Investment Forecast for Power Grid Technical Renovation Projects Based on Angle of Inclination Correlation and Improved Gray Model

Yue Zhang[1*], Yang Xiao[2], Lilan Deng[3], Wenyu Yan[4]

{zytj14@163.com[1*], 83801555@qq.com[2], 79602018@qq.com[3], yan.wenyu@qq.com[4] }

Electric Power Economic Research Institute, Beijing Electric Power Company, No. 15, Zhanzhan West Street, Xicheng District, Beijing, China

Abstract. This paper introduces an innovative method for forecasting investments in power grid technical renovation projects. By employing the Angle of Inclination Correlation and an Enhanced Gray Model, our goal is to enhance investment prediction accuracy in the electric power industry. Using data from 2016-2021, we select key indicators, such as maximum grid load, power supply capacity, line length renovations, electricity sales revenue, installed capacity, line loss rate, unit substation capacity cost, and power supply reliability, to evaluate their impact on technical renovation projects. The Improved Gray Model is then used to forecast 2022 investments. Subsequently, a Markov model is applied to refine predictions and calculate the deviation rate, confirming the method's effectiveness in indicator selection. While this approach holds promise, it is essential to acknowledge the limitations of historical data, which do not meet big data requirements. Further precision improvements are necessary. Recommendations to enhance the forecasting model's accuracy and applicability include strengthening comprehensive investment lifecycle management and harnessing big data information technology for improved data collection. These measures will contribute to more refined and efficient investment forecasts, supporting decision-makers in the electric power industry.

Keywords: investment forecast; angle of inclination correlation; improved gray model; power grid technical renovation projects

1 Introduction

With China's growing economy and society, the demand for electrical energy is increasing. Investment in power grid technology upgrades is on the rise, leading to new challenges for project managers in power grid companies. Poorly planned investments can result in surplus funds in technology upgrade projects, impeding efficient management **[1]**. Thus, forecasting investments in power grid infrastructure is essential. Employing diverse management models tailored to different projects optimizes resource allocation, enhancing the efficiency and rationality of project management **[2]**.

Technological upgrades significantly impact power station production and environmental efforts, influencing both the development of power industry processes and the economic performance of power grid enterprises. Managing technology upgrade projects encompasses the entire project lifecycle, from planning and initial review to financial auditing and final

ICEMME 2023, November 17-19, Beijing, People's Republic of China

DOI 10.4108/eai.17-11-2023.2342721

acceptance. Effective investment forecasting enables cost control, fund utilization, and impacts project management units' economic interests and the future of electrical engineering. This paper employs mathematical methods to predict technology upgrade project investments, contributing to cost-effective engineering projects and full lifecycle project management.

Using investment data from completed projects in Beijing Company (2016-2021), this paper introduces a methodology for forecasting power grid technology upgrade investments. It utilizes the angle of inclination correlation and an enhanced grey model to analyze influencing indicators. After thorough calculation and analysis, investment forecasts are generated, improving the model's accuracy and applicability.

2 Construction of an investment forecast model for power grid technological upgrade projects based on angle of inclination correlation and improved grey model

2.1 Brief of the angle of inclination correlation and improved grey model

The angle of inclination correlation introduces a novel approach based on the grey slope correlation model, incorporating cosine distance. Unlike traditional methods that calculate correlation coefficients for each time interval and average them for factor correlation, this approach globally optimizes correlation analysis, reducing local information masking and promoting more rational application of correlation algorithms in engineering.

The improved grey model extends the basic grey GM(1, N) model by investigating the collective influence of N-1 variables on a single variable. In predictive analysis using the grey GM(1, N) model, it has revealed inherent issues such as initial and background value construction. These challenges become more prominent when forecasting data with substantial fluctuations, necessitating further optimization of initial values. Moreover, the grey prediction model may produce significant errors when dealing with highly volatile and random data. To address this, an effective integration of Markov chains with grey models can optimize the grey prediction model, enhancing its accuracy in forecasting power grid renovation investments.

2.1.1 Construction of the correlation model

Grey slope correlation is a commonly used correlation algorithm, and its fundamental concept is defined based on the difference in the relative rate of change between two data sequence curves. Slope correlation is based on discrete data sequences, where a higher correlation indicates that the slopes of two data sequences are more similar during the corresponding time intervals. Conversely, a lower correlation indicates greater differences in the slopes of the two data sequences during the corresponding time intervals. The steps for calculating the slope correlation are as follows **[3]**.

Let the original data sequence be $x_i = (x_i(1), x_i(2),..., x_i(n))$, $(i = 0, 1, 2, ..., m)$. The formula for calculating the slope correlation is indicated as can be seen in **equation (1).**

$$r_{0,i} = \frac{1}{n-1}\sum_{k=1}^{n-1} \xi(k) \tag{1}$$

The calculation of $\xi(k)$ is in **equation (2).**

$$\xi(k) = \frac{1}{1 + \left|\frac{\Delta x_0(k)}{x_0(k)} - \frac{\Delta x_i(k)}{x_i(k)}\right|} \tag{2}$$

Where, $\Delta x_i(k)=x_i(k+1)- x_i(k)$,$i =0,1, 2,\ldots, m$; $k=1, 2,\ldots, n -1$.

Slope correlation primarily evaluates local similarity between two data sequences. Unlike Deng's relational degree, it preserves the relative order of correlation sequences during sequence transformations, making it order-preserving. However, slope correlation has limitations in representing negative correlations between two data sequences. Additionally, it faces challenges when $x_0(k)$ or $x_i(k)$ are very small, resulting in correlation values approaching zero and losing meaning. Furthermore, correlation is unattainable when values are zero. Some scholars have introduced modifications to address these issues, such as using the mean of X_0 and X_i to replace $x_0(k)$ and $x_i(k)$.

In response to the limitations of existing grey correlation methods, this paper proposes an enhanced angle slope correlation based on grey slope correlation for indicator selection in power grid renovation investment prediction, as depicted in **Fig 1**.

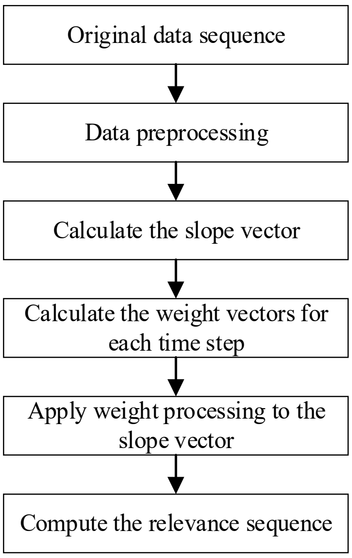

Fig. 1. The process of calculating angular slope relevance algorithm.

The specific construction steps of the angular slope relevance model, as illustrated in **Fig. 1**., are as follows:

For ease of description, let's start by defining and explaining the variables used. The system behavior sequence and relevant indicator data are obtained for a specific time period. We collect the system behavior sequence $Y=\{y(1), y(2),\ldots, y(N)\}$, where $y(n)$ represents the value of the behavior sequence at the nth time point t_n, with n ranging from 1 to N, where N represents the number of time points. Simultaneously, we gather M related factor sequences, denoted as Xi =

{$x_i(1)$, $x_i(2)$,…, $x_i(N)$}, where xi(n) represents the value of the *m*-th related factor at the nth time point t_n, with i ranging from 1 to M.

Preprocessing of data sequences: Both the system behavior sequence Y and the M related factor sequences Xi undergo preprocessing to obtain dimensionless sequences. This results in Y'={y'(1), y'(2),…, y'(N)} for the system behavior sequence and X'_i={$x'_i(1)$, $x'_i(2)$,…, $x'_i(N)$} for each related factor sequence. The calculation formulas for preprocessing are as **equation (3-4)**.

$$y'(n) = \frac{y(n)}{y(1)} \tag{3}$$

$$x_i'(n) = \frac{x_i(n)}{x_i(1)} \tag{4}$$

Calculate the slope vectors by computing the slopes between adjacent time points for both the dimensionless system behavior sequence Y' and the M related factor sequences X'_i. This process yields the slope vector for the system behavior sequence and M slope vectors for the related factors, as **equation (5-6)**:

$$\Delta Y = \{\Delta_y(1), \Delta_y(2), \dots, \Delta_y(N-1)\} \tag{5}$$

$$\Delta X_i = \{\Delta_{xi}(1), \Delta_{xi}(2), \dots, \Delta_{xi}(N-1)\} \tag{6}$$

The formula for calculating the slope is as **equation (7-8)**.

$$\Delta_y(k) = \frac{y(k+1) - y}{t_{k+1} - t_k} \tag{7}$$

$$\Delta_{xi}(k) = \frac{x_i(k+1) - x_i}{t_{k+1} - t_k} \tag{8}$$

Where k = 1, 2, …, N - 1.

Calculate the weight vector for each related factor. Based on the initialized system behavior sequence Y' and the M related factor sequences X'_i, calculate the weight for each historical time point in each related factor sequence. Represent these weights in vector form as Ci = {ci(1), ci(2),…, ci(N-1)}. The calculation formula is as **equation (9)**.

$$c_i(k) = \frac{x_i(k)}{y(k)} \tag{9}$$

Apply weight processing to the slope vectors by multiplying the slope vector of the related factors, iΔX, with the weight vectors for each time point to obtain the weighted slope vector ΔX'. The calculation formula is as **equation (10)**.

$$\Delta X_i'(k) = \Delta X_i \odot C_i = \{\Delta_{xi}'(1), \Delta_{xi}'(2), \dots, \Delta_{xi}'(N-1)\} \tag{10}$$

Where, $\odot$ denotes element-wise multiplication.

Calculate the relevance of each related factor by separately computing the similarity γ between the slope vector ΔY of the system behavior sequence and the vector ΔX' for each of the M related factor sequences. The higher the similarity, the stronger the relationship between the corresponding related factor and the behavior sequence. In this context, we use the cosine of the angle between two slope vectors as the measure of similarity between two data sequences. The calculation formula is as **equation (11)**.

$$\gamma_i = \frac{\Delta X_I' \cdot \Delta_y}{\|\Delta X_I'\| \times \|\Delta_y\|} = \frac{\sum_{k=1}^{n-1} \Delta_{xi}'(k)\Delta_y(k)}{\sqrt{\sum_{k=1}^{n-1}(\Delta_{xi}'(k))^2}\sqrt{\sum_{k=1}^{n-1}(\Delta_y\,(k))^2}} \tag{11}$$

In the angular slope correlation algorithm model, weight vectors play a crucial role by reflecting the spatial proximity between two data sequences. This contradicts the practical interpretation of relevance. When two indicators are geometrically parallel, the closer the associated factor is to the reference sequence in space, the higher the calculated relevance.

The angular slope correlation method utilized in this study considers both the slope factors between sequences and their spatial positioning. It offers a more versatile approach, addressing various aspects of the relationships within data sequences.

2.1.2 Construction of an enhanced grey prediction model with improved initial values

The Grey GM(1,N) model is a result of studying the comprehensive impact of N-1 variables on one variable. When using the model for predictive analysis of a system, some shortcomings of the model itself were observed, particularly in constructing initial values and background values. Errors in predicting data with a certain degree of fluctuation were relatively significant, necessitating further optimization **[4]**.

This paper primarily focuses on optimizing the initial values of the Grey GM(1,N) model, with the following steps in the construction plan:

(1) Determine technical indicators related to power grid renovation investment

Using the angular slope correlation method described in Section 2.2.1 of this paper, calculate the relevance between various technical indicators in the power grid and the investment in power grid renovation. Arrange the power grid technical indicators in descending order of relevance and, based on the collective expertise of power experts, select N-1 power grid technical indicators as the relevant technical indicators for power grid renovation investment.

(2) Obtain historical data for power grid renovation investment

Gather the values of power grid renovation investment and N-1 power grid renovation-related technical indicators corresponding to M time points. Let the vector representing power grid renovation investment be $X_1 = (x_1(1), x_1(2),\ldots, x_1(M))$, where $x_1(j)$ represents the investment in power grid renovation at the jth time point, with j ranging from 1 to M. Also, let the vectors representing N-1 power grid renovation-related technical indicators be $X_i = (x_i(1), x_i(2),\ldots, x_i(M))$, where $x_i(j)$ represents the value of the ith-1 power grid renovation-related technical indicator at the jth time point, with i ranging from 2 to N.

(3) Establish a Grey model optimized by the least squares method

In this paper, the vector of power grid renovation investment, $X_1 = (x_1(1), x_1(2),\ldots, x_1(M))$, is considered as the system characteristic sequence of the Grey model, also known as the behavior variable. The vectors representing N-1 power grid renovation-related technical indicators, $X_i =$

($x_i(1)$, $x_i(2)$,…, $x_i(M)$), are considered as factor variables with a high degree of correlation to the system characteristic sequence. Let $X_i(1) = (x_i(1)(1), x_i(1)(2),…, x_i(1)(M))$ be the first-order cumulative sequence of the power grid renovation-related technical indicator vector X_i, where $x_i(1)(j)$ represents the first-order cumulative value of the ith-1 power grid renovation-related technical indicator at the jth time point. $Z_1(1) = (z_1(1)(1), z_1(1)(2),…, z_1(1)(M))$ is the first-order cumulative sequence of the vector representing power grid renovation investment $X_1(1)$'s adjacent mean value sequence. $z_1(1)(j)$ represents the adjacent mean value of power grid renovation investment at the jth time point. The Grey model GM(1,N) for power grid renovation investment prediction is defined as **equation (12)**.

$$x_1(k) + az_1^{(1)}(k) = \sum_{i=2}^{N} b_i x_i^{(1)}(k) \tag{12}$$

In this context, "a" denotes the development coefficient, while "bi" signifies the coordination coefficient. Let P = [a, b1, b2, ..., bN] T represent the parameter column.

Based on the historical data of grid technology transformation investments in Step S2, the estimated values of the parameter column P are obtained as **equation (13)**.

$$P = (B^T B)^{-1} B^T Y \tag{13}$$

Where,

$$B = \begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \vdots & \vdots & \cdots & \vdots \\ -z_1^{(1)}(M) & x_2^{(1)}(M) & \cdots & x_N^{(1)}(M) \end{bmatrix} \tag{14}$$

$$Y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(M) \end{bmatrix} \tag{15}$$

Therefore, the approximate response of the grey GM(1, N) model is as **equation (16)**.

$$\hat{x}_1^{(1)}(k) = \left(x_1^{(1)}(1) - \frac{1}{a}\sum_{i=2}^{N} b_i x_i^{(1)}(k) \right) e^{-a(k-1)} + \frac{1}{a}\sum_{i=2}^{N} b_i x_i^{(1)}(k) \tag{16}$$

Where $\hat{x}_1^{(1)}(k)$ represents the accumulated investment value of grid technology transformation at the k-th time point, and e denotes the base of the natural logarithm.

The original model assumes that $\hat{x}_1^{(1)}(k)$ passes through the initial point ((1, $x_1^{(1)}(1)$). However, the actual grid technology transformation investment prediction model may not necessarily pass through this point. In order to improve the accuracy of the prediction, this study adopts the principle of minimizing the sum of squared errors between simulated and real values. The specific method is as **equation (17)**.

$$f(c)=\left(x_1^{(1)}(k)-\left(\hat{x}_1^{(1)}(k)\right)\right)^2=\left[x_1^{(1)}(k)-\left(c-\frac{1}{a}m_k\right)e^{-a(k-1)}-\frac{1}{a}m_k\right]^2 \quad (17)$$

Taking the derivative of the above equation and setting f'(c) = 0, we obtain **equation (18)**.

$$c=\frac{\sum_{j=1}^{M}\left(x_1^{(1)}(j)-\frac{m_j}{a}\right)e^{-a(j-1)}+\sum_{j=1}^{M}\frac{m_j}{a}e^{-a(j-1)}}{\sum_{j=1}^{M}e^{-a(j-1)}} \quad (18)$$

Where,

$$m_j=\sum_{i=2}^{N}b_i x_i^{(1)}(j) \quad (19)$$

Substituting the obtained c into Equation (16), we obtain the optimized grey model as **equation (20)**.

$$\hat{x}_1^{(1)}(k)=\left(c-\frac{1}{a}\sum_{i=2}^{N}b_i x_i^{(1)}(k)\right)e^{-a(k-1)}+\frac{1}{a}\sum_{i=2}^{N}b_i x_i^{(1)}(k) \quad (20)$$

(4) Grid Technology Transformation Investment Prediction:

Obtain the first-order cumulative value of the relevant technical indicators for grid technology transformation investments at the predicted time point k and the preceding time point k-1, denoted as $x_i^{(1)}(k-1)$. For the indicator value at time point k, this paper employs the GM(1,1) model. It is then incorporated into the grey model from Step S3 to acquire the first-order cumulative values of grid technology transformation investment predictions for time points k and k-1, denoted as $\hat{x}_1^{(1)}(k)$ and $\hat{x}_1^{(1)}(k-1)$ respectively. Consequently, the grid technology transformation investment prediction at time point k is given by **equation (21)**.

$$\hat{x}_1(k)=\hat{x}_1^{(1)}(k)-\hat{x}_1^{(1)}(k-1) \quad (21)$$

2.1.3 Construction of an enhanced grey prediction model for improved forecasting

When forecasting data characterized by high volatility and randomness, the Grey Prediction Model may sometimes yield significant errors, which exhibit a stochastic nature. The Markov model, dedicated to studying transition patterns among random states, provides a method to effectively integrate with the Grey Model. By employing specific techniques to merge the Markov chain with the Grey Model, the Grey Prediction Model is optimized. This enhancement further elevates the forecasting precision of grid technology transformation investments. The process of refining grid technology transformation investment predictions using the Markov model is depicted in **Fig 2**.

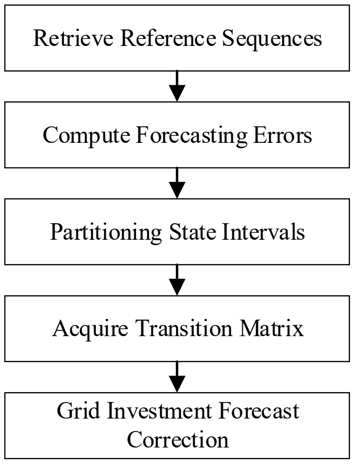

Fig. 2. The process of Markov model for grid technology transformation investment forecast correction.

(1) Acquisition of Reference Data

Set the number of reference data time points to be Q, selecting Q time points from historical data. Typically, to enhance prediction accuracy, it is advisable to choose the Q time points nearest to the predicted time point, denoted as k. Let the actual electric power grid renovation investment values at these Q time points be denoted as y_q, where q = 1, 2, ..., Q. Employ parameter-optimized grey models to compute the corresponding electric power grid renovation investment predictions, denoted as $\hat{y}_q$.

(2) Calculation of Prediction Errors

Calculate the prediction errors corresponding to each time point for electric power grid renovation investments as **equation (22)**.

$$e_q = \frac{\hat{y}_q - y_q}{y_q} \tag{22}$$

(3) Division of State Intervals

Let the minimum and maximum values of the electric power grid renovation investment prediction errors be denoted as e_{min} and e_{max}, respectively. Divide the prediction error interval $[e_{min}, e_{max}]$ into H state intervals $[E_{1h}, E_{2h}]$, where h = 1, 2, ..., H. To facilitate predictions, the value of H should not be less than the difference between the nearest time point k' among Q time points and the prediction time point k, i.e., $H \geq k-k'$.

(4) Acquisition of State Transition Probability Matrix:

Record the state transitions of prediction errors between two time points. Let the number of times state E_j transitions to state $E_{j'}$ after h steps be denoted as $a_{jj'}(h)$, where j, j' = 1, 2, ... The number of times transitions originate from state E_j is denoted as β_j. Therefore, the probability of prediction error transitioning from state E_j to state $E_{j'}$ after h steps is defined as $p_{jj'(h)} = a_{jj'(h)}/\beta_j$. This leads to the h-step state transition probability matrix P(h) as **equation (23)**.

$$P(h) = \begin{matrix} p_{11}(h) & p_{12}(h) & \cdots & p_{1H}(h) \\ p_{21}(h) & p_{22}(h) & \cdots & p_{2H}(h) \\ \vdots & \vdots & & \vdots \\ p_{H1}(h) & p_{H2}(h) & \dots & p_{HH}(h) \end{matrix} \tag{23}$$

In the practical computation process, it is common to first calculate the one-step state transition matrix based on the error states between real values and predictions from historical data, where h takes the value of 1. Then, the two-step to H-step state transition matrices are derived based on the relationships between the state transition matrices.

(5) Correction of Electric Power Grid Renovation Investment Predictions:

Select G time points from the Q time points where the difference between their time and the prediction time point k is less than H. Let the number of such time points be G, and denote these time points as kg, where g = 1, 2, ... , G. Using the prediction error states at these G time points and the H state transition probability matrices P(h), calculate the probabilities of transitioning to various state intervals from these G time points kg to the prediction time point k after k-kg steps. Determine the total probability for each state interval and select the state interval E_{h*} corresponding to the maximum total probability as the prediction error state interval for the electric power grid renovation investment prediction at time point k. Calculate the average value of the prediction errors in state interval E_{h*} as e*. Then, adjust the electric power grid renovation investment prediction $\hat{x}_1(k)$ based on the following formula to obtain the corrected electric power grid renovation investment prediction $\hat{x}_1(k)$.

3 Case study analysis

3.1 Indicator selection

One of the most critical aspects in the process of estimating electric power grid renovation investment is the selection of relevant indicators. The appropriateness of indicator selection directly affects the effectiveness of investment predictions. The selection of indicators for electric power grid renovation investment must begin with a clear understanding of the general structure of the relevant indicators. The indicator extraction plan for electric power grid renovation investment constructed in this paper primarily involves three aspects.

First, a preliminary organization and categorization of the data is performed. This involves clarifying the physical meanings of each indicator and consolidating and organizing data that are duplicated or closely related. Data sequences that show minimal variation over time are removed. For data with a small number of missing values, appropriate methods are used for data imputation to fill in the gaps. Common methods include mean imputation and interpolation. Mean imputation involves calculating the average of available data from neighboring years to estimate missing data. This method is relatively simple but may have lower accuracy. On the other hand, interpolation is a more complex method that uses mathematical techniques to estimate missing data based on the surrounding data points. It often yields higher accuracy compared to mean imputation. However, in cases where data is severely missing, it's advisable to delete entire data records with substantial gaps as they might not provide meaningful information for analysis.

After the preliminary organization of data and the selection of relevant indicators, the next step is to perform a correlation analysis using the "angle slope correlation coefficient algorithm" proposed in this paper. This analysis will help determine the degree of correlation between each indicator and the amount of electric power grid renovation investment.

Finally, by integrating the insights of relevant experts in the field of electric power and considering the specific requirements of the predictive model, the ultimate selection of indicators most closely associated with electric power grid renovation investment is determined.

3.1.1 Indicator organization

The process of organizing indicators begins with understanding their physical meanings, logical relationships, and the statistical data among them, with input from relevant professionals. The aim is to select practical indicators that offer comprehensive coverage and effectively represent the correlation with electric power grid renovation investment. Due to the extensive range of power-related data indicators associated with these investments, they cannot all be listed here.

When dealing with data containing a small number of missing values, it is advisable not to delete entire data records. Instead, appropriate data imputation methods should be employed for completion. Common imputation methods include mean imputation and interpolation/extrapolation. Mean imputation entails averaging surrounding years' data to fill in missing data for a given year, offering a simple solution but potentially lacking precision. In contrast, interpolation involves fitting missing data based on surrounding years' available data using suitable methods. While interpolation methods are more intricate compared to mean imputation, they generally offer higher accuracy in simulating missing data.

After the initial analysis, the data matrix, which compiles data from different systems with varying formats, is denoted as "D." The format of the data matrix D is as **equation (24)**.

$$D = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ x_{11} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{24}$$

Where, "y" denotes the data sequence for electric power grid renovation investment amounts; "x" represents the data sequence group of relevant indicators; "n" stands for the number of years with available historical data from the electric power grid company; "m" refers to the count of selected indicators following the initial organization.

3.1.2 Indicator selection based on the correlation model

In this section, the specific steps for calculating the angular correlation between relevant indicators and electric power grid renovation investment are provided.

Taking the example of calculating the correlation between the line loss rate indicator and electric power grid renovation investment, let's denote the historical data of line loss rate for the past "n" years as "X" and the electric power grid renovation investment for the past "n" years as "Y." We calculate the slope vectors for electric power grid renovation investment and line loss rate for each time interval as ΔY and ΔX, respectively. The weight vector for line loss rate at each time moment is calculated as "C."

Table 1. Electric Power Grid Renovation Investment Forecast Related Indicators.

Indicator names	Indicator Units	Attributes	Significance	Correlation
Electricity Sales Revenue	Yuan	Benefit	Revenue Generated from Electricity Sales in the Area of the Technical Reform Project	0.8967
Line Loss Rate	%	Cost	One of the effective measures reflecting grid efficiency.	0.8014
Equipment Utilization Hours	h	Benefit	Operating Hours at Rated Power for Equipment Over a Specified Period (Usually Calculated Annually)	0.7729
Unit Electricity Transmission and Distribution Cost	Yuan/kwh	Cost	The Total of Prices Provided by the Electric Grid Operating Company for Access Systems, Electricity Transmission, and Sales Services, as a Ratio to the Volume of Electricity Transmitted and Distributed	0.6743
Share of Clean Energy	%	Benefit	Proportion of Clean Energy in the Regional Power Generation Equipment of the Technical Reform Project Area	0.6086
Cost of Unit Substation Capacity	Yuan/kVA	Cost	The Ratio of the Total Cost of Transformers to Transformer Substation Capacity	0.5893
Electric Grid Maximum Load)	Ten thousand·kW	Benefit	The Maximum Annual Load of the Entire Society within the Supply Administrative Area	0.5187
Workforce Size	People	Cost	Workforce Scale and Costs Invested by the Electric Power Grid Company in the Process of Grid Engineering Construction for the Technical Reform Project	0.4978

Following initial indicator organization, angular correlation between various indicators and electric power grid renovation investment is computed. Eight specific indicators (in **Table 1**) are selected for correlation analysis, considering both positive and negative correlations with electric power grid renovation investment. Angular correlation is initially calculated using historical data from 2016 to 2020, and subsequently from 2017 to 2021.

Once the correlation sequence is established through angular correlation calculations, an in-depth evaluation is conducted by experts in collaboration with professionals from the electric power grid industry. These experts provide further insights into the indicators, although this paper does not delve into the specifics.

Following the angular correlation calculations and the expert evaluations, eight indicators that are closely related to electric power grid renovation investment are ultimately chosen. The specific information about these indicators is presented in **Table 1**.

3.2 Appli electric power grid renovation investment forecast based on improved grey models cability analysis of existing cost calculation methods

Following the method steps outlined in Section 2.2.2, we sequentially selected 1 to 8 relevant indicators from **Table 2** and established corresponding multidimensional grey models. We calculated the relative errors between the real values and simulated values for each model, and the average relative residuals for each model are presented in **Table 2**.

Table 2. Average Relative Residuals for Each Model.

Number of Indicators	Grey Model	Average Residual
1	GM(1,2)	30.57%
2	GM(1,3)	25.31%
3	GM(1,4)	15.68%
4	GM(1,5)	10.75%
5	GM(1,6)	14.83%
6	GM(1,7)	32.19%
7	GM(1,8)	89.56%
8	GM(1,9)	121.26%

Comparing various grey models, we find that the Grey GM(1, N) model exhibits larger average relative residuals with fewer indicators. As the number of indicators increases, these residuals gradually decrease. This underscores that forecasting electric power grid renovation investment is influenced by multiple indicators. Utilizing multiple indicators provides a more accurate reflection of investment patterns. However, an excessive number of indicators significantly raises the model's average relative residuals. Excessive indicators result in a large condition number for the internal matrix due to the model's unique construction. In ill-posed equations, minor changes in observed values lead to substantial errors in model predictions.

Using the above method, the selection of Electricity Sales Revenue, Line Loss Rate, Equipment Utilization Hours, and Unit Electricity Transmission and Distribution Cost was made to establish a Grey GM(1,5) model for predicting electric power grid renovation investment. Specific calculation results are presented in **Table 3**.

Table 3. Model Simulation Results.

Year	Real Value	Simulated Value	Relative Error
2016	5.57	5.57	0
2017	9	8.76	-2.67%
2018	12.21	13.04	6.80%
2019	15.57	17.47	12.20%
2020	6.56	6.88	4.88%
2021	8.92	9.98	11.88%
2022	8.61	9.27	7.67%

Following the method described in section 2.2.3, further adjustments were made to the parameter-optimized grey model. As indicated in **Table 4**, the minimum relative error is -2.67%, and the maximum relative error is 11.88%. Considering that there are 6 historical data points, they are divided into three equally weighted states, as shown in **Table 4**.

Table 4. State Division.

State	E1	E2	E3
State Boundaries	[-2.67%,2.18%)	[2.18%,7.03%)	[7.03%,11.88%)

Based on the state division table, the simulated predicted values for the years 2016 and 2017 are in state E1, for the years 2018 and 2020 are in state E2, and for the years 2019 and 2021 are in state E3. Therefore, the one-step state transition probabilities are calculated as **equation (25)**.

$$P(1)=\begin{Bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{Bmatrix} \tag{25}$$

Calculating the two-step state transition matrix based on the one-step state transition matrix as **equation (26)**.

$$P(2)=P(1)*P(1)=\begin{Bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{Bmatrix} \tag{26}$$

Obtaining the three-step state transition matrix by combining the one-step and two-step state transition matrices as **equation (27)**.

$$P(3)=P(2)*P(1)=\begin{Bmatrix} \frac{1}{4} & \frac{3}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{Bmatrix} \tag{27}$$

Because the Markov states are divided into three intervals, the prediction of the 2022 grid technology renovation investment state, based on historical data from 2019, 2020, and 2021, is presented in **Table 5**.

Based on the state prediction results, the 2022 grid technology upgrade investment is in state E2. The corrected error, denoted as e*, is calculated as (2.18% + 7.03%) / 2, resulting in 4.61%. The parameter-optimized grey model predicts the investment for the year 2022 to be 9.27 billion RMB, and the corrected forecast value is 8.86 billion RMB, yielding an error of 2.90%. This relatively small error further substantiates the effectiveness of the selected indicators proposed in this study, demonstrating their capacity to better reflect the impact on grid technology upgrade investments. It also underscores the suitability of improving the grey model forecasting method for grid technology upgrade investment prediction.

Table 5. Predicted 2022 Grid Technology Renovation Investment State Results.

Year	Initial State	Transition Steps	State Transition Probability		
			E1	E2	E3
2021	E3	1	0	1	0
2020	E2	2	0	3/4	1/4
2019	E3	3	0	3/4	1/4
Total			0	10/4	1/2

4 Conclusions

Based on data from grid technology upgrade project investments spanning 2016-2021, this paper employs angle slope correlation to select key indicators. These indicators encompass maximum grid load, power supply capacity, line length renovations, electricity sales revenue, installed capacity, line loss rate, unit substation capacity cost, and power supply reliability, which are analyzed for their influence on technology upgrade projects. An enhanced grey model is then developed to predict 2022's grid technology upgrade project investment. Additionally, the application of the Markov model for correction and deviation rate calculation further validates the effectiveness of the proposed indicator selection method and underscores the suitability of the improved grey model for investment prediction in grid technology upgrades.

However, it's crucial to note that the historical data used in this study still falls short of big data requirements, and result precision requires improvement. In response to the shortcomings of the proposed method, this paper suggests reinforcing the entire investment cycle management to enhance the decision-making capability for technological renovation projects **[5]**. Furthermore, it is recommended to leverage big data information technology to strengthen the collection of fundamental data, thus further improving the accuracy and applicability of the predictive model.

Reference

[1] Li, J., Chen, W., Chen, Y., Sheng, K., Du, S., Zhang, Y., & Wu, Y. (2021). A survey on investment demand assessment models for power grid infrastructure. IEEE Access, 9, 9048-9054.

[2] Dunnan, L., Erfeng, X., Guangyu, Q., & Ying**g, H. (2019). Research on reasonable grid investment scale measurement under the reform of transmission and distribution electricity price. Energy Procedia, 156, 412-416.

[3] Oyewole, A. (2015, September). Extension of the gray correlation to inclination angles. In SPE Annual Technical Conference and Exhibition? (p. D023S099R004). SPE.

[4] Cheng, M., & Shi, G. (2022). Improved methods for parameter estimation of gray model GM (1, 1) based on new background value optimization and model application. Communications in Statistics-Simulation and Computation, 51(2), 647-669.

[5] Eash-Gates Philip, Klemun Magdalena M.,Kavlak Goksin,McNerney James,Buongiorno Jacopo & Trancik Jessika E..(2020).Sources of Cost Overrun in Nuclear Power Plant Construction Call for a New Approach to Engineering Design. (11),2348-2373.