Mathematics

AIMS Mathematics, 10(12): 30507–30527.
DOI: 10.3934/math.20251338
Received: 29 October 2025
Revised: 13 December 2025
Accepted: 19 December 2025
Published: 25 December 2025

https://www.aimspress.com/journal/Math

Research article

Improved elastic net algorithm: A novel parameter identification method for grey system models

Qinzi Xiao[1,2], Mingyun Gao[3,4,*] and Congjun Rao[5]

[1] School of Management, Wuhan Institute of Technology, Wuhan, 430205, China
[2] Research Center for Coordinated Development of Enterprises and Environment, Wuhan Institute of Technology, Wuhan 430205, China
[3] School of Information Management, Central China Normal University, Wuhan, 430079, China
[4] Data Governance and Intelligent Decision Research Center, Central China Normal University, Wuhan, 430079, China
[5] School of Mathematics and Statistics, Wuhan University of Technology, Wuhan, 430070, China

* **Correspondence:** Email: Wh14_gao@126.com.

Abstract: Grey system model is widely employed to address uncertainties in systems characterized by small samples and poor information. In grey modeling, multicollinearity among variables often leads to ill-conditioned estimation, which undermines model stability. While the elastic net regularization method mitigates this issue by combining ridge and lasso regression, it still lacks the oracle property and adaptive group effect. This limitation restricts the applicability of grey models in handling larger datasets and capturing more flexible variable relationships. To overcome these shortcomings, this paper proposes an improved elastic net algorithm within the framework of a grey multivariate power model. The improvement in this paper aims to integrate adaptive lasso regression and correlation-driven penalty, optimize the weight adjustment mechanism of the penalty term, and enhance the adaptability and robustness of the grey model under its specific structure. Theoretically, the proposed algorithm is demonstrated to possess both the oracle property and adaptive group effect. Through an empirical analysis of the annual average of fine particulate matter ($PM_{2.5}$) concentrations in Beijing and Shanghai, China, with a comparison to parameter estimation results based on the least squares method and traditional regularization methods. The results show that the improved elastic net algorithm performs well in handling multicollinearity data, obtains more stable and accurate parameter estimates, and effectively improves the goodness of fit and prediction accuracy of the grey prediction model. This research provides a more powerful and reliable new approach for parameter identification of grey models.

Keywords: grey system; GM(1,N) power model; parameter identification; elastic net; $PM_{2.5}$ prediction
Mathematics Subject Classification: 60G70, 62G32, 65C05

Abbreviation

Abbreviation	Full name
GM(1,1)	Univariate grey model with the first-order accumulation generation
GM(1,N)	Grey multivariate model with the first-order accumulation generation
GM(1,N) power	Grey multivariate power model
1-AGO	First-order accumulation generation
IEN	Improved elastic net
OLS	Ordinary least squares
CP	Correlation-based penalty
L_1 regularization	Least absolute deviations norm regularization
L_2 regularization	Least squares norm regularization
ADMM	Alternating Direction Method of Multipliers
$PM_{2.5}$	Particulate matter 2.5
SO_4^{2-}	Sulfate
NO_3^-	Nitrate
NH_4^+	Ammonium
OM	Organic matter
BC	Black carbon
MAPE	Mean absolute percentage error
STD	Standard deviation of absolute percentage error
RMSE	Root mean square error
R^2	Coefficient of determination
PSO	Particle swarm optimization
GA	Genetic algorithm
WOA	Whale optimization algorithm

1. Introduction

Grey systems theory is a new method for studying the problem of uncertainty with limited samples and poor information [1]. Its core idea is to discover hidden patterns through data generation (such as accumulation) to achieve modeling with limited data [2]. Compared with other traditional statistical models and artificial intelligence models, the grey prediction models demonstrate better performance in short-term trend prediction of uncertain systems and have been widely applied in fields such as agriculture [3], energy [4], transportation [5], economy [6], and product quality [7].

The GM(1,1) model and GM(1,N) model are two typical grey prediction models. The former is a single-variable prediction model that achieves short-term trend prediction by constructing a first-order grey differential equation. In contrast, the latter is a multivariate analysis model that aims to depict the dynamic interactions between the core system variable and several influencing factors. It focuses more on systemic factor analysis than on accurate forecasting. Considering that most practical data exhibit nonlinear characteristics, Wang et al. [8] introduced a power exponent into the grey multivariate model and proposed the GM(1,N) power model. This model provides a unified and generalized framework that encompasses several fundamental grey system models. Specifically, it reduces to the standard GM(1,1) and GM(1,N) models under particular parameter configurations while simultaneously extending them to represent a broader class of nonlinear relationships. By integrating a tunable power exponent, the model gains the flexibility to adapt to diverse data patterns, which significantly enhances its applicability across various fields. Consequently, the GM(1,N) power model is regarded as a representative and comprehensive structure within grey system modeling, capable of capturing both univariate trends and complex multivariate interactions under nonlinear mechanisms. Based on this model, Xu et al. [9] utilized a linear combination structure to improve the background value of the model, thereby forecasting the nonlinear data of railway passenger traffic. At the same time, considering the impact of the driving factors' action mechanism on the prediction accuracy of the grey power model, Ding et al. [10] constructed a novel multivariate discrete grey power model by incorporating a linear driving control coefficient, which exhibits superior adaptability to the nonlinear relationships within the characteristic sequence.

The conventional GM(1,N) power model utilizes the OLS method to estimate its linear parameters [11]. However, when multicollinearity exists among the explanatory variables or when the original data are excessively large or small, the least squares estimation can become unstable with high variance, leading to significant weakening of the prediction accuracy, which is named as the ill-conditioned problems [12]. Beyond the OLS method, the parameter estimation framework for grey system models has also been extended to include nonlinear least squares approaches. Wei et al. demonstrated that grey models can be formulated within a state-space representation, where parameter estimation translates into a nonlinear regression problem [13]. Although offering advantages in estimation accuracy and robustness against noise, nonlinear least squares does not resolve the ill-conditioned problems from multicollinearity among explanatory variables or data with extreme scales. Hirose et al. noted that the introduction of regularization methods can effectively resolve the ill-conditioned problems of models [14]. L_1 regularization and L_2 regularization are the most commonly used regularization methods. L_1 regularization can perform variable selection and parameter estimation simultaneously, thereby endowing the model with high-precision sparsity. For example, considering that the selected variables may lack consistency, Zou et al. [15] consequently proposed the adaptive lasso regression, a novel method characterized by the oracle property. Meinshausen et al. [16] addressed the problem of slow convergence of lasso for high-dimensional data and designed the relaxed lasso method, effectively overcoming the contradiction between efficient computation and the rapid convergence rate of loss. Tibshirani et al. [17] combined the local constancy of the coefficient profile and designed a fused lasso method, which can adapt to the problem of sequences sorted in a specific way. Fan et al. [18] constructed the weighted lasso penalty quantile regression to address heavy-tailed high-dimensional data, which has strong robustness. Park and Sakaori [19] proposed a lag-weighted lasso regression to capture the lag effects in time series models. In summary, L_1 regularization promotes sparsity and variable selection but may induce excessive sparsity and underfitting, particularly when predictors are highly correlated. Conversely, L_2 regularization improves numerical stability and handles multicollinearity effectively, yet it retains all explanatory variables, which can limit interpretability and increase model complexity. Given that real-world data, such as environmental drivers in $PM_{2.5}$ forecasting, often exhibit both multicollinearity and potential redundancy, neither L_1 nor L_2 regularization alone provides a balanced solution.

The elastic net, a parameter estimation strategy that integrates both L_1 and L_2 regularization, offers a potential solution to the issues mentioned above [20]. Compared to OLS and single regularization parameter estimation methods, the elastic net demonstrates unique comprehensive advantages. These are primarily reflected in its adaptability to the complex structure of high-dimensional data and its balancing mechanism for model generalization performance, thereby embodying characteristics of both lasso and ridge regression. However, existing elastic net methods still face two key theoretical shortcomings.

First, the variable selection mechanism lacks the oracle property, resulting in inconsistent selected variables and affecting the accuracy of simulating real data. This deficiency limits the interpretative reliability and asymptotic efficiency of grey models under high-dimensional settings.

Second, the elastic net lacks an adaptive grouping effect, so it fails to account for correlations among variables, which can induce variable selection bias and inconsistent parameter estimation in short-term forecasting models, ultimately resulting in lower prediction accuracy and higher decision-making risks. The absence of this effect hinders the ability to automatically group correlated predictors, constraining its flexibility and scalability in grey system modeling, especially when handling larger datasets with intricate variable relationships.

To address these challenges, research in statistics has provided promising avenues. Ghosh took into account the oracle property and introduced adaptive lasso regression, which makes the variables selected by the improved elastic net consistent [21]. Van Le et al. assigned different parameters to the lasso and ridge penalties [22], respectively, effectively enhancing the accuracy. Furthermore, Anbari and Mkhadri combined the L_1 norm with a CP norm to address feature selection among multiple variables [23]. Daye et al. proposed weighted fusion for correlated data, and the new penalized regression can integrate the information among redundant correlated variables [24]. To simultaneously address both deficiencies of the elastic net, Fu et al. integrated adaptive lasso regression with weight-fused lasso and proposed a novel group variable selection algorithm with the oracle property [25], called WFAEN. These considerations confirm that the standard elastic net framework provides the necessary balance between variable selection and collinearity management, directly addressing the ill-conditioned estimation problems outlined earlier. Although the WFAEN algorithm represents an advanced iteration that incorporates both the oracle property and an adaptive grouping mechanism, its practical application remains constrained by computational intensity due to the exponential L_2 penalty and result variability stemming from inherent stochasticity.

To address the aforementioned issues, this paper proposes an IEN parameter estimation method that integrates adaptive lasso regression with CP based on the GM(1,N) power model. This estimation method is applied to the modeling and prediction of the annual average $PM_{2.5}$ concentrations in Beijing and Shanghai, China. The main innovations are as follows:

(1) In response to the problems of the existing elastic net regression in the regularization process, an IEN algorithm is proposed by integrating adaptive lasso regression with CP to optimize its penalty term weight adjustment mechanism. The novel method simultaneously possesses the oracle property and an adaptive grouping effect, meaning it considers both variable consistency and correlation during variable selection.

(2) By introducing the IEN algorithm and a parameter optimization algorithm into the GM(1,N) power model, a novel parameter estimation method for the grey prediction models is developed. This method considers the different characteristics of linear and nonlinear parameters, effectively enhancing the adaptability and robustness of the grey models under their specific structure and obtaining more

stable and precise parameter estimations.

(3) The GM(1,N) power model with the IEN is applied to the modeling and prediction of the annual average $PM_{2.5}$ concentrations in Beijing and Shanghai, China. Results reveal that the proposed IEN can effectively address the ill-conditioned problem caused by multicollinearity factors in the GM(1,N) power model, thereby improving the prediction accuracy and robustness of GM(1,N) power model.

The remainder of this paper is structured as follows: Section 2 introduces the GM(1,N) power model and traditional parameter identification methods. Section 3 presents the IEN parameter estimation method and explains its properties. Section 4 details the modeling and prediction of the annual average $PM_{2.5}$ concentrations in Beijing and Shanghai, China. Finally, Section 5 presents the conclusion.

2. GM(1,N) power model

Assuming the $i-\text{th}$ original sequence $X_i^{(0)}$ follows $X_i^{(0)}=\left\{x_i^{(0)}(1),x_i^{(0)}(2),\cdots,x_i^{(0)}(n)\right\}$, where $i=1,2,\cdots,N$, then these original sequences can make up the original sequence set $X^{(0)}=\left\{X_1^{(0)},X_2^{(0)},\cdots,X_N^{(0)}\right\}$. For $i-\text{th}$ sequence in this set, its 1-AGO is $X_i^{(1)}=\left\{x_i^{(1)}(1),x_i^{(1)}(2),\cdots,x_i^{(1)}(n)\right\}$, where $x_1^{(1)}(k)=\sum_{j=1}^{k}x_1^{(0)}(j),i=1,2,\cdots,N$. Then, the GM(1,N) power model can be defined in Eq (1).

$$x_1^{(0)}(k)+az_1^{(1)}(k)=\sum_{i=2}^{N}b_i\left(x_i^{(1)}(k)\right)^{r_i}, \tag{1}$$

where $x_1^{(0)}(k)$ means the grey derivative, $z_1^{(1)}(k)=\frac{1}{2}\left(x_1^{(1)}(k)+x_1^{(1)}(k-1)\right)$ means the background value, b_i is the driven coefficient on $x_i^{(1)}(k)$, and $r_i(i=2,\cdots,N)$ means the power index.

As defined in Eq (1), the GM(1,N) power model generalizes several classical grey models through specific parameter configurations. When $r_i=1$, this model reduces to the basic GM(1,N) model [26], i.e., $x_1^{(0)}(k)+az_1^{(1)}(k)=\sum_{i=2}^{N}b_ix_i^{(1)}(k)$. When $r_i=0$, it simplifies further to the basic GM(1,1) model [27], i.e., $x_1^{(0)}(k)+az_1^{(1)}(k)=b$. This structural flexibility allows the GM(1,N) power model to adapt to various modeling scenarios.

For the GM(1,N) power model, its whitening differential equation follows Eq (2).

$$\frac{dx_1^{(1)}(t)}{dt}+ax_1^{(1)}(t)=\sum_{i=2}^{N}b_i\left(x_i^{(1)}(k)\right)^{r_i}. \tag{2}$$

Based on Eq (2), the time response function of the GM(1,N) power model follows Eq (3).

$$\hat{x}_1^{(1)}(k)=\left[x_1^{(0)}(0)-\frac{1}{a}\sum_{i=2}^{N}b_i\left(x_i^{(1)}(k-1)\right)^{r_i}\right]e^{-a(k-1)}+\frac{1}{a}\sum_{i=2}^{N}b_i\left(x_i^{(1)}(k-1)\right)^{r_i}. \tag{3}$$

Through this function, the prediction of $x_1^{(0)}$ during the period k follows $\hat{x}_1^{(0)}(k+1)=\hat{x}_1^{(1)}(k+1)-\hat{x}_1^{(1)}(k)$. As listed in Eq (3), the prediction main depends on the parameters set $\left[a,b_2,b_3,\cdots,b_N\right]^T$ and power index set $\left[r_2,r_3,\cdots,r_N\right]^T$. For the parameters set $\left[a,b_2,b_3,\cdots,b_N\right]^T$, it is estimated through the OLS method in existing reference [6], seen in Eq (4).

$$\hat{P} = \left(B^T B\right)^{-1} B^T Y \,, \tag{4}$$

where

$$P = \begin{bmatrix} a \\ b_2 \\ \vdots \\ b_N \end{bmatrix}, \quad B = \begin{bmatrix} -z_1^{(1)}(2) & \left(x_2^{(1)}(2)\right)^{r_2} & \cdots & \left(x_2^{(1)}(2)\right)^{r_N} \\ -z_1^{(1)}(3) & \left(x_2^{(1)}(3)\right)^{r_2} & \cdots & \left(x_2^{(1)}(3)\right)^{r_N} \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(n) & \left(x_2^{(1)}(n)\right)^{r_2} & \cdots & \left(x_N^{(1)}(n)\right)^{r_N} \end{bmatrix}, \quad Y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{bmatrix}.$$

3. Parameter identification for GM (1,N) power model

The GM(1,N) power model contains a total of $2N-1$ parameters, categorized into linear parameters (including a, b_i) and nonlinear parameters (including r_i, $i = 2,3,\cdots,N$). This section presents different identification strategies for each type. First, the limitations of existing estimation methods are discussed. Subsequently, the IEN estimator is introduced for linear parameter identification. Finally, an optimization-based methodology is provided for determining these nonlinear parameters.

3.1. Linear parameter identification based on IEN estimator

As shown in Eq (4), the conventional GM(1,N) power model uses the OLS method for linear parameter estimation. This method requires low correlation among variables. However, this condition is often not met in real applications with small sample sizes. Moreover, the variability of nonlinear parameter values further exacerbates multicollinearity issues. Therefore, a new parameter estimator is proposed to address these issues.

3.1.1. IEN estimator

To address the limitations of the OLS estimation, regularization methods have been widely adopted to enhance model generalizability [28]. Commonly used approaches include the lasso estimator, ridge estimator, and elastic net estimator. These techniques mitigate overfitting and improve numerical stability by incorporating specific penalty terms into the estimation process.

The lasso estimator enhances the OLS framework by incorporating an L_1 regularization penalty term into the objective function $\text{Lasso}(P,\lambda_1)$. This formulation is explicitly given in Eq (5).

$$\text{Lasso}(P,\lambda_1) = \min_P \left[\|Y - B \cdot P\|_2^2 + \lambda_1 I_N^T \text{abs}(P) \right], \ \ \lambda_1 > 0 \,, \tag{5}$$

where λ_1 is the regularization parameter, $I_N = [1,1,\cdots,1]^T$ is a N-by-1 matrix of ones, and $\text{abs}(P)$ is the absolute value matrix of P. Usually, the L_1 penalty (lasso) promotes sparsity. It can drive the coefficients of irrelevant or highly noise-corrupted variables to exactly zero, effectively performing feature selection and building a more interpretable model that is robust to redundant noisy predictors.

The ridge estimator enhances the OLS framework by introducing an L_2 regularization penalty term. The modified objective function, $\text{Ridge}(P,\lambda_2)$, is formulated in Eq (6).

$$\text{Ridge}(P,\lambda_2) = \min_P \left\{ \|Y - B \cdot P\|_2^2 + \lambda_2 P^T P \right\}, \ \ \lambda_2 > 0 \,. \tag{6}$$

This penalty promotes parameter shrinkage and effectively mitigates multicollinearity issues in the estimation process. Meanwhile, this method can also reduce the sensitivity of parameter estimators to measurement noise by uniformly shrinking all coefficients toward zero, thereby stabilizing the solution and decreasing model variance at the cost of introducing slight bias.

Zou and Hastie [20] integrated the properties of L_1 and L_2 regularization to propose the elastic net estimator. Its mathematical formulation is expressed in Eq (7).

$$\min_P S^{\mathrm{EN}} = \min_P \left\{ \|Y - B\cdot P\|_2^2 + \lambda_1 I_N^T \mathrm{abs}(P) + \lambda_2 P^T P \right\},\ \ \lambda_1 > 0, \lambda_2 > 0\,. \tag{7}$$

Combing the strengths of both L_1 and L_2 regularization, the elastic net mitigates the effects of measurement noise. Moreover, this estimator effectively performs variable selection while mitigating underfitting, making it a powerful regularization tool. However, it suffers from two key limitations:

(1) It lacks the oracle property, which compromises variable selection consistency.

(2) It does not incorporate adaptive group effects to account for inter-variable correlations.

To address these issues, Zou [15] developed the adaptive lasso, which ensures oracle properties, while Wang et al. [29] integrated correlation-based weights into the L_1 regularization penalty to explicitly model variable dependencies. Building on these advances, this study combines both enhancements into a unified penalty function, yielding the IEN estimator formulated in Eq (8).

$$\min_P S^{IEN} = \min_P \left\{ \|Y - B\cdot P\|_2^2 + \lambda_1 W^T \mathrm{abs}(P) + \lambda_2 P^T P + \lambda_3 P^T M P \right\},\ \ \lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0\,, \tag{8}$$

where $W = [w_1, w_2, \cdots, w_N]^T$ means the weight vector, the quadratic form $P^T MP$ means the correlation based penalty, and M is a positive matrix with the element as

$$m_{ij} = \begin{cases} \sum_{s\neq i} \dfrac{2}{1-\rho_{is}^2}, if \quad i = j, \\ -\dfrac{2\rho_{ij}}{1-\rho_{ij}^2}, if \quad i \neq j. \end{cases} \tag{9}$$

In Eq (9), ρ_{ij} is the element from the correlation matrix $\boldsymbol{\rho}$, $\boldsymbol{\rho} = (\mathrm{diag}(C))^{-\frac{1}{2}} I_N^T C I_N (\mathrm{diag}(C))^{-\frac{1}{2}}$, $C = \dfrac{1}{n-2}\left(B^T B - \dfrac{1}{n-1} B^T I_{(n-1)\times(n-1)} B \right)$ means the covariance matrix, $I_{(n-1)\times(n-1)}$ is an $(n-1)\times(n-1)$ matrix of ones, and $(\mathrm{diag}(C))^{-\frac{1}{2}}$ represents the diagonal matrix whose entries are the element-wise reciprocal square roots of the diagonal elements of C.

To simplify the parameter notation, let $\varepsilon_1 = a$ and $\varepsilon_i = b_i$, $i = 2,3,\cdots N$, then $P = \varepsilon$, where $\varepsilon = [\varepsilon_1, \varepsilon_2, \varepsilon_3, \cdots, \varepsilon_N]^T$. Therefore, the IEN estimator for these linear parameters in the GM(1,N) power model follows Eq (10).

$$\hat{\varepsilon}^{\mathrm{IEN}} = \arg\min_\varepsilon \left\{ \|Y - B\varepsilon\|_2^2 + \lambda_1 W^T \mathrm{abs}(\varepsilon) + \lambda_2 \varepsilon^T \varepsilon + \lambda_3 \varepsilon^T M \varepsilon \right\}. \tag{10}$$

In Eq (10), $\lambda_2 \varepsilon^T \varepsilon$ and $\lambda_3 \varepsilon^T M \varepsilon$ are both the 2-norm on ε. Assuming that $Q = M + D$, where $D = diag(\lambda_2, \lambda_2, \cdots, \lambda_2)$, it is obvious that Q is the real symmetric positive definite square matrix when $\rho_{ij}^2 \neq 1$. Thus, Q can be decomposed into L^T and L, i.e., $Q = L^T L$. As for the weight vector W, it satisfies the following two conditions.

Condition 1. w_j is a monotonic nondecreasing function on j.

Condition 2. The weight function w_j should assign smaller penalties to $|\varepsilon_j|$ with larger absolute values and larger penalties to $|\varepsilon_j|$ with smaller absolute values.

Condition 1 is derived from the new information priority principle in grey modeling, and earlier data points exhibit diminishing predictive power for future states. Condition 2 ensures that significant parameters are preserved for their substantial contributions to model prediction, while less influential ones are more strongly regularized to enhance model sparsity and generalizability. Based on the above conditions and that $w_j = \frac{2}{\left(1+j^{-\gamma_1}\right)\left|\hat{\varepsilon}_j^*\right|^{\gamma_2}}$ in this study, γ_1 and γ_2 are undetermined positive constants.

3.1.2. Properties of IEN estimator

Theorem 3. For data matrixes $\boldsymbol{Y}, \boldsymbol{B}$ and penalty parameters (including $\lambda_1, \lambda_2, \lambda_3$), assuming that the loss function of the IEN estimator is $L(\lambda_1,\lambda_2,\lambda_3,\varepsilon)$, namely $L(\lambda_1,\lambda_2,\lambda_3,\varepsilon) = \|Y - B\varepsilon\|_2^2 + \lambda_1 W^T \text{abs}(\varepsilon) + \lambda_2 \varepsilon^T \varepsilon + \lambda_3 \varepsilon^T M \varepsilon$, it follows Eq (11).

$$L(\lambda_1,\lambda_2,\lambda_3,\varepsilon) = \left\|\bar{Y} - B^*\delta^*\right\|_2^2 + k_\lambda \sum_{j=1}^{N} \left|\delta_j^*\right|, \tag{11}$$

where $\bar{Y}_{(n+N-1)\times 1} = \begin{pmatrix} Y_{(n-1)\times 1} \\ O_{N\times 1} \end{pmatrix}$, $\bar{B}_{(n+N-1)\times N} = (1+\lambda_3)^{-\frac{1}{2}} \begin{pmatrix} B_{(n-1)\times N} \\ \sqrt{\lambda_3} L^T{}_{N\times N} \end{pmatrix}$, $b_j^* = \frac{\bar{b}_j}{\hat{w}_j}$, , $\delta = \sqrt{1+\lambda_3}\,\varepsilon$,and $\delta_j^* = \delta_j \cdot \hat{w}_j$.

Proof. Based on the above assumptions, we hold that

$$\begin{aligned}
\left(\bar{Y} - B^*\delta^*\right)^T \left(\bar{Y} - B^*\delta^*\right) &= \bar{Y}^T\bar{Y} - \bar{Y}^T B^*\delta^* - \left(B^*\delta^*\right)^T \bar{Y} + \left(B^*\delta^*\right)^T B^*\delta^* = \bar{Y}^T\bar{Y} - 2\bar{Y}^T B^*\delta^* + \left(B^*\delta^*\right)^T B^*\delta^* \\
&= \begin{pmatrix} Y \\ O \end{pmatrix}^T \begin{pmatrix} Y \\ O \end{pmatrix} - 2\begin{pmatrix} Y \\ O \end{pmatrix}^T \begin{pmatrix} B\varepsilon \\ \sqrt{\lambda_3} L^T \varepsilon \end{pmatrix} + \begin{pmatrix} B\varepsilon \\ \sqrt{\lambda_3} L^T \varepsilon \end{pmatrix}^T \begin{pmatrix} B\varepsilon \\ \sqrt{\lambda_3} L^T \varepsilon \end{pmatrix} \\
&= Y^T Y - 2Y^T B\varepsilon + (B\varepsilon)^T (B\varepsilon) + \lambda_3 \varepsilon^T L L^T \varepsilon,
\end{aligned}$$

and $k_\lambda \sum_{j=1}^{N} \left|\delta_j^*\right| = \frac{\lambda_1}{\sqrt{1+\lambda_3}} \sum_{j=1}^{N} \hat{w}_j \left|\sqrt{1+\lambda_3}\,\varepsilon_j\right|$.

Considering that $\left\|\bar{Y} - B^*\delta^*\right\|_2^2 + k_\lambda \sum_{j=1}^{N} \left|\delta_j^*\right| = \left(\bar{Y} - B^*\delta^*\right)^T \left(\bar{Y} - B^*\delta^*\right) + k_\lambda \sum_{j=1}^{N} \left|\delta_j^*\right|$, we have

$$\begin{aligned}
\left\|\bar{Y} - B^*\delta^*\right\|_2^2 + k_\lambda \sum_{j=1}^{N} \left|\delta_j^*\right| &= Y^T Y - 2Y^T B\varepsilon + (B\varepsilon)^T (B\varepsilon) + \lambda_3 \varepsilon^T L L^T \varepsilon + \lambda_1 \sum_{j=1}^{N} \hat{w}_j \left|\varepsilon_j\right| \\
&= \|Y - B\varepsilon\|_2^2 + \lambda_1 W^T \text{abs}(\varepsilon) + \lambda_2 \varepsilon^T \varepsilon + \lambda_3 \varepsilon^T M \varepsilon.
\end{aligned}$$

Theorem 3 establishes that the proposed IEN estimator can be equivalently transformed into an L_1 regularized formulation. This reformulation bridges the proposed method with the well-established theoretical framework of the lasso estimator, seen in Eq (12).

$$L(\lambda_1,\lambda_2,\lambda_3,\varepsilon) = \text{Lasso}(\delta^*, k_\lambda). \tag{12}$$

Consequently, existing analytical tools and computational strategies for the lasso estimator can be directly leveraged to analyze the statistical properties and optimize the implementation of the IEN estimator.

As listed in Eq (11), the $\left\|\bar{Y}-B^{*}\delta^{*}\right\|_2^2$ and $\sum_{j=1}^{N}\left|\delta_j^{*}\right|$ are both the convex function on δ^{*}. Thus, the IEN estimator can be solved through ADMM.

Theorem 4. For the proposed IEN estimator, if $\hat{\varepsilon}_i\hat{\varepsilon}_j>0$, then Eq (13) holds.

$$\left|\hat{\varepsilon}_i-\hat{\varepsilon}_j\right|\le\frac{\|Y\|_2\sqrt{2\left(1-\rho_{ij}\right)}+\frac{\lambda_1}{2}\left|\hat{w}_i-\hat{w}_j\right|}{\lambda_2+\lambda_3\rho_i^{*}}+\left(\|Y\|_2+\frac{\lambda_1}{2}\hat{w}_j\right)H_{i,j}+\lambda_3H_{i,j}\left|\sum_{k=1}^{N-1}\left(\frac{1}{1-\rho_{ik}}+\frac{1}{1+\rho_{ik}}\right)\hat{\varepsilon}_k\right|, \tag{13}$$

where $H_{i,j}=\left|\frac{1}{\lambda_2+\lambda_3\rho_i^{*}}-\frac{1}{\lambda_2+\lambda_3\rho_j^{*}}\right|$.

Proof. Based on Eq (10), it holds that

$$L\left(\hat{\varepsilon},\lambda_1,\lambda_2,\lambda_3\right)=\|Y-B\varepsilon\|_2^2+\lambda_1\sum_{j=1}^{N}w_j\left|\varepsilon_j\right|+\lambda_2\varepsilon^{T}\varepsilon+\lambda_3\sum_{j=1}^{N-1}\sum_{i>j}\left\{\frac{\left(\varepsilon_i-\varepsilon_j\right)^2}{1-\rho_{ij}}+\frac{\left(\varepsilon_i+\varepsilon_j\right)^2}{1+\rho_{ij}}\right\}.$$

As $\hat{\varepsilon}_i\hat{\varepsilon}_j>0$, it is obvious that $\hat{\varepsilon}_i\ne0$ and $\hat{\varepsilon}_j\ne0$, and it can hold that

$$\frac{\partial L\left(\hat{\varepsilon},\lambda_1,\lambda_2,\lambda_3\right)}{\partial\hat{\varepsilon}_i}=-2b_i^{T}\left(Y-B\varepsilon\right)+\lambda_1\hat{w}_i sign\left(\varepsilon_i\right)+2\lambda_2\varepsilon_i+\lambda_3\sum_{k=1}^{N-1}\left(\frac{2\left(\varepsilon_i-\varepsilon_k\right)}{1-\rho_{ik}}+\frac{2\left(\varepsilon_i+\varepsilon_k\right)}{1+\rho_{ik}}\right),$$

$$\frac{\partial L\left(\hat{\varepsilon},\lambda_1,\lambda_2,\lambda_3\right)}{\partial\hat{\varepsilon}_j}=-2b_j^{T}\left(Y-B\varepsilon\right)+\lambda_1\hat{w}_j sign\left(\varepsilon_j\right)+2\lambda_2\varepsilon_j+\lambda_3\sum_{k=1}^{N-1}\left(\frac{2\left(\varepsilon_j-\varepsilon_k\right)}{1-\rho_{jk}}+\frac{2\left(\varepsilon_j+\varepsilon_k\right)}{1+\rho_{jk}}\right).$$

Thus, $\hat{\varepsilon}_i=\dfrac{2b_i^{T}\left(Y-B\varepsilon\right)-\lambda_1\hat{w}_i sign\left(\varepsilon_i\right)+2\lambda_3\sum_{k=1}^{N-1}\left(\frac{1}{1-\rho_{ik}}+\frac{1}{1+\rho_{ik}}\right)\hat{\varepsilon}_k}{2\lambda_2+2\lambda_3\rho_i^{*}}$,

$$\hat{\varepsilon}_j=\frac{2b_j^{T}\left(Y-B\varepsilon\right)-\lambda_1\hat{w}_j sign\left(\varepsilon_j\right)+2\lambda_3\sum_{k=1}^{N-1}\left(\frac{1}{1-\rho_{jk}}+\frac{1}{1+\rho_{jk}}\right)\hat{\varepsilon}_k}{2\lambda_2+2\lambda_3\rho_j^{*}}.$$

Assuming that $D=\left|\dfrac{b_i^{T}\left(Y-B\varepsilon\right)-\frac{\lambda_1}{2}\hat{w}_i sign\left(\varepsilon_i\right)}{\lambda_2+\lambda_3\rho_i^{*}}-\dfrac{b_j^{T}\left(Y-B\varepsilon\right)-\frac{\lambda_1}{2}\hat{w}_j sign\left(\varepsilon_j\right)}{\lambda_2+\lambda_3\rho_j^{*}}\right|$, we have

$$\left|\hat{\varepsilon}_i-\hat{\varepsilon}_j\right|\le D+\lambda_3\left|\sum_{k=1}^{N-1}\left(\frac{\frac{1}{1-\rho_{ik}}+\frac{1}{1+\rho_{ik}}}{\lambda_2+\lambda_3\rho_i^{*}}-\frac{\frac{1}{1-\rho_{jk}}+\frac{1}{1+\rho_{jk}}}{\lambda_2+\lambda_3\rho_j^{*}}\right)\varepsilon_k\right|.$$

Considering that $\left|\frac{a}{b}-\frac{c}{d}\right|\le\left|\frac{a-c}{b}\right|+|c|\left|\frac{1}{b}-\frac{1}{d}\right|$ and $\hat{\varepsilon}_i\hat{\varepsilon}_j>0$, it holds that

$$D \le \left| \frac{\left(b_i - b_j\right)^T \left(Y - B\varepsilon\right) - \frac{\lambda_1}{2}\left(\hat{w}_i sign\left(\varepsilon_i\right) - \hat{w}_j sign\left(\varepsilon_j\right)\right)}{\lambda_2 + \lambda_3 \rho_i^*} \right| + \left| b_j^T \left(Y - B\varepsilon\right) - \frac{\lambda_1}{2} \hat{w}_j sign\left(\varepsilon_j\right) \right| \left| \frac{1}{\lambda_2 + \lambda_3 \rho_i^*} - \frac{1}{\lambda_2 + \lambda_3 \rho_j^*} \right|.$$

Since $L\left(\hat{\varepsilon}, \lambda_1, \lambda_2, \lambda_3\right) \le L\left(\mathbf{0}, \lambda_1, \lambda_2, \lambda_3\right) = \|Y\|_2$, it holds that $\left| b_j^T \left(Y - B\varepsilon\right) - \frac{\lambda_1}{2} \hat{w}_j sign\left(\varepsilon_j\right) \right| \le \|Y\|_2 + \frac{\lambda_1}{2} \hat{w}_j$ and $\left(b_i - b_j\right)^T \left(Y - B\varepsilon\right) \le \left\|b_i - b_j\right\|_2 \left\|Y - B\varepsilon\right\|_2 \le \|Y\|_2 \sqrt{2\left(1 - \rho_{ij}\right)}$.

Then, $D \le \frac{\|Y\|_2 \sqrt{2\left(1 - \rho_{ij}\right)} + \frac{\lambda_1}{2}\left|\hat{w}_i - \hat{w}_j\right|}{\lambda_2 + \lambda_3 \rho_i^*} + \left(\|Y\|_2 + \frac{\lambda_1}{2} \hat{w}_j\right) \left| \frac{1}{\lambda_2 + \lambda_3 \rho_i^*} - \frac{1}{\lambda_2 + \lambda_3 \rho_j^*} \right|$. The proof of Theorem 4 is complete.

Theorem 4 establishes an upper bound for the absolute difference between parameters $\hat{\varepsilon}_i$ and $\hat{\varepsilon}_j$ through the IEN estimator. When the correlation coefficient $\rho_{ij} \to 0$, this upper bound converges to zero, demonstrating that the IEN estimator possesses the group effect property. This property directly addresses the previously noted limitation of the standard elastic net, which lacks an adaptive grouping mechanism. By ensuring that highly correlated predictors receive similar coefficient estimates, the IEN estimator supports more stable variable selection and parameter estimation. Within the context of grey system modeling, this characteristic contributes in several ways: It mitigates variable selection bias in high-dimensional settings, promotes consistency in parameter estimation for short-term forecasting, and improves the interpretability of models when applied to data with complex variable relationships.

Theorem 5. Let $\mathrm{A} = \left\{ j : \varepsilon_j \ne 0 \right\} = \left\{1, 2, \cdots, k\right\}$, $k < N$, assuming there exists a sequence $\left\{a_n\right\}$, where $\left\{a_n\right\}$ satisfies $a_n^{\gamma_2} \cdot \hat{\varepsilon}_j^* = O_N(1)$, $\lambda_1 = o\left(\sqrt{n}\right)$, and $\frac{\lambda_1 \cdot a_n^{\gamma_2}}{\sqrt{n}} \to \infty$. Then, for the $\hat{\varepsilon}$, it satisfies the following properties:

(1) Asymptotic normality, i.e., $\sqrt{n}\left(\hat{\varepsilon}_A - \varepsilon_A\right) \xrightarrow{d} N\left(0, I_1^{-1}\right)$, where $I\left(\varepsilon_0\right)$ is the Fisher information matrix, i.e., $I\left(\varepsilon_0\right) = \begin{pmatrix} I_1 & I_3 \\ I_2 & I_4 \end{pmatrix}$, and I_1 is the k-order square matrix.

(2) Consistency of variable selection, i.e., $\left\{ j : \hat{\varepsilon}_j \ne 0 \right\} \xrightarrow{P} \mathrm{A}$.

Proof. (1) Assuming that $\tilde{\varepsilon} = \varepsilon_0 + \frac{u}{\sqrt{n}}$ and

$$\phi_n\left(u\right) = \left\| Y - B\left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right) \right\|_2^2 + \lambda_1 \sum_{j=1}^{N} w_j \left| \varepsilon_{0j} + \frac{u_j}{\sqrt{n}} \right| + \lambda_2 \left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right)^2 + \lambda_3 Pc\left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right),$$

it holds that

$$\phi_n\left(u\right) - \phi_n\left(0\right) = A_1 + A_2 + A_3 + A_4,$$

where $A_1 = \left\| Y - B\left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right) \right\|_2^2 - \left\| Y - B\varepsilon_0 \right\|_2^2$, $A_2 = \lambda_1 \sum_{j=1}^{N} w_j \left| \varepsilon_{0j} + \frac{u_j}{\sqrt{n}} \right| - \lambda_1 \sum_{j=1}^{N} w_j \left| \varepsilon_{0j} \right|$, $A_3 = \lambda_2 \left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right)^2 - \lambda_2 \varepsilon_0^{\ 2}$, and $A_4 = \lambda_3 Pc\left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right) - \lambda_3 Pc\left(\varepsilon_0\right)$.

Assuming that $L\left(\varepsilon_0\right) = \left\| Y - B\varepsilon_0 \right\|_2^2$, based on Taylor's formula, it holds that

$$A_1 = L\left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right) - L(\varepsilon_0) = \frac{L'(\varepsilon_0)}{1!}\cdot\frac{u}{\sqrt{n}} + \frac{L''(\varepsilon_0)}{2!}\cdot\left(\frac{u}{\sqrt{n}}\right)^2 + \frac{L'''(\varepsilon_0)}{3!}\cdot\left(\frac{u}{\sqrt{n}}\right)^3.$$

Considering that $\frac{\partial L(\varepsilon_0)}{\partial \varepsilon_0} = -2B^T \cdot (Y - B\varepsilon_0)$, $\frac{\partial^2 L(\varepsilon_0)}{\partial {\varepsilon_0}^2} = 2B^T B$, A_1 follows

$$A_1 = -2B^T \cdot (Y - B\varepsilon_0)\cdot\frac{u}{\sqrt{n}} + B^T B\cdot\left(\frac{u}{\sqrt{n}}\right)^2.$$

Assuming that $T_1 = -2B^T \cdot (Y - B\varepsilon_0)\cdot\frac{u}{\sqrt{n}}$ and $T_2 = B^T B\cdot\left(\frac{u}{\sqrt{n}}\right)^2$, then $A_1 = T_1 + T_2$. As $Y = B\varepsilon_0$, then $EY = B\varepsilon_0$, so it holds that

$$E\left(2B^T \cdot (Y - B\varepsilon_0)\cdot u\right) = 2B^T \cdot E(Y - B\varepsilon_0)\cdot u = 0,$$

$$D\left(B^T \cdot (Y - B\varepsilon_0)\cdot u\right) = u^T E\left(B^T B\right) u = -\frac{1}{2}u^T I(\varepsilon_0) u.$$

According to the central limit theorem, it obtains $T_1 \xrightarrow{d} N\left(0, -u^T I(\varepsilon_0) u\right)$. Based on the law of large numbers, it holds that $\frac{B^T B}{n} \xrightarrow{p} \frac{1}{2} I(\varepsilon_0)$. Thus, $T_2 \xrightarrow{d} \frac{1}{2} u^T I(\varepsilon_0) u$.

For A_2, it holds that $A_2 = \frac{\lambda_1}{\sqrt{n}}\sum_{j=1}^{N} w_j \sqrt{n}\left(\left|\varepsilon_{0j} + \frac{u_j}{\sqrt{n}}\right| - \left|\varepsilon_{0j}\right|\right)$, where $w_j = \frac{\gamma_1 \cdot j}{\left|\hat{\varepsilon}_j^*\right|^{\gamma_2}}$.

Situation 1. If $\varepsilon_{0j} = 0,\ u_j = 0$, and $\sqrt{n}\left(\left|\varepsilon_{0j} + \frac{u_j}{\sqrt{n}}\right| - \left|\varepsilon_{0j}\right|\right) = 0$, then $A_2 = 0$.

Situation 2. If $\varepsilon_{0j} = 0,\ u_j \neq 0$, $\sqrt{n}\left(\left|\varepsilon_{0j} + \frac{u_j}{\sqrt{n}}\right| - \left|\varepsilon_{0j}\right|\right) = \left|u_j\right|$, and $A_2 = \frac{\lambda_1}{\sqrt{n}}\sum_{j=1}^{N} w_j \left|u_j\right|$, then for $n \to \infty$, it holds that $\frac{\lambda_1}{\sqrt{n}}\frac{\gamma_1 \cdot j}{\left|\hat{\varepsilon}_j^*\right|^{\gamma_2}}\left|u_j\right| = \frac{\lambda_1 \cdot a_n^{\gamma_2}}{\sqrt{n}}\frac{\gamma_1 \cdot j}{\left|a_n \cdot \hat{\varepsilon}_j^*\right|^{\gamma_2}}\left|u_j\right| \to \infty$.

Situation 3. If $\varepsilon_{0j} \neq 0,\ u_j = 0$, and $\sqrt{n}\left(\left|\varepsilon_{0j} + \frac{u_j}{\sqrt{n}}\right| - \left|\varepsilon_{0j}\right|\right) = 0$, then $A_2 = 0$.

Situation 4. If $\varepsilon_{0j} \neq 0,\ u_j \neq 0$, for $n \to \infty$ and $\sqrt{n}\left(\left|\varepsilon_{0j} + \frac{u_j}{\sqrt{n}}\right| - \left|\varepsilon_{0j}\right|\right) \to u_j sign(\varepsilon_{0j})$, then $A_2 = \frac{\lambda_1}{\sqrt{n}}\sum_{j=1}^{N} w_j \left|u_j\right| \to 0$.

Combine the above four situations, and it holds that $A_2 \xrightarrow{p} \begin{cases} 0, & \varepsilon_{0j} \in A^c \text{ and } u_j \neq 0, \\ \infty, & \text{otherwise.} \end{cases}$

For A_3, as $A_3 = \lambda_2\left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right)^2 - \lambda_2 {\varepsilon_0}^2$, it holds that $A_3 = \lambda_2\left(2\varepsilon_0\frac{u}{\sqrt{n}} + \frac{u^T u}{n}\right)$. Thus, $A_3 \to 0$ if $n \to \infty$.

For A_4, it holds that

$$A_4 = \lambda_3 Pc\left(\varepsilon_0 + \frac{u}{\sqrt{n}}\right) - \lambda_3 Pc(\varepsilon_0) = \lambda_3\left(\varepsilon^T + \frac{u^T}{\sqrt{n}}\right) M\left(\varepsilon + \frac{u}{\sqrt{n}}\right) - \lambda_3 \varepsilon^T M \varepsilon = \frac{\lambda_3}{n} u^T M u + 2\frac{\lambda_3}{\sqrt{n}}\varepsilon^T M u.$$

Thus, $A_4 \to 0$ if $n \to \infty$.

Then, when $n \to \infty$, it holds that $\phi_n(u) - \phi_n(0) \xrightarrow{d} \begin{cases} -u^T N_{I_1} + \frac{1}{2} u^T I_1 u, & \varepsilon_{0j} \in A^c \text{ and } u_j \neq 0, \\ \infty, & \text{otherwise,} \end{cases}$ where $N_{I_1} = N(0, I_1)$.

Assuming that $\Gamma_n(u) = -u^T N_{I_1} + \frac{1}{2} u^T I_1 u$, then $\Gamma_n(u)$ is the convex function on u. Thus, $\Gamma_n(u)$ has a unique minimum point $u = I_1^{-1} N_{I_1}$. When $\varepsilon_{0j} \in A^c$, it holds that $u \xrightarrow{d} 0$. When $\varepsilon_{0j} \in A$, it holds that $u \xrightarrow{d} I_1^{-1} N_{I_1}$. Thus, $\sqrt{n}(\hat{\varepsilon}_A - \varepsilon_A) \xrightarrow{d} N(0, I_1^{-1})$.

(2) For $H(\varepsilon) = \|Y - B\varepsilon\|_2^2 + \lambda_1 \sum_{j=1}^{N} w_j |\varepsilon_j| + \lambda_2 \varepsilon^2 + \lambda_3 \varepsilon^T M \varepsilon$, based on the stability of Karush–Kuhn–Tucker conditions, it holds that if $\hat{\varepsilon}_j \neq 0$, then $\frac{\partial H(\varepsilon)}{\partial \varepsilon} = 0$. Namely,

$$2B^T(Y - B\varepsilon) + \lambda w_j sign(\varepsilon_j) + 2\lambda_2 \varepsilon + 2\lambda_3 \varepsilon^T M = 0,$$

$$p(\hat{\varepsilon}_j \neq 0) = p\left(2B^T(Y - B\varepsilon) + \lambda w_j sign(\varepsilon_j) + 2\lambda_2 \varepsilon + 2\lambda_3 \varepsilon^T M = 0\right).$$

Thus,

$$\frac{2}{\sqrt{n}} B^T(Y - B\varepsilon) + \frac{\lambda_1}{\sqrt{n}} w_j sign(\varepsilon_j) + \frac{2}{\sqrt{n}} \lambda_2 \varepsilon + \frac{2}{\sqrt{n}} \lambda_3 \varepsilon^T M = 0.$$

For $n \to \infty$, it holds that $\frac{\lambda_1}{\sqrt{n}} w_j sign(\varepsilon_j) = \frac{\lambda_1 \cdot a_n^{\gamma_2}}{\sqrt{n}} \frac{\gamma_1 \cdot j}{|a_n \cdot \hat{\varepsilon}_j^*|^{\gamma_2}} sign(\varepsilon_j) \xrightarrow{p} \infty$, $\frac{2}{\sqrt{n}} \lambda_2 \varepsilon \xrightarrow{p} 0$, and $\frac{2}{\sqrt{n}} \lambda_3 \varepsilon^T M \xrightarrow{p} 0$.

Based on Taylor's formula, $\frac{2}{\sqrt{n}} B^T(Y - B\varepsilon) = \frac{2}{\sqrt{n}} B^T(Y - B\varepsilon_0) + \frac{1}{\sqrt{n}} B^T B(\varepsilon - \varepsilon_0)$.

Based on (1), we hold that $\frac{2}{\sqrt{n}} B^T(Y - B\varepsilon_0) \xrightarrow{d} N(0, I(\varepsilon_0))$ and $\frac{1}{\sqrt{n}} B^T B \xrightarrow{d} I(\varepsilon_0)$. Then, for $n \to \infty$, we hold that $p(\hat{\varepsilon}_j \neq 0) = 0$ and $p(\hat{\varepsilon}_{A^c} = 0) = 1$. The proof is complete.

Theorem 5 establishes that the IEN estimator satisfies the oracle property. This result directly addresses the fundamental limitation identified in the standard elastic net framework, which lacks this property. Specifically, Theorem 5 ensures that the IEN estimator achieves consistent variable selection by asymptotically identifying the correct subset of informative predictors and effectively eliminating noninformative variables. Concurrently, it yields coefficient estimates that converge asymptotically to their true values.

For grey system modeling, the fulfillment of the oracle property by the IEN estimator offers distinct contributions. First, it resolves the issue of inconsistent variable selection, thereby enhancing the accuracy and reliability of models when simulating real data structures. Second, it improves the asymptotic efficiency of parameter estimation in high-dimensional settings, which strengthens the interpretative foundation of the resulting models. When considered together with the group effect property established in Theorem 4, the IEN estimator provides a more robust statistical foundation for grey system analysis. This combination supports more dependable model identification and forecasting, particularly in scenarios involving complex, high-dimensional data.

3.1.3. The selection of regularization parameters

The selection of regularization parameters is critical to the performance of the estimator. As established in Theorem 3, the proposed IEN estimator can be reformulated into a standard lasso problem, enabling the use of established lasso parameter selection methods. Among these, the discrepancy principle is widely adopted for parameter calibration. In cases where prior information on residuals is unavailable, the compensated discrepancy principle offers a feasible solution. Therefore, this method is employed in our study to determine the regularization parameter. Assuming the regularization parameter set is $\Lambda=\{\gamma_1,\gamma_2,\lambda_1,\lambda_2,\lambda_3\}$, the evaluation function of discrepancy principle is given in Eq (14).

$$\nu(\Lambda)=\frac{Rss(\Lambda)}{N_h-\rho(\Lambda)}, \tag{14}$$

where $Rss(\Lambda)$ means the residual sum of squares on regularization parameter set Λ, N_h is sample size, $\rho(k)=trace\left(B\left(B^TB+kL^-\right)^{-1}B^T\right)$, L^- is the generalized inverse matrix of L, $L=diag\left\{\left|\varepsilon_j^*\right|\right\}$, and ε^* is obtained through the elastic net estimator.

3.2. Nonlinear parameters identification

The GM(1,N) power model also contains a power index set. This parameter set includes $N-1$ nonlinear parameters, which have been treated as exogenous information in the IEN estimator. Namely, the estimation of the linear parameters can be regarded as implicit functions of these nonlinear parameters. This section focuses on identifying the optimal values of such nonlinear parameters. To enhance both model performance and prediction accuracy, the MAPE of the model results is adopted as the optimization objective. The corresponding formulation is presented in Eq (15), and solving this optimization model yields the final estimates for the nonlinear parameters.

$$\underset{\{r_2,r_3,\cdots,r_N\}}{\text{Min}}\ \text{MAPE}=\frac{1}{n}\sum_{k=1}^{n}\frac{\left|\hat{x}_1^{(0)}(k)-x_1^{(0)}(k)\right|}{x_1^{(0)}(k)}\times 100\%. \tag{15}$$

$$s.t.\begin{cases}\min\limits_{\Lambda}\nu(\Lambda)=\dfrac{Rss(\Lambda)}{N_h-\rho(\Lambda)}\\ \Lambda=\{\gamma_1,\gamma_2,\lambda_1,\lambda_2,\lambda_3\}\\ \lambda_1>0,\lambda_2>0,\lambda_3>0,\gamma_1>0,\gamma_2>0\\ \hat{\varepsilon}^*=\arg\min\left\{\left\|\bar{Y}-B^*\delta^*\right\|_2^2+k_\lambda\sum\limits_{j=1}^{N}\left|\delta_j^*\right|\right\}\\ a=\hat{\varepsilon}_1^*,\\ b_i=\hat{\varepsilon}_i^*,\quad i=2,3,\cdots N\\ \hat{x}_1^{(0)}(1)=\hat{x}_1^{(1)}(1)\\ \hat{x}_1^{(0)}(k)=\left[x_1^{(0)}(0)-\dfrac{1}{a}\sum\limits_{i=2}^{N}b_i\left(x_i^{(1)}(k-1)\right)^{r_i}\right]e^{-a(k-1)}+\dfrac{1}{a}\sum\limits_{i=2}^{N}b_i\left(x_i^{(1)}(k-1)\right)^{r_i},\quad k=2,3,\cdots,n.\end{cases}$$

The optimization model in Eq (15) aims to minimize the MAPE but involves implicit functions that preclude an analytical solution. Therefore, heuristics algorithm methods such GA, PSO, and WOA [30]) are employed to solve this problem numerically. In this paper, the GA is adopted to identify these nonlinear parameters, with a population size of 100, 500 iterations, a crossover rate of 0.8, and a mutation rate of 0.01. Thus, the overall modelling process of the GM(1,N) power model is summarized as follows:

Step 1. The original datasets, $X^{(0)}=\left\{X_1^{(0)},X_2^{(0)},\cdots,X_N^{(0)}\right\}$, are divided into training and testing sets. The training set is used for parameter estimation and model specification, while the testing set serves for model validation.

Step 2. 1-AGO is applied to the training sequences to construct accumulated generating operation sequences, $X_i^{(1)}=\left\{x_i^{(1)}(1),x_i^{(1)}(2),\cdots,x_i^{(1)}(n)\right\}$ and $i=1,2,\cdots,N$.

Step 3. The nonlinear parameters are determined by solving the optimization model in Eq (13) using heuristic algorithms.

Step 4. The linear parameters of the GM(1,N) power model are estimated using the proposed IEN estimator.

Step 5. These estimated parameters are substituted into the GM(1,N) power model to generate predictions results for $X_1^{(1)}$, i.e., $\hat{X}_1^{(1)}=\left\{\hat{x}_1^{(1)}(1),\hat{x}_1^{(1)}(2),\cdots,\hat{x}_1^{(1)}(n)\right\}$. These results are then transformed back to the original observation set, $\hat{X}_1^{(0)}=\left\{\hat{x}_1^{(0)}(1),\hat{x}_1^{(0)}(2),\cdots,\hat{x}_1^{(0)}(n)\right\}$. Finally, the prediction accuracy is evaluated against the actual observations.

4. Application: Air quality prediction

4.1. Data description

This study selects Beijing and Shanghai as case studies. Both cities are among the largest and most economically significant in China, yet they have followed different trajectories in energy consumption, industrial structure, and urban development. These systemic differences are likely reflected in distinct chemical profiles of air pollution. A comparative analysis of these two major urban centers therefore supports the development of a more robust and generalizable model for assessing air quality dynamics in megacities. The dataset utilized in this analysis provides a comprehensive long-term record, covering the 23-year period from 2000 to 2022. It includes annual average concentrations of $PM_{2.5}$, which is the key dependent variable for this research. The dataset also contains annual averages for its major chemical components, which serve as the independent variables. These components are sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+), organic matter (OM), and black carbon (BC). These five substances are the primary constituents of $PM_{2.5}$, and their relative proportions are critical for identifying specific pollution sources and atmospheric processes.

To construct and validate a predictive model, the dataset is divided into two distinct parts. The data from the first 18 years, spanning from 2000 to 2017, are designated as the training set. The model will be developed and its parameters will be calibrated using this portion of the data. The remaining five years of data, from 2018 to 2022, are reserved as the testing set. This setup allows for a rigorous evaluation of the model's predictive power on unseen, more recent data, ensuring that its performance can be reliably assessed for future applications.

To comprehensively evaluate the effectiveness of these forecasting models, four metrics are adopted in this study: MAPE, RMSE, STD, and R ? The corresponding formulations are provided in Eqs (16)–(19).

$$\text{MAPE} = \frac{1}{n}\sum_{k=1}^{n}\frac{\left|\hat{x}_1^{0}\ \ k\ -x_1^{0}\ \ k\right|}{x_1^{0}\ \ k}\times 100\%\ , \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\hat{x}_1^{(0)}(k)-x_1^{(0)}(k)\right)^2}\ , \tag{17}$$

$$\text{STD} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\frac{\left|\hat{x}_1^{(0)}(k)-x_1^{(0)}(k)\right|}{x_1^{(0)}(k)}-\text{MAPE}\right)^2}\ , \tag{18}$$

$$R^2 = 1-\frac{\sum_{k=1}^{n}\left(\hat{x}_1^{(0)}(k)-x_1^{(0)}(k)\right)^2}{\sum_{k=1}^{n}\left(\hat{x}_1^{(0)}(k)-\overline{x}_1^{(0)}(k)\right)^2}\ , \tag{19}$$

where $\overline{x}_1^{(0)} = \frac{1}{n}\sum_{k=1}^{n}x_1^{(0)}(k)$.

4.2. Comparison of results

This study employs the GM(1,N) power model as the foundational framework for forecasting air quality. To validate the advantages of the proposed IEN algorithm, this method is specifically applied to handle the parameter identification of the GM(1,N) power model. The complete model utilizing the proposed algorithm is designated as IEN. For a comprehensive comparison, this research also identifies the parameters of the same GM(1,N) power model using four existing estimators: OLS, ridge estimation, lasso estimation, and the classical elastic net algorithm. The corresponding models are labeled as OLS, Ridge, Lasso, and EN, respectively. The key metrics of all five models are illustrated in Figures 1 and 2 and summarized in Table 1.

The GM(1,N) power model demonstrates considerable potential in forecasting annual average $PM_{2.5}$ concentrations. As observed in the upper-left subplots of Figures 1 and 2, the trends generated by all five models align closely with the actual $PM_{2.5}$ values, which are represented by yellow-green bars. Given that these models share the same GM(1,N) power structure and differ only in parameter estimation methods, this consistency underscores the robustness of the model framework. Furthermore, residual analysis supports this conclusion: The scatter plots in the lower portions of both Figures 1 and 2 show that most points are distributed along the diagonal line in the first quadrant, indicating accurate trend capture by all five models.

The choice of the parameter estimation method significantly influences the predictive performance of the GM(1,N) power model, despite the overall trend consistency observed across all variants. As illustrated in the upper-left subplots of Figures 1 and 2, noticeable deviations between the modeled and actual $PM_{2.5}$ values are identified. Specifically, the OLS-based model exhibits considerable discrepancies in both the fitting and forecasting phases. Given that the OLS is widely used in grey system models, its suboptimal performance here highlights a clear need for improvement. In contrast, the Ridge-based model shows pronounced fluctuations during the fitting phase, yet captures the forecasting trend more accurately, suggesting its underlying potential. The IEN, EN, and Lasso models consistently align closely with the actual $PM_{2.5}$ values, indicating their superior suitability for this modeling framework.

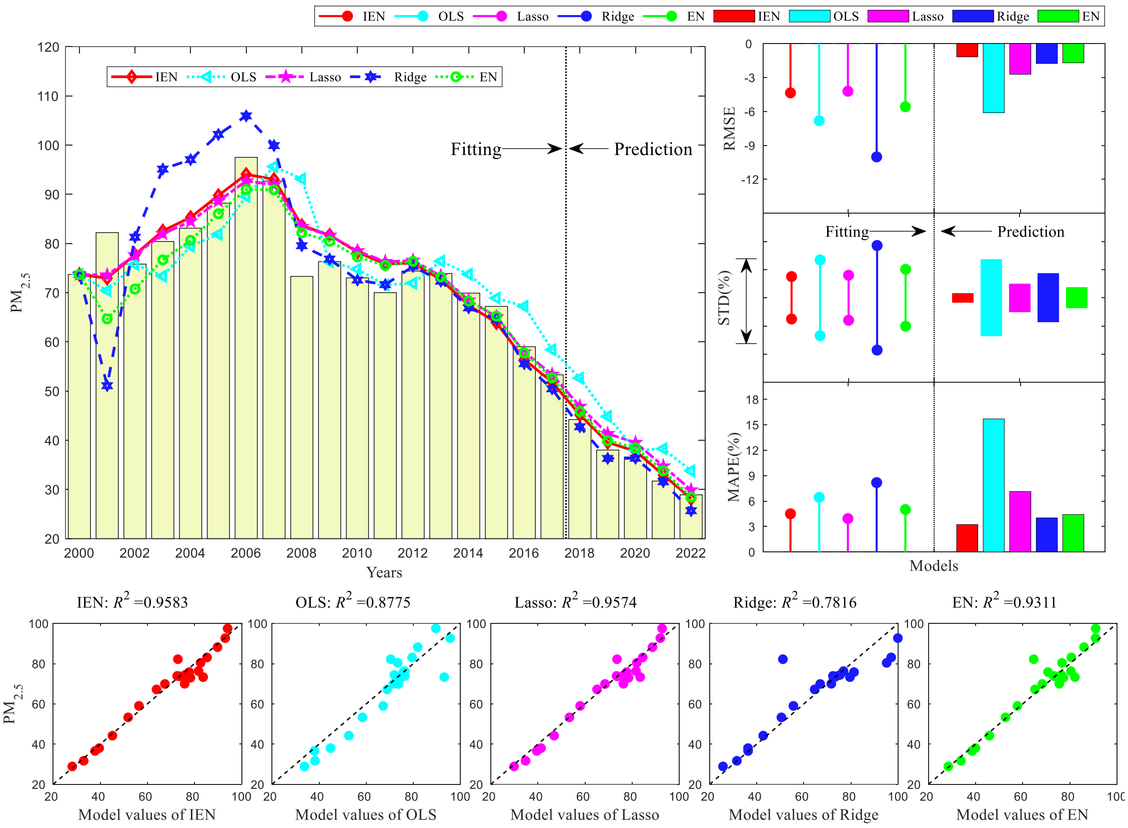


Figure 1. The performance of the five models for predicting air quality in Beijing.

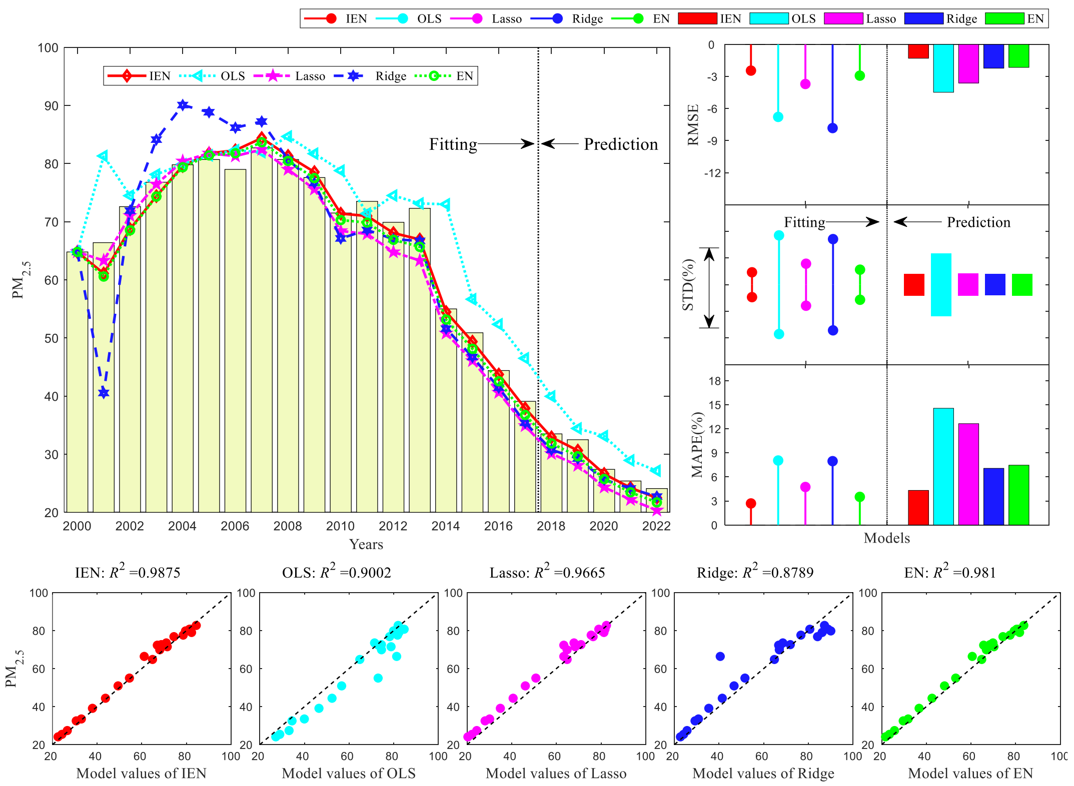


Figure 2. The performance of the five models for predicting air quality in Shanghai.

The proposed IEN estimator achieves superior predictive accuracy compared to other estimators in this study. As shown in the subplots of the RMSE and MAPE in Figures 1 and 2, the stems and bars corresponding to the IEN model (marked in red) are consistently the shortest, indicating that this model maintains the lowest error levels in both fitting and forecasting phases. This advantage is further supported by the numerical results in Table 1. On the Shanghai dataset, the IEN model consistently achieves the lowest MAPE and MSE values in both phases. On the Beijing dataset, it also attains the best or second-best performance in terms of both metrics. These consistent results across datasets and evaluation phases confirm the prediction accuracy of the GM(1,N) power model using the IEN estimator.

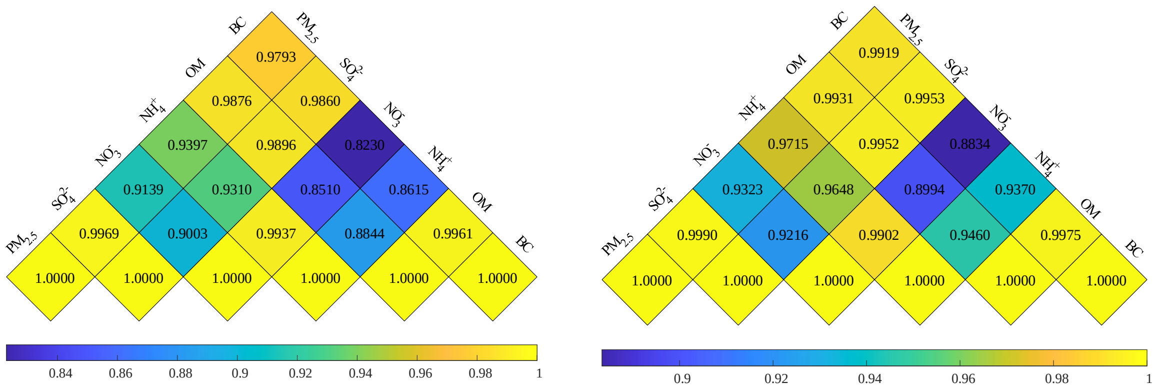
The proposed IEN estimator significantly enhances the robustness of the GM(1,N) power model. In this study, the STD is employed to assess model robustness, reflecting the fluctuation of forecasting accuracy. The results in Table 1 show that the IEN-based model achieves the lowest STD values on both datasets, indicating superior stability in its outputs. Furthermore, the STD subplots in Figures 1 and 2 consistently display the shortest red stems and bars, representing the IEN model. This visual evidence confirms the ability of the IEN estimator to improve the robustness of the GM(1,N) power model.

Table 1. Metrics of the five models for fitting and prediction.

Models / Metrics	Beijing					Shanghai				
	IEN	OLS	Lasso	Ridge	EN	IEN	OLS	Lasso	Ridge	EN
$MAPE_{fit}$	<u>4.50%</u>	6.44%	**<u>3.94%</u>**	8.18%	5.01%	**<u>2.69%</u>**	8.05%	4.74%	7.96%	<u>3.51%</u>
$MAPE_{pre}$	**<u>3.22%</u>**	15.69%	7.13%	<u>4.02%</u>	4.41%	**<u>4.32%</u>**	14.56%	12.64%	<u>7.07%</u>	7.46%
$RMSE_{fit}$	<u>4.357</u>	6.815	**<u>4.211</u>**	10.021	5.582	**<u>2.448</u>**	6.781	3.700	7.824	<u>2.922</u>
$RMSE_{pre}$	**<u>1.174</u>**	6.111	2.712	1.763	<u>1.700</u>	**<u>1.294</u>**	4.475	3.607	2.220	<u>2.148</u>
STD_{fit}	**<u>3.66%</u>**	6.54%	<u>3.89%</u>	9.02%	4.91%	**<u>2.27%</u>**	9.00%	3.83%	8.32%	<u>2.75%</u>
STD_{pre}	3.46%	13.46%	3.54%	<u>2.72%</u>	**<u>2.46%</u>**	**<u>2.83%</u>**	9.85%	6.69%	<u>2.98%</u>	4.61%
R^2	**<u>0.960</u>**	0.841	<u>0.954</u>	0.851	0.923	**<u>0.988</u>**	0.897	0.970	0.907	<u>0.982</u>

*Note: In this table, the metrics with the subscript *fit* represent the model performance on the fitting dataset, while those with *pre* represent the performance on the prediction dataset. To enhance readability, the best-performing value in each comparative experiment is highlighted in both bold and underlined, and the second-best value is only underlined.

The IEN estimator effectively addresses the ill-conditioned nature of the classical GM(1,N) power model. Conventional models utilizing the OLS estimation fail to handle multicollinearity among predictor variables. As revealed in the correlation analysis of the Beijing and Shanghai datasets (Figure 3), most variable pairs exhibit correlation coefficients exceeding 0.85, indicating severe multicollinearity. This high intervariable correlation results in a rank-deficient matrix $B^T B$ in Eq (4), creating an ill-conditioned system in the standard modeling approach. By incorporating both L_1 and L_2 regularization terms along with a CP, the IEN estimator successfully mitigates this pathological matrix condition and enhances numerical stability.



(a) Multicollinearity analysis in Beijing. (b) Multicollinearity analysis in Shanghai.

Figure 3. Multicollinearity analysis of the GM(1,N) power model in air quality prediction.

The observed performance advantages of the IEN estimator can be critically understood through the theoretical properties established earlier. Its superior and stable predictive accuracy, as evidenced by the lowest error metrics and STD values, is a direct empirical manifestation of its integrated oracle and group effect properties. The oracle property ensures the consistent selection of the most relevant variables during model identification, which directly enhances the interpretative reliability of the GM(1,N) power model and its fidelity in simulating real data structures. Concurrently, the adaptive group effect promotes the assignment of similar coefficients to highly correlated predictors. This mechanism mitigates estimation variance and selection bias induced by multicollinearity, a common challenge in complex grey systems, thereby fundamentally underpinning the model's robustness and output stability. Compared to single-penalty or nonadaptive regularization techniques, the IEN framework provides a more balanced solution by simultaneously addressing accurate variable selection and stable coefficient estimation. This synergy is particularly beneficial for grey system modeling in high-dimensional settings, as it improves generalization capability and reduces the decision-making risk associated with model-based forecasts.

In summary, the IEN estimator is recommended for constructing GM(1,N) power models due to its comprehensive advantages. It effectively resolves the ill-conditioned problem caused by multicollinearity in multivariate datasets through integrated regularization mechanisms. This capability directly contributes to enhanced prediction accuracy across both fitting and forecasting phases. Furthermore, improved robustness is consistently demonstrated through reduced variability in model errors. The ability of the IEN estimator to maintain numerical stability while delivering superior performance makes it a preferable alternative to traditional parameter estimation approaches for grey multivariate modeling under complex variable relationships.

5. Conclusions

The GM(1,N) power model is widely employed as a grey multivariate forecasting technique. However, its conventional parameter estimation that relies on OLS often leads to ill-conditioned solutions, particularly under multicollinearity. To address this limitation, an IEN estimator integrating adaptive lasso and CP is proposed in this paper. The main findings are summarized as follows:

First, the proposed IEN estimator fundamentally addresses the ill-conditioned problem inherent in conventional models that rely on ordinary least squares. By integrating both L_1 and L_2 regularization with a correlation-aware penalty, the method effectively mitigates the adverse effects of multicollinearity, which is frequently encountered in real-world data. This structural enhancement ensures numerical stability and reliability during parameter identification.

Second, from a theoretical perspective, the IEN estimator possesses two key properties that underpin its robustness and effectiveness. As established in Theorem 4, the estimator exhibits an adaptive group effect, which allows it to automatically identify and group highly correlated explanatory variables, thereby enhancing model interpretability and stability. Furthermore, Theorem 5 confirms that the IEN estimator satisfies the oracle property, ensuring that it asymptotically selects the correct subset of predictors and provides consistent parameter estimation. These theoretical guarantees distinguish the IEN framework from conventional regularization methods and form the foundation for its improved performance.

Third, from an applied perspective, the superiority of the IEN estimator in enhancing the parameter estimation for the GM(1,N) power model is demonstrated. Its application to $PM_{2.5}$ concentration forecasting in Beijing and Shanghai empirically validates the theoretical advantages. This IEN-based model consistently achieves superior predictive accuracy and robustness, as evidenced by leading performance in key metrics, including the RMSE, MAPE, and STD, across both fitting and forecasting phases. This empirical success directly stems from the estimator's ability to manage high variable correlations and maintain numerical stability in real-world, collinear data environments. Therefore, the IEN estimator is established as a robust and effective parameter estimation alternative for grey forecasting models, particularly when confronting complex and ill-conditioned data structures.

Despite the promising results, several limitations of this study should be addressed. First, the proposed IEN algorithm primarily focuses on estimating linear parameters, while nonlinear parameters are still solved using conventional heuristic methods. Furthermore, the applicability of the proposed parameter estimation approach has only been validated for grey multivariate forecasting models; its effectiveness for other generalized forecasting architectures, including deep learning and artificial intelligence models, has not been explored. Future research will aim to develop a nonlinear elastic net formulation for the simultaneous estimation of all model parameters. Additionally, the potential extension and adaptation of the proposed estimator to other advanced modeling frameworks should be investigated.

Author contributions

Qinzi Xiao: Data curation, writing–original draft preparation; Mingyun Gao: Software, visualization, formal analysis, validation, writing–review & editing; Congjun Rao: writing–review & editing, supervision. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was supported by grants from the National Natural Science Foundation of China (No.

72301202, 72401104), the Humanities and Social Science Fund of the Ministry of Education (No. 24YJCZH060), the Natural Science Foundation of Hubei Province of China (No. 2023AFB491), the Philosophy and Social Science Research Project of the Hubei Provincial Department of Education (No. 22Q067), and the Science Fund of Wuhan Institute of Technology (No. 23QD22).

Conflict of interest

The authors confirm no conflicts of interest.

References

1. Q. Xiao, M. Gao, R. Wu, How can new quality productivity alleviate rural energy poverty? Evidence from dynamic lag grey relational analysis, *Energ. Source. Part B*, **21** (2026), 2599195. https://doi.org/10.1080/ 15567249.2025.2599195
2. W. Yang, L. Lin, H. Gao, Simulation evaluation of small samples based on grey estimation and improved bootstrap, *Grey Syst.*, **12** (2022), 376–388. https://doi.org/10.1108/GS-09-2020-0121
3. B. Li, S. Zhang, W. Li, Y. Zhang, Application progress of grey model technology in agricultural science, *Grey Syst.*, **12** (2022), 744–784, https://doi.org/10.1108/GS-05-2022-0045
4. Y. Lv, M. Gao, X. Xiao, Unbiased forecasting of seasonal wind power generation based on a novel seasonal multivariable grey model, *Renew. Energ.*, **258** (2026), 124952. https://doi.org/10.1016/j.renene.2025.124952
5. R. M. K. T. Rathnayaka, D. Seneviratna, Predicting of aging population density by a hybrid grey exponential smoothing model (HGESM): A case study from Sri Lanka, *Grey Syst.*, **14** (2024), 601–617. https://doi.org/10.1108/GS-01-2024-0002
6. Q. Xiao, M. Gao, L. Chen, J. Jiang, Dynamic multi-attribute evaluation of digital economy development in China: A perspective from interaction effect, *Technol. Econ. Dev. Eco.*, **29** (2023), 1728–1752. https://doi.org/10.3846/tede.2023.20258
7. Q. Xiao, M. Gao, L. Chen, M. Goh, Multi-variety and small-batch production quality forecasting by novel data-driven grey Weibull model, *Eng. Appl. Artif. Intel.*, **125** (2023), 106725. https://doi.org/10.1016/ j.engappai.2023.106725
8. Z. Wang, Grey multivariable power model GM(1,N) and its application, *Syst. Eng. Theory Pract.*, **34** (2014), 2357–2363.
9. K. Xu, X. Bao, Q. Wang, Railway passenger volume forecasting based on GA-GM(1,N,α) power mode, *Railway Standard Design*, **62** (2018), 6–10.
10. S. Ding, Y. G. Dang, N. Xu, J. J. Wang, S. S. Geng, Construction and optimization of a multi-variables discrete grey power model, *Syst. Eng. Electrons.*, **40** (2018), 1302–1309.
11. Q. Xiao, M. Gao, L. Chen, M. Goh, Small-batch product quality prediction using a novel discrete Choquet fuzzy grey model with complex interaction information, *Inform. Sciences*, **678** (2024), 120997. https://doi.org/10.1016/j.ins.2024.120997
12. H. Zhu, C. Liu, W. Z. Wu, W. L. Xie, T. Lao, Weakened fractional-order accumulation operator for ill-conditioned discrete grey system models, *Appl. Math. Model.*, **111** (2022), 349–362. https://doi.org/10.1016/j. apm.2022.06.042
13. B. Wei, N. M. Xie, Parameter estimation for grey system models: A nonlinear least squares perspective, *Commun. Nonlinear Sci.*, **95** (2021), 105653. https://doi.org/10.1016/j.cnsns.2020.105653

14. Y. Hirose, Regularization methods based on the Lq-likelihood for linear models with heavy-tailed errors, *Entropy*, **22** (2020), 1036. https://doi.org/10.3390/e22091036
15. H. Zou, The adaptive lasso and its Oracle properties, *J. Am. Stat. Assoc.*, **101** (2006), 1418–1429. https://doi.org/10.1198/016214506000000735
16. N. Meinshausen, B. Yu, Lasso-type recovery of sparse representations for high-dimensional data, *Ann. Stat.*, **37** (2009), 246–270. https://doi.org/10.1214/07-AOS582
17. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, *J. R. Stat. Soc. B*, **67** (2005), 91–108. https://doi.org/10.1111/j.1467-9868.2005.00490.x
18. J. Fan, Y. Fan, E. Barut, Adaptive robust variable selection, *Ann. Stat.*, **42** (2014), 324. https://doi.org/10.1214/13-AOS1191
19. H. Park, F. Sakaori, Lag weighted lasso for time series model, *Computation Stat*., **28** (2013), 493–504. https://doi.org/10.1007/s00180-012-0313-5
20. H. Zou, T. Hastie, Regularization and variable selection via the Elastic Net, *J. R. Stat. Soc. B*, **67** (2005), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x
21. S. Ghosh, On the grouped selection and model complexity of the adaptive Elastic Net, *Stat. Comput*., **21** (2011), 451–462. https://doi.org/10.1007/s11222-010-9181-4
22. C. V. Le, How to choose tuning parameters in Lasso and Ridge regression? *Asian J. Econ. Bank.*, **4** (2020), 61–76.
23. M. E. Anbari, A. Mkhadri, Penalized regression combining the L_1 norm and a correlation based penalty, *Sankhya B*, **76** (2014), 82–102. https://doi.org/10.1007/s13571-013-0065-4
24. Z. J. Daye, X. J. Jeng. Shrinkage and model selection with correlated variables via weighted fusion, *Comput. Stat. Data. An*., **53** (2009). 1284–1298. https://doi.org/10.1016/j.csda.2008.11.007
25. G. H. Fu, W. M. Zhang, L. Dai, Y. Z. Fu. Group variable selection with Oracle property by weight-fused adaptive Elastic Net model for strongly correlated data, *Commun. Stat.-Simul. C.*, **43** (2014), 2468–2481. https://doi.org/10.1080/03610918.2012.752841
26. T. S. Kumar, K.V. Rao, M. Balaji, P. Murthy, D. V. Kumar, Online monitoring of crack depth in fiber reinforced composite beams using optimization Grey model GM (1,N), *Eng. Fract. Mech*., **271** (2022). https://doi.org/10.1016/j.engfracmech.2022.108666
27. T. Tang, W. Jiang, H. Zhang, J. Nie, Z. Xiong, X. Wu, et al., GM (1,1) based improved seasonal index model for monthly electricity consumption forecasting, *Energy*, **252** (2022), 124041. https://doi.org/10.1016/j.energy.2022.124041
28. J. C. Obi, I. C. Jecinta, A review of techniques for regularization, *Int. J. Cognitive Res.*, **11** (2023), 360–367.
29. Y. Wang, W. Zhang, M. Fan, Q. Ge, B. Qiao, X. Zuo, et al., Regression with adaptive lasso and correlation based penalty, *Appl. Math. Model.*, **105** (2022), 179–196. https://doi.org/10.1016/j.apm.2021.12.016
30. M. Gao, L. Xia, Q. Xiao, Goh. M, Incentive strategies for low-carbon supply chains with information updating of customer preferences, *J. Clean. Prod.*, **410** (2023), 137162. https://doi.org/10.1016/j.jclepro.2023.137162