting refined analyses of niche or regional markets, (5) performing comparative studies between the Chinese market and international counterparts to broaden the global perspective, (6) leveraging system dynamics simulations to support decision-making, and (7) investigating the environmental and social impacts of the NEV industry on sustainability. These efforts will deepen our understanding of the factors shaping the NEV market development and provide precise references for informed decision-making.

Disclosure statement

The authors declare no conflict of interest.

Reference

[1] Nosyk PN, Bacho R, 2024, Strategic Development Trends in the Automotive Industry of Ukraine. Engineering Proceedings, 79(1): 71.

[2] Feng J, Shi H, Ma D, et al., 2025, Supervised Functional Principal Component Analysis Under the Mixture Cure Rate Model: An Application to Alzheimer's Disease. Statistics in Medicine, 44(3–4): e10324.

[3] Fukushima T, 2018, Fast Computation of Sine/cosine Series Coefficients of Associated Legendre Function of Arbitrary High Degree and Order. Journal of Geodetic Science, 8(1): 162–173.

[4] Li R, Yang J, Cao Z, et al., 2025, A Multiple Linear Regression Model Calculates the Anterior Neck Skin Surface Area in Chinese Individuals: Implications for Neck Resurfacing and Reconstruction. Journal of Plastic, Reconstructive & Aesthetic Surgery, 2025(102): 363–372.

[5] Liu L, Wang G, Lan Y, et al., 2024, Predictive Model of Granular Fertilizer Spreading Deposition Distribution Based on GA-GRNN Neural Network. Drones, 9(1): 16.

[6] Ye J, Cai S, Zhang S, et al., 2025, Space Utilization Analysis of Classrooms for Individual Studies via Back Propagation Neural Network. Journal of Building Engineering, 2025(102): 111946.

[7] Yao D, Shen J, Hou J, et al., 2024, Online Vehicle Velocity Prediction Based on an Adaptive GRNN with Various Input Signals. International Journal of Automotive Technology, 2024(99): 1–13.

[8] Wang Z, Niu S, Fu C, et al., 2024, Advancing Data-driven Sustainable Design: A Novel NEV Form Design Approach in China's Market. Journal of Cleaner Production, 2024(461): 142626.