



Research article

A novel ensemble learning-based grey model for electricity supply forecasting in China

Yubin Cai¹ and Xin Ma^{1,2,*}

¹ School of Science, Southwest University of Science and Technology, Mianyang 621010, China

² School of Economics & Management, Southwest University of Science and Technology, Mianyang 621010, China

*** Correspondence:** Email: cauchy7203@gmail.com.

Abstract: Electricity consumption is one of the most important indicators reflecting the industrialization of a country. Supply of electricity power plays an import role in guaranteeing the running of a country. However, with complex circumstances, it is often difficult to make accurate forecasting with limited reliable data sets. In order to take most advantages of the existing grey system model, the ensemble learning is adopted to provide a new stratagy of building forecasting models for electricity supply of China. The nonhomogeneous grey model with different types of accumulation is firstly fitted with multiple setting of accumulation degrees. Then the majority voting is used to select and combine the most accurate and stable models validated by the grid search cross validation. Two numerical validation cases are taken to validate the proposed method in comparison with other well-known models. Results of the real-world case study of forecasting the electricity supply of China indicate that the proposed model outperforms the other 15 exisiting grey models, which illustrates the proposed model can make much more accurate and stable forecasting in such real-world applications.

Keywords: grey system; ensemble learning; grid search; electricity supply

Mathematics Subject Classification: 91D10

1. Introduction

Energy is the foundation of a country's economic development. Electricity is an essential energy source, and the electric energy policy will guide the development of a country to a certain extent [1]. Sustainable supply of electricity is one of the biggest challenges. Effective forecasting of electric energy supply has therefore become a prerequisite for formulating energy policies. If the electricity supply is underestimated, it will not only fail to meet the regular power demand of the country, but also pose a threat to the security of the electricity system. On the contrast, if the electricity supply

is overestimated, the national economy will suffer severe losses due to the difficulty of large-scale storage of electric energy. Therefore, it is important to have a high-precision system of electricity supply forecasting. Among numerous forecasting models, the complicated calculation process of support vector machines [2] and the demand of artificial neural networks [3] for data magnitude sets cannot always obtain the ideal forecast accuracy. Because the power supply is faced with the problems of small sample size and incomplete information, and grey system proposed by Deng is a tool to deal with them, so the forecasting method of grey system is suitable for solving such problems [4]. The grey system model has been applied in various fields in recent few years. For example, Zeng *et al.* applied a multivariate grey model to China's grain production [5]. Ding *et al.* applied the improved Simpson grey model to the prediction of electric vehicles through an adaptive method of generating dynamic weighted sequences [6]. In terms of environmental pollution, in order to predict carbon dioxide emissions in the BRICS countries, Wu *et al.* established a new conformable fractional-order nonhomogeneous grey model [7]. Hu *et al.* proposed a novel time-delayed fractional grey model, which took into account the time delay effect and applied it to the natural gas consumption forecast of the manufacturing industry in China [8].

Driven by the research of many scholars, the grey system has been developed vigorously, which is mainly reflected in the three aspects of accumulative order, background value optimization and expansion form. Wu *et al.* designed a grey model of fractional accumulation to eliminate the randomness of the original data [9]. Zhou *et al.* used the exponential weighted average method to define the cumulative generation of new information priority with parameters [10]. Ma *et al.* designed a conformable fractional grey model [12]. As more accumulation methods are proposed, the grey model can achieve stronger and more accurate predictions. Zeng *et al.* developed a structurally compatible multivariate grey model to improve compatibility [11]. Ma *et al.* applied the Simpson formula to the model and proposed a background value optimization model [13]. Wei *et al.* also established a new method of optimizing background value by using the integral median theorem [14]. With the discrete grey model proposed by Xie *et al.*, there are more ways to deal with various data types [15]. In order to fit a time series that approximates the nonhomogeneous exponential law, Cui *et al.* designed NGM to solve the limitations of the traditional grey model [16]. Chen *et al.* combined the Bernoulli equation to improve the prediction accuracy by controlling the power exponent n to adjust the curvature of the curve [17]. To overcome the defects of the traditional Verhulst model's misalignment substitution of parameters and the unreasonable selection of initial values, Zeng *et al.* proposed a new prediction model for tight gas production [18]. These methods of improving the grey model greatly expand its application range and enrich the grey system theory.

In response to the formulation of energy policies in developing countries, Li *et al.* applied the adaptive grey model to electricity consumption forecasting in the Asia-Pacific region and obtained effective results [1]. Xu *et al.* designed a grey model with an optimal time response function, used particle swarm optimization to optimize nonlinear parameters, and verified the reliability of the model with an example of China's electricity consumption [19]. Wang *et al.* focused on the problem of forecast stability and developed a hybrid forecast based on an improved grey forecasting model based on a multi-objective ant optimization algorithm, which dynamically selects the best input training set and takes the annual electricity consumption in various regions of China as the research object for practical evaluation [20]. Ding *et al.* used the particle swarm optimization algorithm to optimize the new initial conditions and combined with the rolling prediction mechanism to obtain the future trend of

China's total electricity consumption and industrial electricity consumption [21]. Liu *et al.* predicted the electricity consumption in China and India with a new grey multinomial model of time electricity term score [22]. Yu *et al.* developed a highly flexible grey model with time-delayed power-driven for photovoltaic power generation [23]. The above research shows that although the single improved model performs well in electric energy, the traditional single-model prediction method also has certain limitations in different studies.

As a branch of machine learning, ensemble learning aims to get better results than the single estimator [24]. The integration schemes have many types [25], roughly divided into classification and regression. Yu *et al.* used a simple addition strategy set to output the prediction results [26]. The dynamic ensemble model proposed by Chen applied to the prediction of wind speed showed strong competitiveness [27]. Dong *et al.* first built a decision tree to mine energy consumption patterns and used ensemble learning methods to establish building energy consumption prediction models for different modes [28]. Lin *et al.* designed an air quality monitoring tool by using multiple linear regression technology to integrate multiple deep learning prediction models [29]. At the level of classification, Arangarajan uses voting methods combined with discrete wavelet transform analysis to detect and classify different power quality disturbances [30]. The ensemble idea can not only effectively improve the prediction accuracy, but also has a lot of room for development in the integration of multiple models.

According to the literature study above, although there are many cases of combination of grey system and machine learning in recent years [31], ensemble learning has never been used for grey system models at present. With significant performance of ensemble learning to improve the existing machine learning models, it can also be expected to enhance the performance of the grey system models. Therefore, the novel ensemble strategy is proposed to integrate the different form of grey system models in this work. Then the ensemble learning-based grey model is used of the application of electricity supply in China and the effectiveness of the new model has been verified.

The remainder of this paper is organized as follows. Section 2 describes the accumulation methods and diverse forms of the grey model. Section 3 introduce the construction of the novel ensemble theory. In Section 4, the verification of numerical cases is given. The electricity supply of China is discussed in Section 5. Furthermore, the conclusions are shown in Section 6.

2. Method of grey model

The main methods used in this work will be presented in this section. Section 2.1 presents the different processing methods of the raw data in the modelling process. Section 2.2 presents the establishment of the nonhomogeneous grey model and discusses the form of different models.

2.1. The different accumulation methods

The method of processing raw data in the grey model determines the accuracy of the forecast. For a set of raw time series data $Y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)\}$, the corresponding accumulation generation sequence is defined as $Y^{(\aleph)} = \{y^{(\aleph)}(1), y^{(\aleph)}(2), \dots, y^{(\aleph)}(n)\}$, \aleph is the accumulative generator, the following three accumulation methods are used as the research direction of this paper.

- The first-order accumulation generation:

$$y^{(1)}(k) = \sum_{i=1}^k y^{(0)}(i), k = 1, 2, \dots, n. \quad (2.1)$$

The first-order inverse accumulation:

$$y^{(0)}(k) = y^{(1)}(k) - y^{(1)}(k-1), k = 2, 3, \dots, n. \quad (2.2)$$

- The fractional-order accumulation generation:

$$\begin{aligned} y^{(r)}(k) &= \sum_{i=1}^k \binom{r}{k-i} y^{(0)}(i) \\ &= \sum_{i=1}^k \binom{r+k-i-1}{k-i} y^{(0)}(i) \\ &= \sum_{i=1}^k \frac{(r+k-i-1)!}{(k-i)!(r-1)!} y^{(0)}(i) \end{aligned} \quad (2.3)$$

The fractional-order inverse accumulation:

$$\begin{aligned} y^{(0)}(k) &= \sum_{i=1}^k \binom{-r}{k-i} y^{(r)}(i) \\ &= \sum_{i=1}^k (-1)^{k-i} \binom{r}{k-i} y^{(r)}(i) \\ &= \sum_{i=1}^k (-1)^{k-i} \frac{r!}{(k-i)!(r-k+i)!} y^{(r)}(i) \end{aligned} \quad (2.4)$$

For the particular cases, the binomial coefficients are:

$$\begin{bmatrix} \mu \\ \nu \end{bmatrix} = \begin{cases} 0, \mu = 0 \\ 1, \mu = \nu = 0 \\ 0, \mu < 0, \nu > |\mu| \end{cases} \quad (2.5)$$

- The new information priority accumulation generation:

$$y^{(\lambda)}(k) = \begin{cases} y^{(0)}(1), k = 1 \\ \lambda \cdot y^{(\lambda)}(k-1) + y^{(\lambda)}(k), k = 2, 3, \dots, n \end{cases}, \lambda \in (0, 1). \quad (2.6)$$

The new information priority inverse accumulation:

$$y^{(0)}(k) = \begin{cases} y^{(\lambda)}(k) - \lambda \cdot y^{(\lambda)}(k-1), k = 2, 3, \dots, n \\ y^{(\lambda)}(1), k = 1 \end{cases}, \lambda \in (0, 1). \quad (2.7)$$

2.2. The base estimator

In this work, the nonhomogeneous grey model (NGM) will be selected as the basic estimator for ensemble learning because it has a simple form but is more flexible than the classic grey model. The NGM is based on multiple accumulations, which makes the form of the model more abundant and can fully reflect the various characteristics of the data. Therefore, in the process of ensemble learning, it can comprehensively incarnate the essence of the system to improve forecasting performance. Other common grey models are also given in the form of the NGM conversion.

2.2.1. The nonhomogeneous grey model

For the sequence generated by accumulation $Y^{(N)} = \{y^{(N)}(1), y^{(N)}(2), \dots, y^{(N)}(n)\}$, the first-order differential equation of the NGM is:

$$\frac{dy^{(N)}(t)}{dt} + \alpha y^{(N)}(t) = \beta t + \gamma, \quad (2.8)$$

where α is a development coefficient and $\beta t + \gamma$ is the grey action quantity.

And its discrete differential equation is:

$$y^{(N)}(t) - y^{(N)}(t-1) + \alpha z^{(N)}(t) = \beta t + \gamma, \quad (2.9)$$

where $z^{(N)}(t) = \frac{y^{(N)}(t) + y^{(N)}(t+1)}{2}$ is the sequence mean generated of consecutive neighbors of $y^{(N)}(t)$. Set:

$$\Psi = \begin{pmatrix} y^{(N)}(2) - y^{(N)}(1) \\ y^{(N)}(3) - y^{(N)}(2) \\ \vdots \\ y^{(N)}(n) - y^{(N)}(n-1) \end{pmatrix}, \Theta = \begin{pmatrix} -z^{(N)}(2) & 2 & 1 \\ -z^{(N)}(3) & 3 & 1 \\ \vdots & \vdots & \vdots \\ -z^{(N)}(n) & n & 1 \end{pmatrix}. \quad (2.10)$$

Then the least squares estimate of NGM satisfies:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = (\Theta^T \Theta)^{-1} \Theta^T \Psi. \quad (2.11)$$

The the continuous-time response function is:

$$\hat{y}^{(N)}(k) = (y^{(0)}(1) - \frac{\hat{\beta}}{\hat{\alpha}} - \frac{\hat{\gamma}}{\hat{\alpha}} + \frac{\hat{\beta}}{\hat{\alpha}^2}) e^{-\hat{\alpha}(k-1)} + \frac{\hat{\beta}}{\hat{\alpha}} k + \frac{\hat{\gamma}}{\hat{\alpha}} - \frac{\hat{\beta}}{\hat{\alpha}^2}, k = 1, 2, \dots, n. \quad (2.12)$$

The forecasting results $Y^* = \{\hat{y}^{(0)}(1), \hat{y}^{(0)}(2), \dots, \hat{y}^{(0)}(n)\}$ is obtained by the inverse accumulation.

2.2.2. Relationship between the nonhomogeneous grey model to other existing grey models

- Denote $\beta t + \gamma$ as \mathcal{W} , when the grey action quantity \mathcal{W} :

$$\mathcal{W} = \beta, \quad (2.13)$$

the NGM degenerates into GM (1,1), the continuous-time response function of GM is:

$$\hat{y}^{(N)}(k) = (y^{(0)}(1) - \frac{\hat{\beta}}{\hat{\alpha}})e^{-\hat{\alpha}(k-1)} + \frac{\hat{\beta}}{\hat{\alpha}}, k = 1, 2, \dots, n. \quad (2.14)$$

- Note the equation

$$\hat{y}^{(N)}(k+1) = \varrho_1 \hat{y}^{(N)}(k) + \varrho_2, \quad (2.15)$$

as the discrete form of GM, the recursive function of the discrete grey model (DGM) is:

$$\hat{y}^{(N)}(k) = \varrho_1^{k-1} (y^{(0)}(1) - \frac{\hat{\varrho}_2}{1 - \hat{\varrho}_1}) + \frac{\hat{\varrho}_2}{1 - \hat{\varrho}_1}, k = 1, 2, \dots, n. \quad (2.16)$$

- In the same way, the discrete form of NGM is:

$$\hat{y}^{(N)}(k+1) = \varrho_1 \hat{y}^{(N)}(k) + \varrho_2 k + \varrho_3, \quad (2.17)$$

and the recursive function of the nonhomogeneous discrete grey model (NDGM) is:

$$\hat{y}^{(N)}(k) = \varrho_1^{k-1} y^{(0)}(1) + \hat{\varrho}_2 \sum_{i=1}^{k-1} i \varrho_1^{k-1-i} + \frac{1 - \hat{\varrho}_1^{k-1}}{1 - \hat{\varrho}_1} \hat{\varrho}_3, k = 1, 2, \dots, n. \quad (2.18)$$

- When the grey action quantity \mathcal{W} :

$$\mathcal{W} = \beta \left(y^{(N)}(t) \right)^\tau, \quad (2.19)$$

where τ is the nonlinear parameter, and the continuous-time response function of the nonlinear grey Bernoulli model (NGBM) is:

$$\hat{y}^{(N)}(k) = (((y^{(0)}(1))^{1-\tau} - \frac{\hat{\beta}}{\hat{\alpha}})e^{-\hat{\alpha}(k-1)(1-\tau)} + \frac{\hat{\beta}}{\hat{\alpha}})^{\frac{1}{1-\tau}}, k = 1, 2, \dots, n, \quad (2.20)$$

when $\tau = 2$, the above model is Verhulst grey model.

The basic form and solution of models are shown in the Table 1. To sum up the above contents, it can be seen the NGM has the better generality because the NGM could degenerate into various basic models within such higher generality, it is used as the base estimator of ensemble learning could give full play to the advantages of several grey models.

3. The theory of ensemble grey model

Ensemble learning is a commonly used machine learning method that uses specific strategies to combine multiple learners to complete the learning task. Based on the conventional ensemble method, a novel ensemble strategy is constructed to improve prediction performance. The new strategy will simplify the ensemble method and apply it to the grey system to obtain forecasting results by combining several basic models. This section introduces the necessary elements of the conventional ensemble model in the first part, then gives an improved ensemble strategy, discusses the advantages of the new method in the second part, and ultimately introduces the construction of the ensemble grey model in this paper.

Table 1. The basic forms and the time response functions of the grey models.

Model	Basic form	Solution	Reference
Grey model	$y^{(N)}(t) - y^{(N)}(t-1) + \alpha z^{(N)}(t) = \beta$	$\hat{y}^{(N)}(k) = (y^{(0)}(1) - \frac{\hat{\beta}}{\hat{\alpha}})e^{-\hat{\alpha}(k-1)} + \frac{\hat{\beta}}{\hat{\alpha}}$	[4]
Nonhomogeneous grey model	$y^{(N)}(t) - y^{(N)}(t-1) + \alpha z^{(N)}(t) = \beta t + \gamma$	$\hat{y}^{(N)}(k) = (y^{(0)}(1) - \frac{\hat{\beta}}{\hat{\alpha}} - \frac{\hat{\gamma}}{\hat{\alpha}} + \frac{\hat{\beta}}{\hat{\alpha}^2})e^{-\hat{\alpha}(k-1)} + \frac{\hat{\beta}}{\hat{\alpha}}k + \frac{\hat{\gamma}}{\hat{\alpha}} - \frac{\hat{\beta}}{\hat{\alpha}^2}$	[16]
Discrete grey model	$\hat{y}^{(N)}(k+1) = \varrho_1 \hat{y}^{(N)}(k) + \varrho_2$	$\hat{y}^{(N)}(k) = \varrho_1^{k-1} (y^{(0)}(1) - \frac{\hat{\varrho}_2}{1-\hat{\varrho}_1}) + \frac{\hat{\varrho}_2}{1-\hat{\varrho}_1}$	[15]
Nonhomogeneous discrete grey model	$\hat{y}^{(N)}(k+1) = \varrho_1 \hat{y}^{(N)}(k) + \varrho_2 k + \varrho_3$	$\hat{y}^{(N)}(k) = \varrho_1^{k-1} y^{(0)}(1) + \hat{\varrho}_2 \sum_{i=1}^{k-1} i \varrho_1^{k-1-i} + \frac{1 - \hat{\varrho}_1^{k-1}}{1 - \hat{\varrho}_1} \hat{\varrho}_3$	[32]
Nonlinear grey Bernoulli model	$y^{(N)}(t) - y^{(N)}(t-1) + \alpha z^{(N)}(t) = \beta(z^{(N)}(t))^\tau$	$\hat{y}^{(N)}(k) = (((y^{(0)}(1))^{1-\tau} - \frac{\hat{\beta}}{\hat{\alpha}})e^{-\hat{\alpha}(k-1)(1-\tau)} + \frac{\hat{\beta}}{\hat{\alpha}})^{\frac{1}{1-\tau}}$	[17]

3.1. Conventional ensemble learning strategy

In the ensemble learning theory, the most significant operation is divided into the construction of the base model and the approach of ensemble. This section will present the ensemble learning from these two stages. In the first stage of the ensemble operation, a single data set is used by different base models to obtain a trained base model. In the second stage, the forecasting results of each model are combined by the averaging method to improve the forecasting accuracy.

3.1.1. Construction of the base model

The raw data sequence is assumed as $Y^{(0)} = \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)\}$, the base model is denoted by \hbar_e , and the parametric grid is \mathbb{S} . Divide the data set $Y^{(0)}$ into two parts: the modelling set Y_M to get the trained model and the forecasting set Y_F to compare the forecast performance of the model, which:

$$\begin{aligned} Y_M &= \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(m)\} \\ Y_F &= \{y^{(0)}(m+1), y^{(0)}(m+2), \dots, y^{(0)}(m+\varphi)\}, m+\varphi = n. \end{aligned} \quad (3.1)$$

The parameter χ' is selected from the parameter grid \mathbb{S} of model \hbar_e as a single base model $\hbar_e(\chi = \chi')$. By fitting the data on Y_M , the ensemble base model $\hat{\hbar}_e(\chi = \chi')$ can be obtained. The variety of different base model is obtained by the sampling times of χ in the parameter grid \mathbb{S} .

3.1.2. Ensemble approach

This paper adopts a simple averaging method to combine the forecasting results of base models. After the trained base models are obtained, the set of base models is recorded as \mathbb{E} and the results on the forecasting set Y_F is recorded as $\hat{Y}_F = \{\hat{y}^{(0)}(m+1), \hat{y}^{(0)}(m+2), \dots, \hat{y}^{(0)}(m+\varphi)\}$. Then the final ensemble forecasting results can be expressed by the following mathematical formula:

$$\hat{y}_H(k) = \frac{1}{\text{card}(\mathbb{E})} \sum_{e \in \mathbb{E}} \hat{y}_{\hat{\hbar}_e}^{(0)}(k), k = m+1, m+2, \dots, m+\varphi, \quad (3.2)$$

where $card(\mathbb{E})$ represents the cardinality of \mathbb{E} , $\hat{y}_{\hat{h}_e}^{(0)}(k)$ is the forecasting results of a single base model and $\hat{y}_H(k)$ is the forecasting results of ensemble learning.

3.2. Improved ensemble strategy

In the above-mentioned conventional ensemble operation, as the average prediction result is selected as the ensemble method, if there is a model with poor generalization ability in the input base model, it will pull down the performance of other models on this data set. Furthermore, if the sampling times in the parameter grid are plethoric, the number of base models that need to be input in is exceeding numerous. Therefore, we improved the ensemble strategy from the structure: first, the different parameter values χ is selected by the sample of the parameter grid \mathbb{S} with equal intervals s . Then the forecasting model with optimal parameters χ^* is obtained by setting the objective function and conditions of optimization. This section focuses on the search for the optimal basic model.

For the models with nonlinear parameters (such as r – *order*, λ – *order* or nonlinear parameter τ), the modelling set $Y_{\mathbb{M}}$ continue to be divided into two data blocks: the training set Y_{train} and the validation set Y_{valid} , which:

$$\begin{aligned} Y_{train} &= \{y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(\varsigma)\} \\ Y_{valid} &= \{y^{(0)}(\varsigma + 1), y^{(0)}(\varsigma + 2), \dots, y^{(0)}(\varsigma + \nu)\}, \varsigma + \nu = m. \end{aligned} \quad (3.3)$$

Y_{train} is used to estimate model parameters, Y_{valid} is used for validating the performance of the model out of the training sample. The objecive function is formulated as:

$$\min \Phi = \frac{1}{\nu} \sum_{j=\varsigma+1}^{\varsigma+\nu} |y^{(0)}(j) - \hat{y}_{\hat{h}}^{(0)}(j)|. \quad (3.4)$$

To obtain a relatively superior parameter value in the interval, the parameter values are sampled with equal intervals s to calculate the forecasting results of the model. According to the Section 2, we get the parameter values that satisfying the optimal conditions Eq (3.4) in the parameter interval \mathbb{S} as the approximate optimal parameters χ^* of the model, and take the trained model $\hat{h}_e(\chi = \chi^*)$ as the parametric base model of the ensemble model. The process of parameter searching is presented by Algorethm 1.

3.3. The ensemble of the nonhomogeneous grey model with a different accumulative approach

In this section, the ensemble learning-based grey model (ELGM) will be presented. The three accumulation forms of the nonhomogeneous grey model is used as the base models, i.e. $\mathbb{E} = \{NGM, FNGM, NIPNGM\}$. The ELGM is constructed by combining the forecasting results of each model with averaging method.

For the raw sequence $Y^{(0)}$, according to the Eqs (2.1), (2.3), (2.6), the accumulation sequence could be expressed as $Y^{(\aleph)}$ ($\aleph = 1, r, \lambda$), then construct $\Psi_e, \Theta_e (e \in \mathbb{E})$ with the method described in Section 2.2.1:

Algorithm 1**Input:**

Grey model with parameters \hat{h}_e ;
 Parameters to search χ ;
 The search interval corresponding to the parameter $\mathbb{S} = [\underline{\mathbb{S}}, \bar{\mathbb{S}}]$;
 Sampling interval s ;

Output:

The fitted model with parameters, $\hat{h}_e(\chi = \chi^*)$

- 1: **for** $i = \underline{\mathbb{S}} : s : \bar{\mathbb{S}}$ **do**
- 2: $\chi = i$;
- 3: Use the model $\hat{h}(Y_\chi)$ to fit the training set Y_{train} ;
- 4: Forecast the validation set Y_{valid}^* ;
- 5: Calculate $\Phi_{Y \in Y_{valid}^*}$;
- 6: **end for**
- 7: Use parameter χ^* when $\Phi^* = \min \Phi$ as the optimal search parameter of the model \hat{h}_e ;
- 8: **return** $\hat{h}_e(\chi = \chi^*)$.

$$\Psi_e = \begin{pmatrix} y^{(\aleph)}(2) - y^{(\aleph)}(1) \\ y^{(\aleph)}(3) - y^{(\aleph)}(2) \\ \vdots \\ y^{(\aleph)}(n) - y^{(\aleph)}(n-1) \end{pmatrix}, \Theta_e = \begin{pmatrix} -z^{(\aleph)}(2) & 2 & 1 \\ -z^{(\aleph)}(3) & 3 & 1 \\ \vdots & \vdots & \vdots \\ -z^{(\aleph)}(n) & n & 1 \end{pmatrix}. \quad (3.5)$$

According to Section 2.1, when $e = NGM$, we have $n = m$; when $e = FNGM, NIPNGM$, we have $n = \varsigma$. And:

$$\begin{pmatrix} \hat{\alpha}_e \\ \hat{\beta}_e \\ \hat{\gamma}_e \end{pmatrix} = (\Theta_e^T \Theta_e)^{-1} \Theta_e^T \Psi_e. \quad (3.6)$$

When $e = NGM$, directly calculate the value on Y_F and get the forecasting result $\hat{y}_{\hat{h}_e}^{(0)}$ through Eq (2.2):

$$\begin{cases} \hat{y}^{(1)}(k) = (y^{(0)}(1) - \frac{\hat{\beta}}{\hat{\alpha}} - \frac{\hat{\gamma}}{\hat{\alpha}} + \frac{\hat{\beta}}{\hat{\alpha}^2})e^{-\hat{\alpha}(k-1)} + \frac{\hat{\beta}}{\hat{\alpha}}k + \frac{\hat{\gamma}}{\hat{\alpha}} - \frac{\hat{\beta}}{\hat{\alpha}^2} \\ \hat{y}^{(0)}(k) = \hat{y}^{(1)}(k) - \hat{y}^{(1)}(k-1) \\ \hat{y}_{\hat{h}_e} = \{\hat{y}^{(0)}(m+1), \hat{y}^{(0)}(m+2), \dots, \hat{y}^{(0)}(m+\varphi)\} \end{cases} \quad (3.7)$$

When $e = FNGM, NIPNGM$, solve the optimized parameter \aleph^* ($\aleph = r, \lambda$):

$$\begin{aligned}
\min \Phi = & \frac{1}{v} \sum_{j=\varsigma+1}^{\varsigma+v} |y^{(0)}(j) - \hat{y}^{(0)}(j)| \\
\left\{ \begin{array}{l} \hat{y}^{(1)}(k) = (y^{(0)}(1) - \frac{\hat{\beta}}{\hat{\alpha}} - \frac{\hat{\gamma}}{\hat{\alpha}} + \frac{\hat{\beta}}{\hat{\alpha}^2} e^{-\hat{\alpha}(k-1)} + \frac{\hat{\beta}}{\hat{\alpha}} k + \frac{\hat{\gamma}}{\hat{\alpha}} - \frac{\hat{\beta}}{\hat{\alpha}^2} \\ \hat{y}^{(0)}(k) = \left\{ \begin{array}{l} \sum_{i=1}^k \begin{bmatrix} -\aleph \\ k-i \end{bmatrix} y^{(\aleph)}(i), \aleph = r \\ y^{(\aleph)}(k) - \aleph \cdot y^{(\aleph)}(k-1), \aleph = \lambda \end{array} \right. \\ k = 1, 2, \dots, \varsigma+v. \end{array} \right. \quad (3.8)
\end{aligned}$$

The optimal parameter \aleph^* is used to refit the training set to calculate the value on the Y_F . Then the forecasting result $\hat{y}_{\hat{h}_e}^{(0)}$ is obtained.

The forecasting results $\hat{y}_{\hat{h}_e}^{(0)}$ of \mathbb{E} is combined to get the ultimate results $\hat{y}_H(k)$ of ELGM by averaging method:

$$\hat{y}_H(k) = \frac{1}{\text{card}(\mathbb{E})} \sum_{e \in \mathbb{E}} \hat{y}_{\hat{h}_e}^{(0)}(k), k = m+1, m+2, \dots, m+\varphi. \quad (3.9)$$

And the process of building the ELGM can be represented by Figure 1.

4. Model validation

This section gives several numerical examples to verify the different accumulation forms of NGM under the new ensemble strategy. In each cases, the ELGM we proposed is compared with the different models in different accumulation forms presented in Section 2.2.2. In addition, the first section briefly presents the model evaluation criteria used in this work and the last section gives a summary of numerical examples.

4.1. Evaluation criteria for model prediction performance

To compare the generalization ability between the ELGM and other models, the prediction performance of the model is quantified by MAPEPR (Mean absolute percentage error for the prior-sample period), MAPEPO (Mean absolute percentage error for the post-sample period) and MAPE (Mean absolute percentage error) [33, 34]. The specific mathematical expressions are as follows:

$$MAPEPR = \frac{1}{m} \sum_{y^{(0)}(j) \in Y_{\mathbb{M}}} \left| \frac{\hat{y}^{(0)}(j) - y^{(0)}(j)}{y^{(0)}(j)} \right| \times 100\% \quad (4.1)$$

$$MAPEPO = \frac{1}{\varphi} \sum_{y^{(0)}(j) \in Y_{\mathbb{F}}} \left| \frac{\hat{y}^{(0)}(j) - y^{(0)}(j)}{y^{(0)}(j)} \right| \times 100\% \quad (4.2)$$

$$MAPE = \frac{1}{n} \sum_{y^{(0)}(j) \in Y^{(0)}} \left| \frac{\hat{y}^{(0)}(j) - y^{(0)}(j)}{y^{(0)}(j)} \right| \times 100\% \quad (4.3)$$

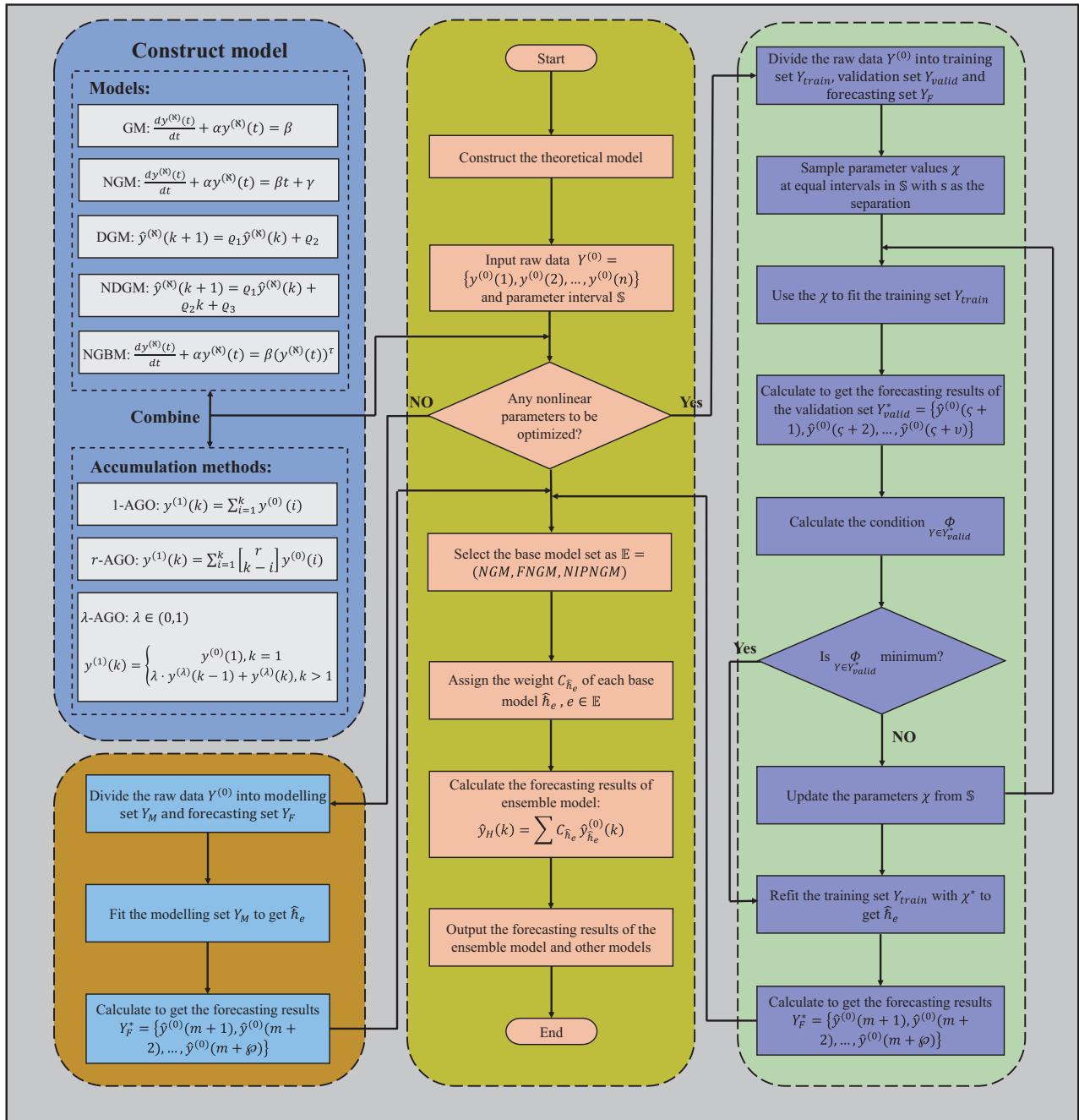


Figure 1. Ensemble process of nonhomogeneous grey model under three accumulation forms.

4.2. Numerical cases

Numerical case 1: (The renewable energy consumption of the Czech Republic) We consider a sample for establishing a grey model provided in the literature [35]. The raw data is divided by the parameter values given in the Table 2. And Figure 2 shows the comparison of the forecasting results of the five models with the smallest MAPE value. The MAPEPR, MAPEPO, and MAPE obtained from the ELGM are 5.93%, 2.82%, and 5.37%, respectively, which are better than the other comparison models. The novel model has the smallest evaluation criterion value in the modelling and forecasting stages, which shows that the ELGM is better than other competing models in this case.

Table 2. The data division parameters setting of the numerical cases.

Parameters	Values	
	Case 1	Case 2
n	11	9
m	9	8
φ	2	1
ς	8	7
v	1	1

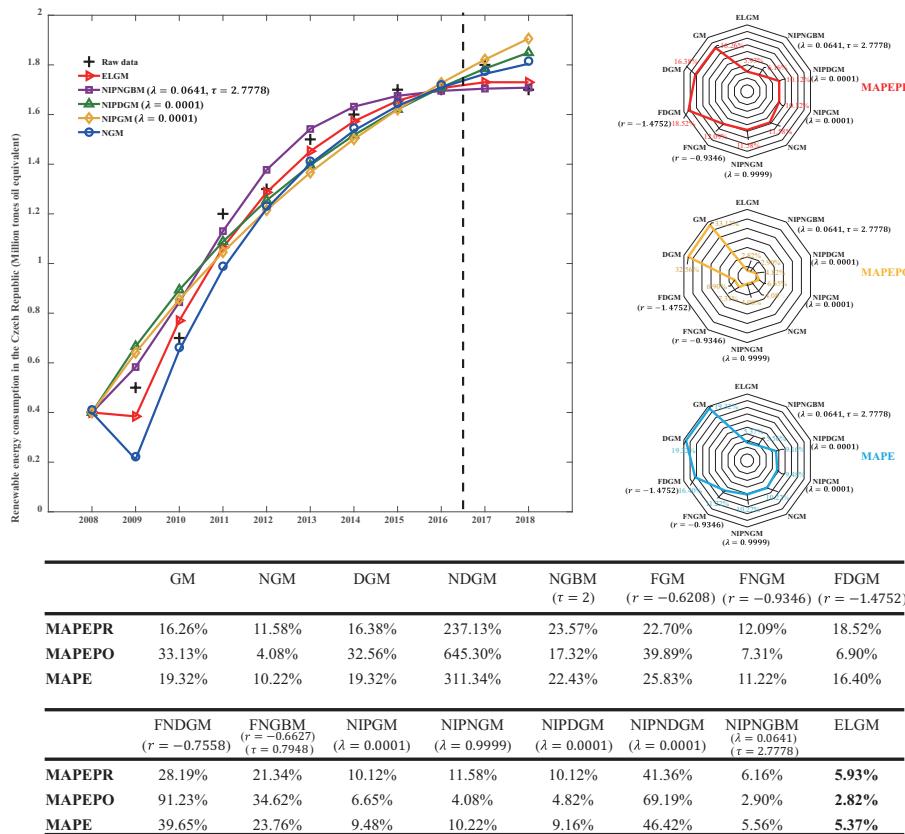


Figure 2. The forecasting results in numerical case 1.

Numerical case 2: (The biodiesel production of the United States) The data on biodiesel energy in the literature [36] is used to verify the grey model again. The data is divided as defined in the Table 2. Similar to numerical case 1, the forecasting results of the five models with the smallest MAPE value are plotted in the Figure 3. From the bottom of the figure, it is shown that the new model has not only achieved good results in the modelling and forecasting stages but also has a MAPEPO value of 0.73% in the forecasting stage, which is much better than other comparable models. It indicates that the ELGM could produce better predictive power than the base model under certain circumstances.

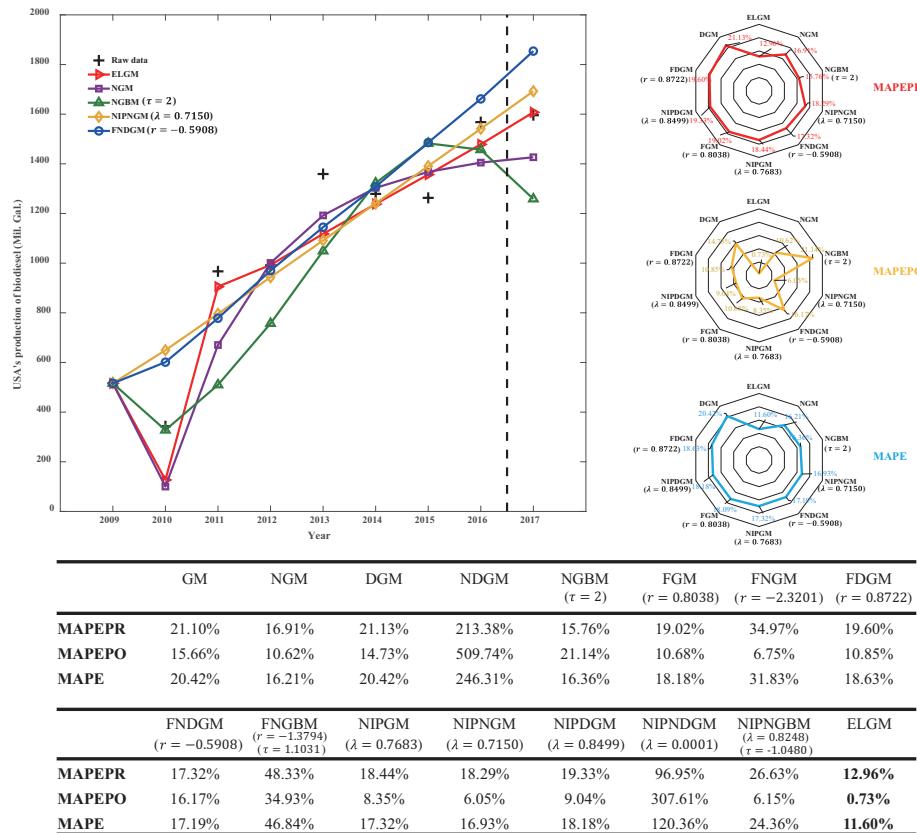


Figure 3. The forecasting results in numerical case 2.

4.3. Summary of the numerical cases

The performance of the ELGM in this section has been evaluated through two real numerical cases. It is worth noticing that these data not only contain monotonous changes in the degree of growth but also contain the characteristics of gradual fluctuations over time. The ELGM can also obtain the relatively best forecasting results among many models. However, there are also significant differences in the results obtained when dealing with changing trends over time. In numerical case 1, except for the ELGM, the NIPNGBM with the optimal parameters can also obtain similar prediction performance. But when we compared the forecasting results of the base model with the ELGM, we can conclude that the application of the novel ensemble strategy is, to a certain extent, optimize the forecasting results so that the prediction performance is further improved. In numerical case 2, when the ELGM

faced disturbance data, although the overall prediction accuracy is not as good as the application of monotonic trend data, it can even achieve ten times the prediction performance of other models.

In a nutshell, on the cases verified in this work, the ELGM has the lowest MAPE compared with the other fifteen models in the modelling and forecasting stages. It can be obviously seen that the ELGM can also show higher prediction accuracy in short-term prediction when processing the data with different feature.

5. Application of forecasting electricity supply in China

5.1. Data collection and division

The raw data of electricity supply in China is collected from the website [37] until the last update, which includes electricity supply data from 2000 to 2018 in China. The data is divided from a total of 19 time nodes from 2000 to 2018 by the division method introduced in Section 2.1 to obtain two data sets: modelling set and forecasting set without optimized parameter models; training set, validation set and forecasting set with optimized parameter models. The parameters of the division method are shown in the Table 3.

Table 3. The parameter settings of data division.

Parameters	Values
n	19
m	16
φ	3
ς	13
v	3

5.2. Forecasting results

In this paper, the three accumulation forms of NGM are used as three basic models to train and integrate the forecasting results of the models to get the final results. At the same time, the forecasting results of five different models under three accumulation forms are analyzed. The forecasting results, optimized parameter values and evaluation metrics values are shown in the Table 4. MAPEPR, MAPEPO and MAPE of the ELGM are 2.34%, 0.20% and 2.55%, respectively, which are the minimum values among the metrics, and it indicates that the ELGM has the best prediction performance.

Taking the MAPE value as the benchmark, the forecasting results of the minimum accumulation mode of MAPE value under the different accumulation of five models and the prediction charts of the ELGM results are given, as shown in Figure 4. It can be seen the prediction ability of the ELGM is slightly better than the optimal situation of other several models, which illustrates that the ensemble strategy can further improve the traditional prediction method and optimize the prediction results.

Table 4. The forecasting result of all models.

Year	Raw Data	ELGM	NIPGGM ($\lambda = 0.3833$)	NIPDGM ($\lambda = 0.3529$)	NIPNGM ($\lambda = 0.8776$)	NGM	NIPNGBM ($\lambda = 0.0010$)	FNGM ($\tau = 3$)	DGM	FDGM ($\tau = -0.4131$)	FNFDGM ($\tau = -1.0564$)	FNGBM ($\tau = 0.6987$)	FGM ($\tau = -1.4114$)	NGBM ($\tau = 2$)	NIPNDGM ($\lambda = 0.6113$)	NDGM	
2000	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70	13472.70		
2001	14724.10	14717.72	13346.77	13537.79	12822.43	12311.84	15465.80	17389.89	17800.53	17836.61	19400.69	15112.16	20937.65	3591.18	12788.38	13575.05	
2002	16466.00	16800.95	16485.92	16681.63	15542.20	15172.29	17731.64	19688.05	19503.90	19542.49	23814.24	17914.79	26029.23	4525.95	22553.03	33254.80	
2003	19032.20	19436.34	19644.99	19841.74	18096.23	18364.99	20274.00	21847.80	21370.26	21411.52	27523.33	20834.97	30013.09	30357.52	5690.64	63083.45	
2004	21972.30	22170.61	22824.12	23018.21	21285.12	21085.07	23100.20	24141.64	23415.22	23459.31	30806.99	23908.89	33263.85	34059.05	7133.95	44279.53	10820.04
2005	24940.80	25011.82	26023.43	26211.10	24298.14	24140.24	26205.13	26596.87	25655.86	25702.94	33796.78	27184.50	36299.78	8910.29	57095.63	175057.34	
2006	28588.40	27954.25	29243.06	29420.52	27400.30	27263.23	29566.49	29199.24	28110.91	28161.15	36566.05	30706.49	38939.24	40163.27	11077.62	7289.06	273101.86
2007	32712.40	30991.90	32483.12	32646.54	30455.53	33140.68	30586.86	31933.32	30800.90	30854.47	39160.30	34517.96	41354.18	42749.53	13693.28	90229.92	415275.02
2008	34540.80	34120.32	35743.76	35889.25	33854.08	33718.68	36860.56	34788.19	33748.29	33805.37	41609.99	38661.43	43592.01	45102.08	16806.92	111042.69	6320386.92
2009	37032.70	37336.35	39025.09	39148.74	37198.18	37054.47	40636.99	37756.60	36977.72	37038.49	43936.73	43179.29	45685.82	47259.89	20449.91	134796.08	916087.03
2010	41936.50	40637.74	42327.26	42425.09	40615.57	40463.88	44365.36	40833.75	40516.18	40580.83	46156.53	48113.92	47659.69	49252.99	24620.81	161565.56	134365.08
2011	47002.70	47022.77	45650.39	45718.40	44102.85	44102.85	47936.84	44016.29	44393.25	44461.96	48281.73	53507.83	49531.84	51105.13	29267.83	191430.77	196516.00
2012	49767.70	47490.12	48994.62	49028.74	47656.74	47511.84	51252.34	47301.80	48641.32	48714.27	50322.11	59403.68	51316.35	52835.48	34270.58	224465.17	2878066.65
2013	54204.10	51038.72	52360.08	52356.22	51274.14	51153.56	54225.33	50688.44	53295.89	53373.27	52285.65	65844.31	53024.43	54459.79	39426.11	260731.80	4230790.17
2014	57830.50	54667.66	55746.90	55700.91	54876.12	56839.85	54174.78	54877.86	58395.87	58477.86	54178.95	54665.13	55991.18	44446.34	300282.88	626069.06	300282.88
2015	58021.30	58376.23	59155.23	59062.90	58687.74	58681.31	59051.99	57759.64	63983.87	64070.64	56007.66	80532.08	56245.90	57440.68	48974.35	343160.69	9347788.22
MAPEPR	2.34%	2.48%	2.66%	3.47%	4.04%	5.44%	3.42%	5.01%	5.02%	16.35%	11.28%	21.07%	22.56%	39.04%	179.14%	2881.10%	
2016	61204.40	62163.80	62585.20	62442.29	62478.42	62570.95	60885.32	61442.03	70106.60	70198.32	57776.58	88865.76	57772.93	58817.67	52624.00	389398.84	14111306.64
2017	65914.00	66029.85	66036.94	65839.17	66321.55	66546.92	62373.04	65221.08	76815.22	76912.04	59489.94	97917.24	59251.48	60130.20	55040.07	439023.49	2157502.55
2018	71509.20	69975.96	69511.59	69253.62	70214.69	70611.14	63559.20	69096.05	84165.80	84267.86	61151.46	107730.21	60686.00	61385.21	55965.88	492054.50	33458456.52
MAPEPO	0.20%	0.28%	0.28%	0.24%	0.23%	0.90%	0.25%	2.57%	2.59%	1.57%	7.60%	1.62%	1.41%	2.75%	88.97%	5383.01%	
MAPE	2.55%	2.75%	2.93%	3.71%	4.28%	4.31%	5.69%	7.58%	7.61%	17.92%	18.88%	22.70%	23.97%	41.79%	268.11%	8264.11%	

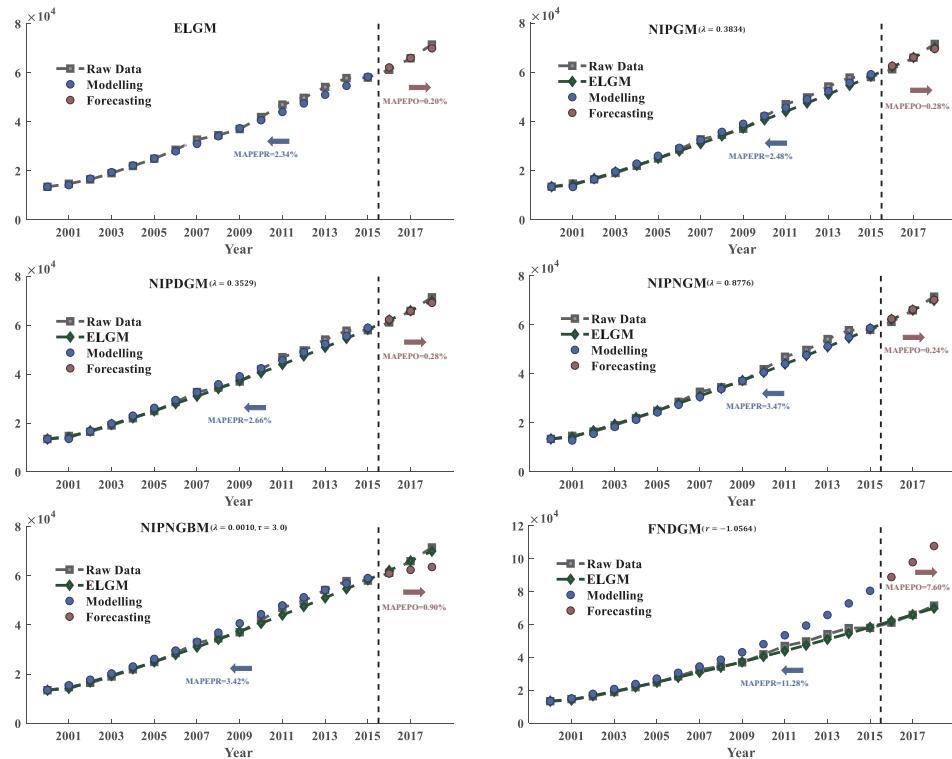


Figure 4. Comparison of optimal forecasting results between ELGM and various theoretical models.

The ELGM is compared with the original three basic models. It can be concluded from Figure 5 that the ELGM can not only ensure the minimum error between the forecasting set and the original data, but also make up for the poor prediction effect of the early samples due to the minimum error in the later samples.

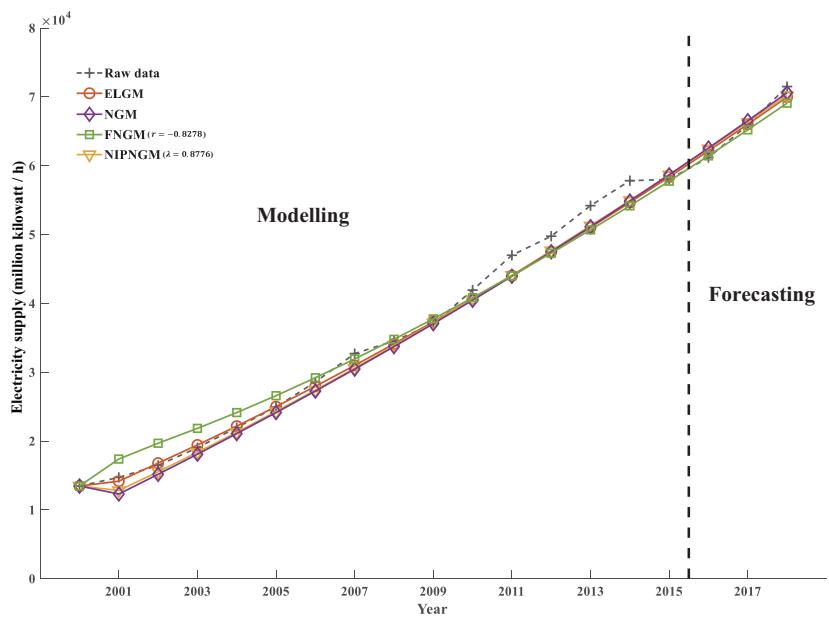


Figure 5. Comparison of forecasting results between ELGM and base models.

The evaluation criteria values of the top ten models with better prediction performance are visualized in Figure 6. From the evaluation criteria values in the figure, it can be clearly observed that the prediction of the ELGM is better than other models, which further verifies the feasibility of the ensemble learning-based grey prediction.

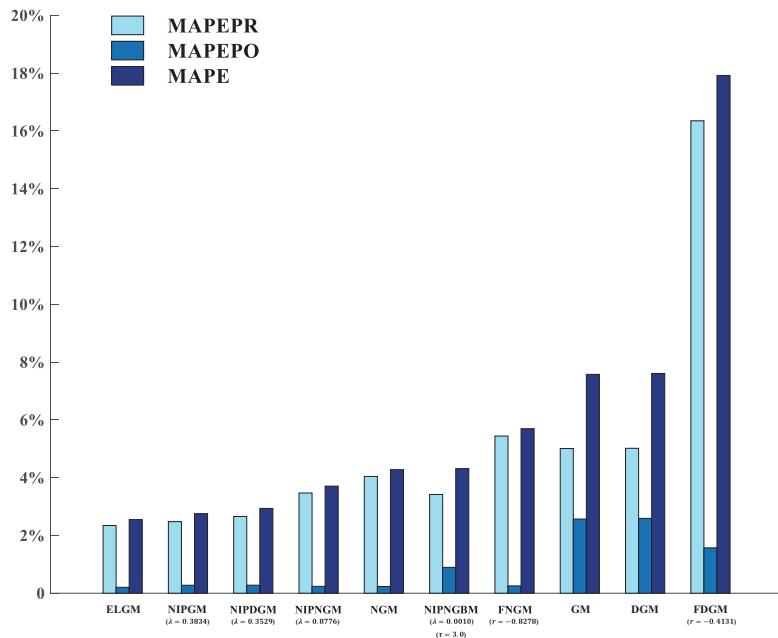


Figure 6. The MAPEPR, MAPEPO, MAPE of models with superior forecast performance.

5.3. Brief summary and short discussion

According to the forecasting results in Section 5.2, first of all, the models with the same basic form but different accumulation are called a group. The group of GM, DGM, or NGBM has a better fitting effect than the NGM group, but all of them are less effective than the NGM group in the forecasting stage, which indicating the NGM group has high stability. And the NDGM group performed worse in modelling and forecasting set. It could be seen the forecasting ability of ELGM benefits from the base estimator.

However, compared with the forecasting results of the basic models, the forecasting accuracy of ELGM obtained by the new ensemble strategy exceeds all basic models, which verifies that the strategy achieves the ensemble effect of ensemble learning while theoretically simplifying the input of the basic model.

The above discussion shows that the proposal of ELGM not only verifies that the effect of ensemble learning is very significant, but also provides a new way for the selection of ensemble strategies. The application results also fully illustrate that ELGM could be used as a reliable forecasting tool in the energy field.

6. Conclusions

A nonhomogeneous grey model based on a novel ensemble strategy is proposed in this work, abbreviated as ELGM. At first, the grid search cross validation is used to obtain the optimal

parameters of each basic model, and the nonhomogeneous grey models under the three accumulation forms are taken as the ensemble objects, and the ensemble is carried out in the way of average prediction results. In the case study, compared with fifteen general models, the results show that the ELGM has the highest generalization ability in small-sample time series forecasting, and can also have the same excellent performance when dealing with different trend data. The applied research of electricity supply shows that the proposed ELGM has the highest prediction accuracy in short-term electricity prediction and is expected to become a reliable tool for energy prediction.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (71901184, 71771033), Humanities and Social Science Fund of Ministry of Education of China (19YJCZH119), and Southwest University of Science and Technology University Student Innovation Fund Project (CX21-100).

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. D. C. Li, C. J. Chang, C. C. Chen, W. C. Chen, Forecasting short-term electricity consumption using the adaptive grey-based approach—An Asian case, *Omega*, **40** (2012), 767–773.
2. P. F. Pai, W. C. Hong, Support vector machines with simulated annealing algorithms in electricity load forecasting, *Energy Convers. Manage.*, **46** (2005), 2669–2688.
3. J. Yuan, C. Farnham, C. Azuma, K. Emura, Predictive artificial neural network models to forecast the seasonal hourly electricity consumption for a University Campus, *Sustainable Cities Soc.*, **42** (2018), 82–92.
4. J. L. Deng, Control problems of grey systems, *Syst. Control Lett.*, **1** (1982), 288–294.
5. B. Zeng, H. Li, X. Ma, A novel multi-variable grey forecasting model and its application in forecasting the grain production in China, *Comput. Ind. Eng.*, **150** (2020), 106915.
6. S. Ding, R. Li, Forecasting the sales and stock of electric vehicles using a novel self-adaptive optimized grey model, *Eng. Appl. Artif. Intell.*, **100** (2021), 104148.
7. W. Wu, X. Ma, Y. Zhang, W. Li, Y. Wang, A novel conformable fractional non-homogeneous grey model for forecasting carbon dioxide emissions of BRICS countries, *Sci. Total Environ.*, **707** (2020), 135447.
8. Y. Hu, X. Ma, W. Li, W. Wu, D. Tu, Forecasting manufacturing industrial natural gas consumption of China using a novel time-delayed fractional grey model with multiple fractional order, *Comput. Appl. Math.*, **39** (2020), 1–30.
9. L. Wu, S. Liu, L. Yao, S. Yan, D. Liu, Grey system model with the fractional order accumulation, *Commun. Nonlinear Sci. Numer. Simul.*, **18** (2013), 1775–1785.

10. W. Zhou, H. Zhang, Y. Dang, Z. Wang, New information priority accumulated grey discrete model and its application, *China J. Manage. Sci.*, **25** (2017), 140–148.
11. B. Zeng, H. Duan, Y. Zhou, A new multivariable grey prediction model with structure compatibility, *Appl. Math. Modell.*, **75** (2019), 385–397.
12. X. Ma, W. Wu, B. Zeng, Y. Wang, X. Wu, The conformable fractional grey system model, *ISA Trans.*, **96** (2020), 255–271.
13. X. Ma, W. Wu, Y. Zhang, Improved GM (1, 1) model based on Simpson formula and its applications, (2019), *ArXiv*: 1908.03493.
14. B. Wei, N. Xie, A. Hu, Optimal solution for novel grey polynomial prediction model, *Appl. Math. Modell.*, **62** (2018), 717–727.
15. N. Xie, R. Wang, N. Chen, Measurement of shock effect following change of one-child policy based on grey forecasting approach, *Kybernetes*, **47** (2018), 559–586.
16. J. Cui, Y. Dang, S. Liu, Novel grey forecasting model and its modeling mechanism, *Control Decis.*, **24** (2009), 1702–1706.
17. C. Chen, H. L. Chen, S. P. Chen, Forecasting of foreign exchange rates of Taiwan’s major trading partners by novel nonlinear Grey Bernoulli model NGBM (1, 1), *Commun. Nonlinear Sci. Numer. Simul.*, **13** (2008), 1194–1204.
18. B. Zeng, X. Ma, M. Zhou, A new-structure grey Verhulst model for China’s tight gas production forecasting, *Appl. Soft Comput.*, **96** (2020), 106600.
19. N. Xu, Y. Dang, Y. Gong, Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China, *Energy*, **118** (2017), 473–480.
20. J. Wang, P. Du, H. Lu, W. Yang, T. Niu, An improved grey model optimized by multi-objective ant lion optimization algorithm for annual electricity consumption forecasting, *Appl. Soft Comput.*, **72** (2018), 321–337.
21. S. Ding, K. W. Hipel, Y. Dang, Forecasting China’s electricity consumption using a new grey prediction model, *Energy*, **149** (2018), 314–328.
22. C. Liu, W. Wu, W. Xie, J. Zhang, Application of a novel fractional grey prediction model with time power term to predict the electricity consumption of India and China, *Chaos Solitons Fractals*, **141** (2020), 110429.
23. L. Yu, X. Ma, W. Wu, X. Xiang, Y. Wang, B. Zeng, Application of a novel time-delayed power-driven grey model to forecast photovoltaic power generation in the Asia-Pacific region, *Sustainable Energy Technol. Assess.*, **44** (2021), 100968.
24. S. B. Cho, Ensemble of structure-adaptive self-organizing maps for high performance classification, *Inf. Sci.*, **123** (2000), 103–114.
25. Q. Xiao, C. Li, Y. Tang, L. Li, Meta-reinforcement learning of machining parameters for energy-efficient process control of flexible turning operations, *IEEE Trans. Autom. Sci. Eng.*, **99** (2019), 1–14.
26. L. Yu, Y. Ma, M. Ma, An effective rolling decomposition-ensemble model for gasoline consumption forecasting, *Energy*, **222** (2021), 119869.

27. C. Chen, H. Liu, Dynamic ensemble wind speed prediction model based on hybrid deep reinforcement learning, *Adv. Eng. Inf.*, **48** (2021), 101290.
28. Z. Dong, J. Liu, B. Liu, K. Li, X. Li, Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification, *Energy Buildings*, **241** (2021), 110929.
29. C. Y. Lin, Y. S. Chang, S. Abimannan, Ensemble multifeatured deep learning models for air quality forecasting, *Atmos. Pollut. Res.*, **12** (2021), 101045.
30. A. Vinayagam, V. Veerasamy, P. Radhakrishnan, M. Sepperumal, K. Ramaiyan, An ensemble approach of classification model for detection and classification of power quality disturbances in PV integrated microgrid network, *Appl. Soft Comput.*, **106** (2021), 107294.
31. X. Ma, M. Xie, J. A. K. Suykens, A novel neural grey system model with Bayesian regularization and its applications, *Neurocomputing*, **456** (2021), 61–75.
32. N. Xie, S. Liu, Interval grey number sequence prediction by using non-homogenous exponential discrete grey forecasting model, *J. Syst. Eng. Electron.*, **26** (2015), 96–102.
33. L. Yu, X. Ma, W. Wu, Y. Wang, B. Zeng, A novel Elastic Net-based NGBMC (1, n) model with multi-objective optimization for nonlinear time series forecasting, *Commun. Nonlinear Sci. Numer. Simul.*, **96** (2021), 105696.
34. X. Xiang, X. Ma, M. Ma, W. Wu, L. Yu, Research and application of novel Euler polynomial-driven grey model for short-term PM10 forecasting, *Grey Syst. Theory Appl.*, (2020).
35. L. Liu, L. Wu, Forecasting the renewable energy consumption of the European countries by an adjacent non-homogeneous grey model, *Appl. Math. Modell.*, **89** (2021), 1932–1948.
36. S. A. Javed, B. Zhu, S. Liu, Forecast of biofuel production and consumption in top CO₂ emitting countries using a novel grey model, *J. Cleaner Prod.*, **276** (2020), 123997.
37. National Bureau of Statistics, Available from: <http://www.stats.gov.cn/>.



AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)