

International Journal of Powertrains

ISSN online: 1742-4275 - ISSN print: 1742-4267

<https://www.inderscience.com/ijpt>

Short-term load forecasting technique for power system based on grey correlation analysis and factor analysis

Xiaoguo Zhang

DOI: [10.1504/IJPT.2025.10071545](https://doi.org/10.1504/IJPT.2025.10071545)

Article History:

Received:	21 January 2025
Last revised:	31 March 2025
Accepted:	27 April 2025
Published online:	10 July 2025

Short-term load forecasting technique for power system based on grey correlation analysis and factor analysis

Xiaoguo Zhang

School of Mathematics and Physics,
Henan University of Urban Construction,
Pingdingshan, 467036, China
Email: zhangxiaoguo0530@163.com

Abstract: Aiming at the problem that traditional load forecasting techniques cannot achieve higher forecasting accuracy, a short-term load forecasting technique for power systems based on grey correlation analysis and factor analysis is proposed. Grey correlation analysis is introduced to determine the key load factors, and the forecasting circuit is designed by combining the data from the grid visualisation system. Secondly, the main influencing factors are extracted to form a composite factor by dimensionality reduction through factor analysis. Experiments show that the model performs well in predicting different date types and zone. The difference between the model predicted and true values was less than 0.6%. The results show that the load forecasting technique proposed in the study has high accuracy and stability, which provides a strong support for the stable operation and planning of the power system.

Keywords: power system; short-term load forecasting; STLF; grey correlation analysis; GCA; factor analysis; grid visualisation management.

Reference to this paper should be made as follows: Zhang, X. (2025) 'Short-term load forecasting technique for power system based on grey correlation analysis and factor analysis', *Int. J. Powertrains*, Vol. 14, No. 1, pp.78–96.

Biographical notes: Xiaoguo Zhang obtained her BE in Science from Henan University in 2005. She obtained her ME in Applied Mathematics from Wuhan University in 2007. Presently, she is working as a Lecturer in the Department of Mathematics and Applied Mathematics, School of Mathematics and Physics, Henan University of Urban Construction. Her areas of interest are mathematical modelling and application, optimisation theory and control.

1 Introduction

Accurate short-term load forecasting (STLF) is essential for energy management, economic dispatching, and the safe and stable functioning of the power system (PS) as the system continues to evolve and intelligence levels rise. However, many traditional methods rely on linear trend analysis of historical data, which makes it difficult to effectively capture the nonlinear change characteristics of power loads, and the prediction accuracy decreases dramatically in the face of rapidly changing economic environment and extreme weather conditions. Second, existing methods are deficient in dealing with

the interaction of multiple factors. The influence of multiple factors such as meteorological factors, social activities, holiday effects, etc. on power load is complex and intertwined, and traditional methods are often difficult to simultaneously consider the combined influence of these factors, leading to large deviations in forecasting results. In the meantime, the power demand exhibits a more complex trend of change because of the economy's quick development and society's ongoing advancement (Sun et al., 2024; Meng et al., 2022). On the one hand, a number of variables, like the weather and season, influence the power load, making it challenging to precisely understand how the load changes. However, the PS has amassed a substantial amount of historical load data and associated influencing factor data as a result of the development of the smart grid and the use of big data technology (Fan et al., 2022b). It has become an important problem in the current PS research to effectively mine the potential information in these data and improve the accuracy and reliability of load forecasting (LF).

In recent years, artificial intelligence and data analysis techniques have made significant progress in various fields, providing new ideas and methods for PS LF (Fan et al., 2022a; Tong et al., 2022). Numerous academics have conducted studies on PS's STLF. Kondaiah et al. proposed an accurate microgrid model for STLF of PS. The study quickly identified and constructed target data by simulating the operating environment of the PS. The usefulness of the suggested model in power forecasting was shown by the experimental findings (Kondaiah et al., 2022). Yi et al. (2023) proposed a STLF model based on the output dimension in an attempt to solve the problem of missing LF features in PSs. The study extracted the long-term behaviour of the data through a long and short-term memory network, emphasising the key input load information. According to the findings, the suggested model showed a high prediction accuracy (Yi et al., 2023). Lee et al. (2023) pointed out the limitations of the traditional feature importance metric based on the output values of machine learning models and proposed a new metric incorporating the coefficient of determination. The outcomes indicated that this method had good stability in LF (Lee et al., 2023). To model load calculation, Panda proposed an intelligent power LF method based on deep neural networks (DNN). The study used spatio-temporal correlation technique to construct neural networks by combining cross-multi-modelling and quadratic decision-making mechanisms. The suggested approach performed better in LF, according to simulation results (Panda and Ray, 2022). A hybrid algorithm model that takes algorithm parameter settings and load volatility into account was presented by Chen et al. The outcomes revealed that the suggested approach might successfully raise the prediction accuracy (Chen et al., 2022).

As effective data analysis tools, grey correlation analysis (GCA) and factor analysis have unique advantages in dealing with complex data relationships and extracting key information. Grey correlation and factor analysis can simplify data, increase the effectiveness and precision of predictive models, and uncover public factors concealed in vast amounts of data. Several scholars have carried out studies on GCA and factor analysis. Rajesh et al. (2024) used grey correlation classification to optimise the electrode coating EDM process of mixed abrasive powders with high material removal rate to assess the effect of their properties on high temperature alloys. The GCA technique based on response surface methodology has been proposed by Gunasekaran et al. (2024) in an attempt to optimise the process parameters in the mixing process of power alloys. The study utilised GCA technique to develop a multi-objective numerical model based on central combinatorial design. The outcomes revealed that this technique had good

optimisation performance (Gunasekaran et al., 2024). Candan et al. used GCA to rank the sustainable industrialisation performance of countries, reveal the level of sustainable industrialisation in these countries, and determine the evaluation criteria weights and areas to be improved. The results indicated the validity of the proposed method (Candan and Cengiz Toklu, 2022). Kavimani et al. (2022) used Taguchi coupling GCA to optimise the multi-response of processability of polymer matrix composites for power transportation. The study explored the output response of the input parameters to obtain the optimal control factors. The outcomes revealed that the suggested approach was successful in enhancing the material's processability (Kavimani et al., 2022). Samuel et al. (2024) used grey correlation and factor analysis in an attempt to optimise natural fibre reinforced composites for wind turbine technology applications. The study developed and optimised natural fibre and synthetic fibre polymer composites through GCA by Taguchi robust optimisation technique. The experimental findings demonstrated that the suggested approach might enhance the composites' performance (Samuel et al., 2024).

Combined with the above, it can be seen that related scholars have carried out multifaceted research on STLF, which has promoted the development of PS to a certain extent. However, some methods have poor model interpretations when dealing with multi-factor interactions, and it is difficult to clarify the actual contribution of each factor to the load. Some methods require high data quality, have insufficient adaptability to data, have high computational complexity when facing large-scale PSs, and are difficult to meet real-time requirements. In addition, the existing methods still have deficiencies in computational efficiency, model interpretability, and data adaptability when dealing with complex PS LF. In view of this, the study innovatively proposes STLF of PS technique based on GCA and factor analysis. Factor analysis of the main factors is conducted to establish a LF model, aims to address the shortcomings of existing methods in complex LFs and to provide more accurate and reliable forecasting support for stable PS operation and planning.

The novelty of the study is the proposal of a PS STLF method based on GCA and factor analysis. The method quantifies the correlation between loads and multiple influencing factors through GCA, which can more accurately identify the key factors that have the greatest influence on the load changes, and makes up for the shortcomings of the traditional methods in dealing with the interaction effects of multiple factors. Meanwhile, factor analysis is used for dimensionality reduction processing to synthesise multiple correlated factors into a few uncorrelated composite factors, which not only simplifies the data structure, but also improves the model's interpretability and prediction accuracy. The study further constructs a comprehensive STLF model, which fully utilises the advantages of the two methods, improves the accuracy and stability of the prediction, and provides more reliable prediction support for the stable operation and planning of the PS.

2 Technical study of STLF of PS based on GCA and factor analysis

2.1 GCA-based STLF of PS technique

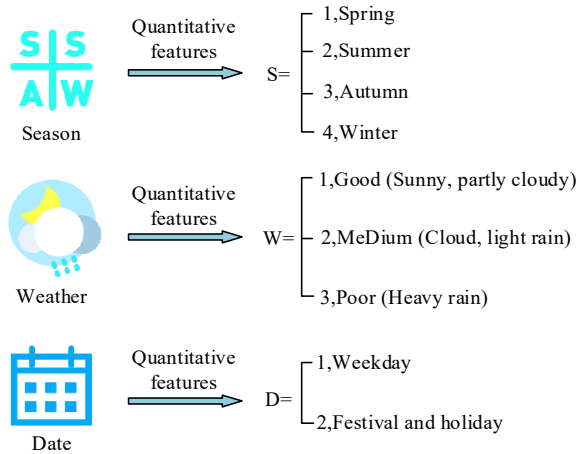
GCA is a multifactor statistical analysis method used to analyse the degree of association between factors. The core idea is to quantify the impact of each factor on load by calculating the degree of correlation between the reference series (historical load data)

and the comparison series (meteorological factors, date types, socio-economic factors, etc.) GCA can reveal the similarity between different factors and load variations, thus identifying the key factors that have a greater impact on load. GCA has unique advantages in STLTF of PS due to its characteristics of low data requirements, no need for a large amount of sample data, and the ability to handle uncertain information and small sample data (Mokayed et al., 2023). First, it is possible to rapidly ascertain the degree of correlation between load and various affecting elements. The GCA features are defined as shown in equation (1).

$$I = \frac{q \left(\sum Q + \beta R_{\max} \right)}{k \cdot \Delta T} \quad (1)$$

In equation (1), I represents the correlation feature. q represents the grey level processing authority in the data information. β represents the forecast volume extraction coefficient. ΔT represents the unit prediction duration of short-term electricity load. Q represents the basic data information correlation coefficient correlation coefficient (COC), which is used to measure the correlation between the data. $R_{\max}$ represents the maximum value of the load during the forecast period, and k represents the STLTF forecast data. It is possible to determine the extent to which various factors affect load by computing the correlation between the reference series and each comparison series. The quantification of correlation factors using GCA can make the complex influence relationship clearer and understandable (Gheisari et al., 2023). The quantification of correlation factors is shown in Figure 1.

Figure 1 Quantification method of related factors (see online version for colours)



In Figure 1, power loads have different characteristics and patterns of change under different seasons, weather and date types. These parameters can be quantified and then added as input variables to the LF model to help the model better capture the link between load and these factors, which will increase the forecast's accuracy. The study coded the four seasons as 1, 2, 3, and 4, representing spring, summer, fall, and winter, which facilitates their use in data analysis and modelling. Then seasonal feature vectors (FVs) are constructed, where the four elements of the vectors correspond to spring,

summer, fall and winter, which can be easily combined and processed with other features. Then the classification and coding are carried out according to common weather types to distinguish different weather conditions. Moreover, a comprehensive weather quantisation index is constructed, and a comprehensive weather quantisation value is calculated, which facilitates the consideration of its impact on load in the analysis. Binary coding is used for weekdays and non-workdays to quickly differentiate between different holiday types. The load of the PS is dynamically changing, and seasonal changes, sudden weather changes, and different date types all have an impact on the load (Zhang et al., 2022; Dong et al., 2022). Quantifying these factors can make the forecasting model better adapt to such dynamic changes, adjust the forecasting results in time, and optimise the resource allocation while facilitating the planning and management of the PS. The constructed eigenvectors are shown in equation (2).

$$T_z = [S, W, d, T_{\max}, T_{\min}, v] \quad (2)$$

In equation (2), T_z is the constructed eigenvector. S is season. W is weather. d represents date type. $T_{\max}$ and $T_{\min}$ are maximum and minimum temperatures, respectively. v represents the wind speed of the day. The grey COCs of the characteristic series of historical and forecast data are shown in equation (3).

$$\zeta(T_{z0}(k), T_{zi}(k)) = \frac{\min_i \min_k |T_{z0}(k) - T_{zi}(k)| + \rho \max_i \max_k |T_{z0}(k) - T_{zi}(k)|}{|T_{z0}(k) - T_{zi}(k)| + \rho \max_i \max_k |T_{z0}(k) - T_{zi}(k)|} \quad (3)$$

In equation (3), ζ represents the grey COC of the feature quantity. T_{z0} represents the FV of the predicted data. T_{zi} represents the FV of the i^{th} historical data. $\min_i \min_k |T_{z0}(k) - T_{zi}(k)|$ and $\max_i \max_k |T_{z0}(k) - T_{zi}(k)|$ are the minimum and maximum difference of two-vector correlation. ρ represents the resolution factor. The influence factor feature grey correlation degree of historical and predicted data is shown in equation (4).

$$R(T_{z0}, T_{zi}) = \frac{1}{n} \sum_{k=1}^n \zeta(T_{z0}(k), T_{zi}(k)) \quad (4)$$

In equation (4), R represents grey correlation degree. n represents the number of grey COC averages. The prediction circuit designed by combining grey analysis is shown in Figure 2.

In Figure 2, the prediction circuit is able to control and regulate the short-term power load level in the grid environment under the GCA law. The prediction circuit has several different levels of resistors, which can regulate the voltage loads on both sides of the resistor parameters and integrate the short-term power loads to be transmitted. The prediction circuit is a key component of the GCA-based design for data conditioning and preliminary processing in the short-term load prediction model. It dynamically reflects the PS short-term load trend by simulating changes in resistance and voltage load. Different resistance levels regulate the voltage loads on both sides of the circuit to simulate the system operation under different load conditions. These regulated data are integrated and inputted into the prediction model to help the model capture the change rule of load more accurately and improve the prediction accuracy and stability. Grid visualisation management system (GVMS) contains rich PS operation data, such as historical load data, meteorological data, and equipment operation status data. Therefore, it is investigated to utilise GVMS to collect and integrate multiple data sources of PS,

including historical load data, meteorological data, and equipment operation status data. After preliminary processing by GVMS, it is transmitted to the prediction model. Figure 3 depicts the GVMS's organisational structure.

Figure 2 Prediction circuit designed using grey analysis method (see online version for colours)

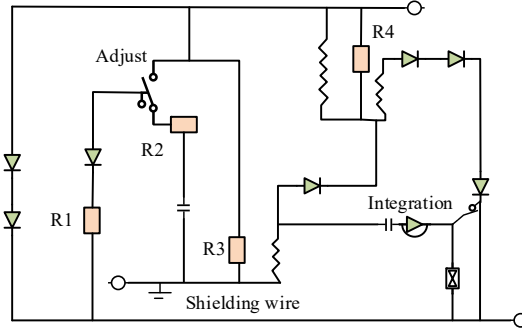


Figure 3 Structure of GVMS (see online version for colours)

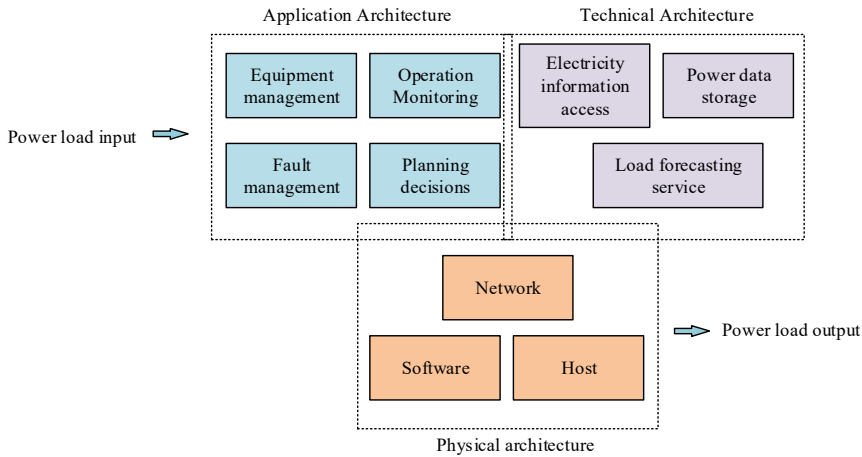


Figure 3 show that GVMS has a comprehensive architecture and functional modules. In terms of application architecture, it integrates several key aspects such as equipment management, operation monitoring, fault management and planning decision support, realising full coverage from daily maintenance of equipment to overall operation analysis of the system. In terms of technical architecture, advanced information technology is utilised to ensure efficient data collection, processing and visualisation display. In terms of data collection, GVMS utilises its advanced sensor network and data acquisition modules to collect real-time PS operational data. In order to eliminate the difference in magnitude between different data sources, GVMS standardises the data. The specific formula is shown in equation (5).

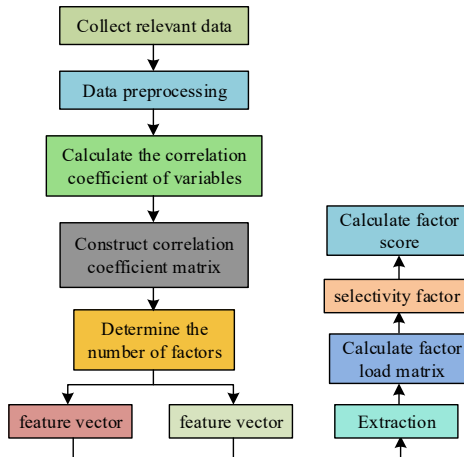
$$x_{norm} = \frac{x - \mu}{\sigma} \quad (5)$$

In equation (5), x denotes the original data; μ denotes the data mean; σ denotes the standard deviation of the data; and x_{norm} denotes the standardised processed data. For outliers, GVMS uses Z-score for detection and processing. On this basis, GVMS fuses data from different data sources to generate comprehensive FVs. The physical architecture, on the other hand, guarantees stable operation and reliable deployment of the system. Among the functional modules, the equipment management module carries out full life cycle management of grid equipment. The operation monitoring module grasps the operation status of the power grid in real time. Fault management module responds quickly to locate, analyse and deal with faults when they occur. The planning decision support module uses both historical and current data to guarantee grid operating safety. The study combines GVMS with GCA. The system can provide basic data support for GCA. Accordingly, the results of GCA, such as the correlation between factors and loads, can be fed back into the system to finally determine the main factors affecting loads.

2.2 STLF of PS model combined with factor analysis

After using GCA to determine the main factors affecting the load, it is found that GCA mainly focuses on examining the degree of association between each factor and the load. For complex PS loads, GCA may face increased computational complexity and difficulty in accurately identifying key factors when there are multiple factors interacting with each other. Moreover, the internal organisation of the influencing elements and the possible common influencing factors are hard to thoroughly examine. Factor analysis, on the other hand, can synthesise multiple correlated influencing factors into a few uncorrelated composite factors. These composite factors can overcome the drawbacks of GCA single component analysis by more accurately reflecting the combined impact of several factors on the load (Veeramsetty et al., 2022). The study combines factor analysis with grey correlation to dig deeper into the internal structure of the influencing factors by means of dimensionality reduction to provide a deeper understanding of LF. Specifically, it is manifested in the factor analysis of the main factors and the extraction of the composite factors. Figure 4 illustrates the factor analysis procedure in PS's STLF.

Figure 4 The process of factor analysis in STLF of PSs (see online version for colours)



In Figure 4, in STLF of PS, factor analysis first collects and pre-processes relevant data. Then the matrix of COCs between variables is calculated and the number of factors is determined. Next, the common factors are extracted and their meanings are interpreted, and the interpretation is made clearer by rotating the factors. During factor analysis, the extracted composite factors not only simplified the data structure, but also provided a deeper explanation of the STLF model. Through factor analysis, the study synthesised several related influencing factors into several unrelated composite factors. These composite factors can more accurately reflect the joint influence of the factors on load, thus improving the prediction accuracy and stability of the model. In addition, factor analysis not only simplifies the data structure, but also reduces data redundancy and improves data processing efficiency. Specifically, factor analysis compresses the information in the original data into a few composite factors by extracting the principal components, and these composite factors can explain most of the variance of the original data.

The hypothetical factor analysis extracts two main synthesising factors: the first synthesising factor mainly reflects the influence of meteorological conditions on loads, for example, changes in temperature directly affect the frequency of air conditioning and heating, thus affecting electricity loads. The second composite factor, on the other hand, mainly reflects the impact of socio-economic activities on loads. For example, on weekdays, commercial and industrial electricity demand is higher, while on holidays, residential electricity demand may increase. By combining these composite factors with GCA, the model not only captures the correlation between each factor and load, but also reflects the joint impact of these factors more accurately through the composite factors. Finally, the factor scores are calculated and applied to the LF model. The factor model is shown in equation (6).

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\ \dots \\ X_m = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_p \end{cases} \quad (6)$$

In equation (6), $X = (X_1, X_2, \dots, X_p)^T$ represents the random vector. $F = (F_1, F_2, \dots, F_m)^T$ represents the common factor. P is the indicator quantity. $A = (a_{ij})^2$ is the factor loading matrix. a_{ij} represents factor loadings. ε represents special influences other than the common factor. Data are then collected on multiple variables related to short-term loadings. For each variable to have equal weight in the analysis, the data must be normalised because distinct variables may have different scales and orders of magnitude. To determine the degree of linear correlation between two variables, the COC between each variable is computed for the standardised data. The COC is shown in equation (7).

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (7)$$

In equation (7), r_{ij} represents the COC. n is the samples' quantity. x_{ki} is the value of the i^{th} variable of the k^{th} sample. $\bar{x}_i$ is the mean value of the variable. The COCs between the variables are then arranged into a matrix to solve for the eigenvalues as shown in equation (8).

$$|R - \lambda I| = 0 \quad (8)$$

In equation (8), R represents the COC matrix. λ represents the eigenvalues. I represents the unit matrix. Eigenvectors are calculated for each eigenvalue. The eigenvalues represent the variance contribution of each factor, and the eigenvectors represent the loadings of each variable on each factor. The factor loading matrix is built to examine the level of correlation between the variables and the factors, and the number of factors is established by the size of the eigenvalues as the original composite factor. The factor loading matrix is also rotated in order to make the meaning of the factors clearer. The rotated factor loading matrix makes the loadings of each variable on a small number of factors more concentrated, thus making it easier to interpret the meaning of each factor. The factor scores are calculated by regression method, and the extracted composite factors are used as input variables to establish the LF model as shown in equation (9).

$$L(t) = L_{\max} P_w(t) P_d(t) P_t(t) \quad (9)$$

In equation (9), $L(t)$ represents the load forecast value at the moment of t . P_w , P_d , and P_t are all used to adjust the forecast, where P_w represents the load ratio factor for the week prior to the forecast. P_d represents the mean load ratio factor for the forecast week, and P_t represents the load ratio factor at the time of the forecast. Figure 5 displays the PS model's STLF structure.

Figure 5 Structure of STLF model for PS (see online version for colours)

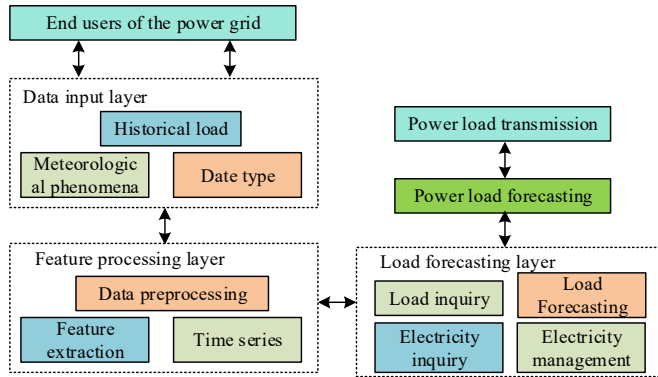


Figure 5 shows that the structure of STLF of PS model mainly includes data input layer, feature processing layer and LF layer. The data input layer collects multi-source information such as historical load data, meteorological data, and date types. These data enter the feature processing layer and undergo pre-processing such as data cleaning and normalisation. Then key features are extracted and important factors related to load are identified through GCA and factor analysis. At the LF layer, the study introduces long short-term memory network (LSTM), random forest (RF) and support vector machine (SVM) for prediction. The model is trained according to the processed feature data,

learns the change rule of load, and finally outputs the LF value, which provides a decision-making basis for the operation and planning of the PS. Among them, LSTM is a 2-layer neural network and has 128 neurons in each layer, using Adam optimiser with a learning rate of 0.001, a batch size of 32, and a loss function of mean square error (MSE). Meanwhile, an RF model containing 100 decision trees was constructed. The maximum depth of each tree was set to 10, the minimum number of sample splits was 2, and the minimum number of sample leaf nodes was 1. The SVM used the radial basis function as the kernel function, with its parameter γ set to 0.1 and the value of C set to 1.0, and used grid search combined with cross-validation to optimise the parameters. The predictions of LSTM, RF and SVM were weighted and averaged using integrated learning to obtain the final predictions.

On this basis, the study selects MSE and mean absolute error (MAE) as the cost function of the prediction model. MSE has high sensitivity to larger errors and can effectively penalise larger deviations in the model, thus improving the overall fitting accuracy of the model. The formula of MSE is shown in equation (10).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

In equation (10), y_i denotes the true value (TV); $\hat{y}_i$ denotes the predicted value; and n denotes the sample size. Electricity load data are usually volatile, and large errors may have a significant impact on the model performance. The sensitivity of MSE to large errors makes it an ideal choice for handling such data. However, to ensure the robustness of the model, the study further introduces MAE to avoid over-penalising certain outliers. The MAE calculation formula is shown in equation (11).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

In real-world applications, predictive models need to perform well in different scenarios. The combined use of MSE and MAE can provide a more comprehensive performance evaluation and optimise the model to maintain high prediction accuracy in various scenarios.

2.3 Analysis of model computational complexity

The main computational steps of GCA include data normalisation and correlation calculation. Among them, the complexity of data normalisation is $O(n)$, where n is the number of data points. The complexity of relevance calculation is $O(m \times n)$, where m is the number of comparison series. Therefore, the overall computational complexity of GCA is $O(m \times n)$. Since GCA does not require a large amount of sample data and has strong processing ability for small sample data, it has high efficiency in practical applications.

The main computational steps of factor analysis include the computation of the correlation coefficient matrix, the extraction of eigenvalues and eigenvectors, and factor rotation. The computational complexity of the correlation coefficient matrix is $O(p^2 \times n)$, where p is the number of variables. The complexity of eigenvalue and eigenvector

extraction is $O(p^3)$. The complexity of factor rotation is $O(p^2)$. Therefore, the overall computational complexity of factor analysis is $O(p^2 \times n + p^3)$.

Combining the above, the synthesised model proposed in the study is mainly composed of two parts: the GCA is used to initially screen the key factors, and its computational complexity is $O(m \times n)$; the factor analysis is used to further extract the synthesised factors, and its computational complexity is $O(p^2 \times n + p^3)$. The total computational complexity of the model is $O(m \times n + p^2 \times n + p^3)$.

3 Effectiveness analysis of STLTF of PS techniques

3.1 Evaluation experiments of STLTF of PS models

To analyse the effectiveness of the research-designed STLTF of PS model based on GCA and factor analysis, the study collects historical load data of a power company system in one year in 2023 and performs outlier detection and normalisation on the load data to obtain the dataset used for the experiment. The dataset covers the historical load data of the power grid of a large city in eastern China during the past year, including hourly load records, meteorological data such as temperature, humidity, wind speed, and date types such as weekdays, weekends, and holidays. The time range of the dataset is from January 2023 to December 2023, covering the load changes during different seasons and special events. Two training sets (TrSs) and two test sets (TeSs) make up the dataset. Table 1 displays the settings for the experimental environment.

Table 1 Experimental environment setup

<i>Parameter variables</i>	<i>Parameter selection</i>
Data pre-processing tool	OpenRefine
Software	Jupyter Notebook
Memory	16G
CPU dominant frequency	2.62 Hz
Data analysis library	Scikit-learn
Operating system	Windows10
Central processing unit	Intel i5-13400
Operating environment	MATLAB2018b

The experimental time is set to 50 minutes from the current moment and the accuracy of the prediction model is evaluated using MAE and root mean square error (RMSE). The two datasets are predicted using the STLTF of PS model and the obtained MAE is shown in Figure 6.

In Figure 6, the MAE of both the TrS and TeS increases to some extent as the prediction time increases. In Figure 6(a), the MAE of the two TrS data increases rapidly with increasing time initially. When the predicted time point reaches 35 min, the MAE stabilises until the end of the experiment. In this case, the difference between the MAEs of the two TrSs at each time point is small, and the error of TrS 1 is 0.01 higher than that of TrS 2. The combined MAE of the two TrSs is 0.09, which is in the acceptable range interval. In Figure 6(b), the MAE trend of the two TeSs is similar to that of the TrS, which is stabilised at 30 min. Moreover, the error curves of the two TeSs are more close

to each other, and the difference at each time point is less than 0.01. The combined MAE of the two TeSs is 0.08. In summary, the prediction model is able to carry out STLTF of PS more accurately to some extent. The error trends of the TrS and TeS are similar, indicating that the model has better stability and generalisation ability. Meanwhile, the error gap between the TrSs and between the TeSs is small, indicating that the model has more consistent prediction performance for different data samples. The RMSEs of the predictions of the TrS and the TeS are shown in Figure 7.

Figure 6 Prediction MAE, (a) MAE of training set, (b) MAE of the test set (see online version for colours)

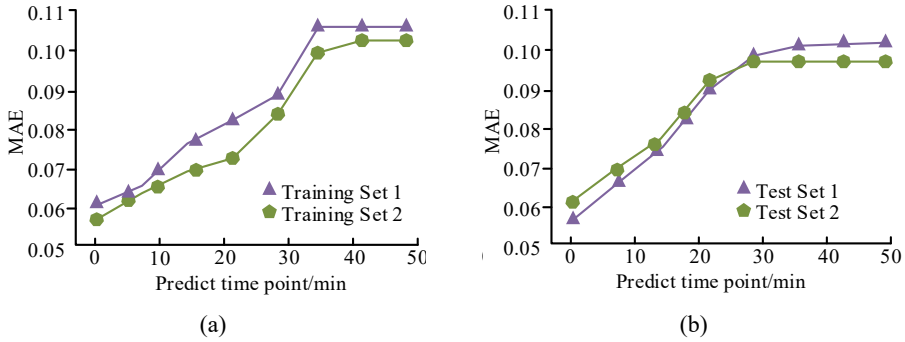
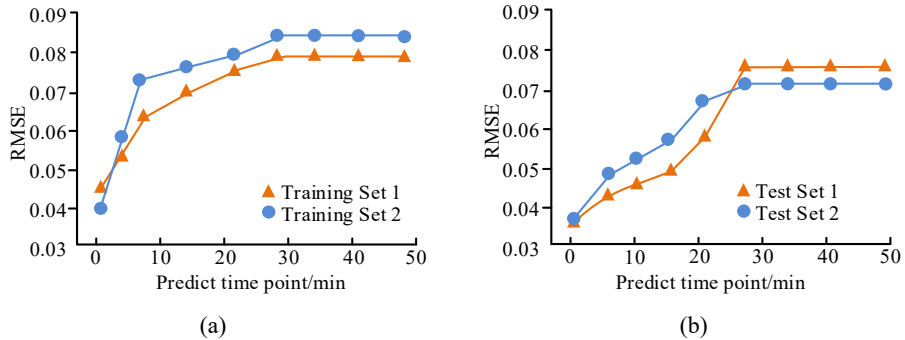


Figure 7 RMSE of prediction, (a) RMSE of training set, (b) RMSE of the test set (see online version for colours)



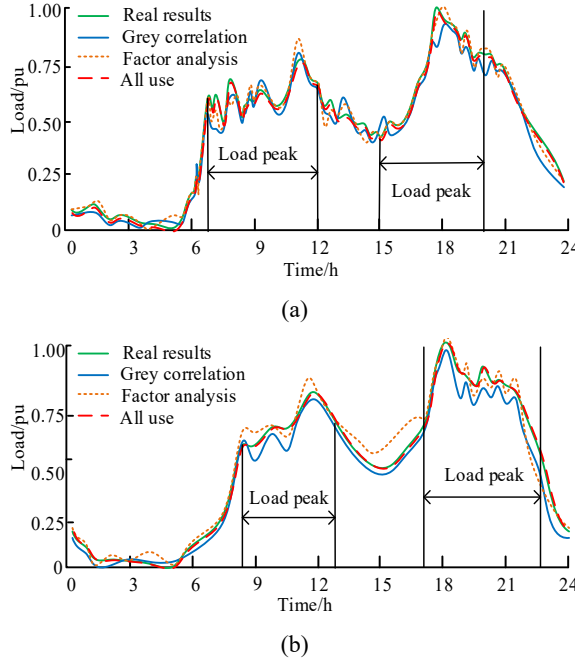
In Figure 7, the RMSEs of both the TrS and TeS show an increasing trend as the prediction time keeps increasing. In Figure 7(a), the RMSE values of the two TrSs increase rapidly in the prediction time from 0 to 10 min, and then slow down slightly until the time reaches 30 min when it almost no longer changes. The final RMSE difference between the two TrSs is no more than 0.01, and the average RMSE value is 0.07. In Figure 7(b), the increasing trend of RMSE in the TeS is similar to that of the TrS. However, at the prediction time point of 30 min, the RMSE value stops increasing. The final average RMSE is 0.06, which is a decrease of 0.01 from the TrS. Overall, the prediction model has better stability and reliability for STLTF of PS to some extent. The difference in RMSE between the TrSs is very small, indicating that the model performs consistently for different training samples. The increasing trend of RMSE in the TeS is

similar to that in the TrS and the final average RMSE is slightly lower than that in the TrS. This suggests that the model can better adjust to the prediction of unknown data and has some generalisation potential.

3.2 Comparative experiments on STLF of PS models

The study offers a thorough evaluation of the STLF of PS models' correctness using factor analysis and GCA. To further explore in depth the effect of different analysis methods on STLF of PS, the study is analysed using comparative experiments. LF is compared with the real results using GCA only, factor analysis only and both methods respectively. The role and value of GCA and factor analysis in LF is understood by comparing the difference in performance of different methods with the same dataset and forecasting time period. The comparison of daily forecasting results under different date types is shown in Figure 8.

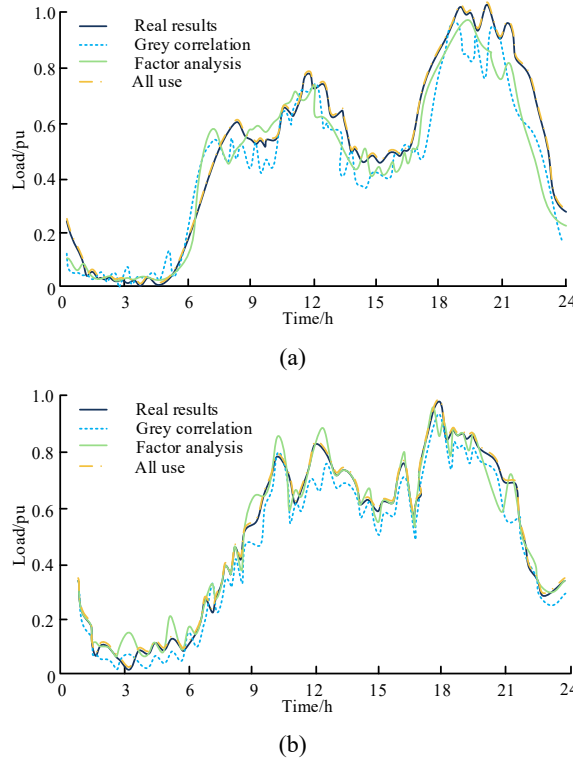
Figure 8 Comparison of daily forecast results under different date types, (a) forecast results for working days, (b) forecast results for festival and holiday (see online version for colours)



In Figure 8, the load forecasts for weekdays and holidays from the different methods show significant differences. In Figure 8(a), on weekdays, the building's electrical demand is typically higher in the middle of the day and the evening, with a longer duration from morning to midday. The predictions obtained using GCA only and factor analysis only are 3.5% lower versus 2.6% higher than the true data, respectively. The prediction results when both methods are used differ from the true by only 0.4%. In Figure 8(b), during holidays, the building's electrical load is highest in the evening. The prediction results using only one method are 5.2% and 6.4% lower and higher,

respectively, and the prediction results when both are used differ from the TV by 0.5%. The results show that the combination of the two methods can better capture the changing pattern of electricity load and reduce the prediction error than using only a single method. A comparison of the daily prediction results for different building types is shown in Figure 9.

Figure 9 Daily forecast results for different types of buildings, (a) residential LF, (b) LF for commercial areas (see online version for colours)



In Figure 9, the electrical loads in residential and commercial areas have different results under different methods of prediction. From Figure 9(a), it can be seen that the peak loads in the residential area are within six hours in the evening. The predicted value obtained under GCA is 8.5% lower than the TV, while factor analysis is 4.4% higher, and the difference between the predicted value and the TV is less than 0.6% when both methods are used. This may be due to the fact that residential loads are affected by a variety of factors such as weather, seasons, holidays, etc. GCA underestimates the impact of certain factors on loads when dealing with these complex factors, resulting in low predicted values. While factor analysis overestimated the impact of certain factors when extracting the complex factors, resulting in high predicted values. In Figure 9(b), the predicted loadings differ from the TVs by 0.5% when both methods are used. The PV for GCA is then 4.2% lower and factor analysis method is 5.5% higher. The results show that for both residential and commercial areas, the deviation between predicted and TVs is large when using a single method for prediction, which indicates that a single method has some limitations in dealing with complex load changes. However, the deviation between

predicted and TVs is significantly reduced when using GCA and factor analysis in combination, which indicates that the integrated method can better adapt to different load conditions and improve the stability and reliability of the predictions. This suggests that the method exhibits high prediction accuracy and stability under different loading conditions. The results of daily prediction for different seasons are shown in Figure 10.

Figure 10 Daily forecast results in different seasons (see online version for colours)

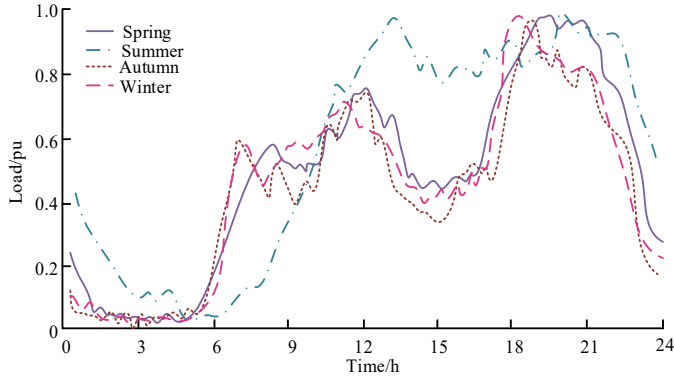
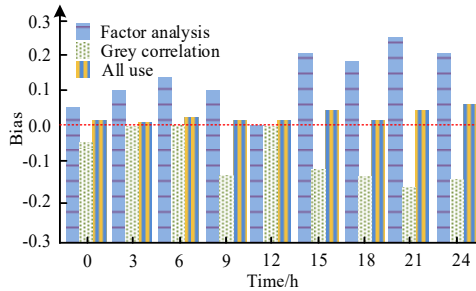
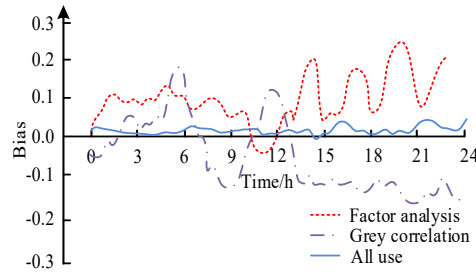


Figure 11 Prediction bias of each method, (a) prediction deviation column chart, (b) prediction deviation curve chart (see online version for colours)



(a)



(b)

In Figure 10, in the spring, the building's electrical load is relatively smooth, with peak hours concentrated during the daytime working hours. In summer, due to rising temperatures and increased use of air conditioners and other electrical appliances, the

power load rises significantly, and the peak load is usually in the evening hours. In the fall, power loads gradually fall back to a more stable state. In winter, due to the use of heating equipment, the power load rises again, and the power load of the building peaks in the morning and evening hours. The findings demonstrate that the forecast model can more accurately capture how the electric load varies with the seasons. The prediction deviation of each method is shown in Figure 11.

In Figure 11(a), the deviation of the obtained prediction results is usually positive when only factor analysis is used; indicating that factor analysis tends to overestimate the electricity load to some extent. In Figure 11(b), the bias of the forecast results obtained when only GCA is used is usually negative, indicating that GCA underestimates electricity load. The lower bias exhibited when both are used indicates that the combination of the two methods improves the accuracy of the predictions. On this basis, the study introduced a DNN-based prediction method proposed in Panda and Ray (2022) Panda et al. and a hybrid prediction method proposed in Chen et al. (2022) Chen et al. for performance comparison. Real data under three weather conditions, namely, sunny, cloudy and heavy rain, were selected for the prediction performance comparison. The details are shown in Table 2.

Table 2 Comparison of PS prediction effect of different models for different weather

<i>Weather</i>	<i>Average load (MW)</i>	<i>Predictive accuracy (%)</i>		
		<i>Panda et al.</i>	<i>Chen et al.</i>	<i>This paper</i>
Sunny	120.53	98.54	99.88	99.97
Cloudy	115.24	99.21	98.56	99.94
Heavy rain	82.13	98.13	98.92	99.95

As can be seen from Table 2, the accuracy of STLF prediction by the proposed method of the study is more than 99.94% for all the three weather conditions. The prediction accuracies under stormy weather are increased by 1.85% and 1.04%, respectively, compared to the methods proposed by other scholars. This may be due to the lower prediction accuracy of the DNN method proposed by Panda et al. when dealing with load changes under extreme weather conditions, which is limited by the complexity of the model structure and the limitation of the training data. And the hybrid prediction method proposed by Chen et al. may not be able to adequately adapt to rapid load changes under extreme weather due to over-reliance on a single optimisation algorithm. Overall, the model proposed in the study can consider the combined effects of multiple factors more comprehensively, improve the quality and availability of data through effective data pre-processing and feature extraction, and thus improve the model prediction efficiency.

4 Discussion and conclusions

To address the problem of STLF of PS, the study explored techniques based on GCA and factor analysis. Among them, to identify the factors that are highly correlated with the load accurately, GCA was used to determine the degree of correlation between different variables and the load. Combined with grey analysis to design the prediction circuit, then GVMS was combined with GCA to provide data support to the methodology. Meanwhile, the factor analysis method was developed in order to minimise the

dimensionality of the data and uncover the main driving elements. Factor analysis is used to reduce the dimension of data and reduce the number of input features, thereby reducing the sensitivity of the model to changes in input parameters. In addition, regularisation techniques are introduced during the model training process to limit the size of model parameters and prevent the model from overfitting the training data. GCA can identify the key factors that have the greatest impact on load changes, and are particularly suitable for processing small sample data and uncertain information (Tao et al., 2023). By combining GCA with factor analysis, the model can not only identify key factors more accurately, but also reduce the complexity of data through dimensionality reduction, thereby improving the prediction accuracy and stability of the model. It is worth mentioning that the proposed model refers to cutting-edge research in related fields during the design and optimisation process. For example, the hybrid algorithm proposed by Chen et al. (2022) demonstrated the potential of hybrid methods in improving prediction accuracy, which promoted the practice of combining factor analysis to optimise data processing and feature extraction. The cited research not only provides a theoretical basis, but also helps identify the shortcomings of existing methods and inspires innovative ideas in model design, so that the proposed model shows higher prediction accuracy and stability under different weather conditions.

The results revealed that in terms of forecasting performance, the MAE and RMSE of the model in different scenarios were 0.08 and 0.06, respectively, which were within the acceptable range. For the weekday period, the prediction results using only one method were 3.5% lower and 2.6% higher, respectively. The difference between the forecast and the TV when both methods are used was only 0.4%. During the holiday period, the forecasts using only one method were 5.2% lower and 6.4% higher, respectively, and when both methods were used, they differed from the TV by 0.5%. The structure demonstrates that the model is far superior to a single method in terms of accuracy and stability, and has a better integrated strategy for LF. The results show that the proposed model effectively solves the sensitivity problems of input parameter changes and data quality, enhances the generalisation ability and credibility of the model, and shows significant superiority in dealing with complex PS LF problems. By improving the accuracy of STLF, the proposed model can enable PS operators to plan power resources more accurately and optimise power generation plans, thereby improving the operating efficiency and stability of the power grid. In addition, accurate LF helps to reduce the demand for spare capacity, reduce operating costs, and improve economic benefits. In terms of PS management, the proposed model can reduce the system's spare capacity demand caused by forecast errors; reduce power generation costs and operating costs. Accurate forecast information can also support the integration of renewable energy, optimise the energy structure of the power grid, reduce carbon emissions, and promote the sustainable development of PS.

However, the study did not consider the impact of special accidents such as power outages and equipment failures on LF. In the future, the study will consider introducing fault detection and recovery mechanisms, combining real-time monitoring data, and dynamically adjusting the prediction model. Machine learning algorithms will be used to classify and predict fault data, warn of potential fault risks in advance, and introduce corresponding adjustment factors into the model to improve the adaptability and robustness of the model under special accident conditions.

Declarations

The author declares that they have no conflicts of interest.

References

- Candan, G. and Cengiz Toklu, M. (2022) 'Sustainable industrialization performance evaluation of European Union countries: an integrated spherical fuzzy analytic hierarchy process and grey relational analysis approach', *International Journal of Sustainable Development & World Ecology*, Vol. 29, No. 5, pp.387–400.
- Chen, Z., Jin, T., Zheng, X., Liu, Y., Zhuang, Z. and Mohamed, M.A. (2022) 'An innovative method-based CEEMDAN-IGWO-GRU hybrid algorithm for short-term load forecasting', *Electrical Engineering*, Vol. 104, No. 5, pp.3137–3156.
- Dong, X., Deng, S. and Wang, D. (2022) 'A short-term power load forecasting method based on k-means and SVM', *Journal of Ambient Intelligence and Humanized Computing*, Vol. 13, No. 11, pp.5253–5267.
- Fan, C., Li, Y., Yi, L., Xiao, L., Qu, X. and Ai, Z. (2022a) 'Multi-objective LSTM ensemble model for household short-term load forecasting', *Memetic Computing*, Vol. 14, No. 1, pp.115–132.
- Fan, G.F., Peng, L.L. and Hong, W.C. (2022b) 'Short-term load forecasting based on empirical wavelet transform and random forest', *Electrical Engineering*, Vol. 104, No. 6, pp.4433–4449.
- Gheisari, M., Hamidpour, H., Liu, Y., Saedi, P., Raza, A., Jalili, A., Rokhsati, H. and Amin, R. (2023) 'Data mining techniques for web mining: a survey', *Artificial Intelligence and Applications*, Vol. 1, No. 1, pp.3–10.
- Gunasekaran, J., Sevvel, P., John Solomon, I. and Vasanthe Roy, J. (2024) 'Multi objective optimization of parameters during FSW of AZ80A-AZ31B Mg alloys using grey relational analysis', *Journal of Mechanical Science and Technology*, Vol. 38, No. 9, pp.4971–4982.
- Kavimani, V., Gopal, P.M., Sumesh, K.R. et al. (2022) 'Multi response optimization on machinability of SiC waste fillers reinforced polymer matrix composite using Taguchi's coupled grey relational analysis', *Silicon*, Vol. 14, No. 1, pp.65–73.
- Kondaiah, V.Y., Saravanan, B., Sanjeevikumar, P. and Khan, B. (2022) 'A review on short-term load forecasting models for micro-grid application', *The Journal of Engineering*, Vol. 2022, No. 7, pp.665–689.
- Lee, Y.G., Oh, J.Y., Kim, D. and Kim, G. (2023) 'Shap value-based feature importance analysis for short-term load forecasting', *Journal of Electrical Engineering & Technology*, Vol. 18, No. 1, pp.579–588.
- Meng, Z., Xie, Y. and Sun, J. (2022) 'Short-term load forecasting using neural attention model based on EMD', *Electrical Engineering*, Vol. 104, No. 3, pp.1857–1866.
- Mokayed, H., Quan, T.Z., Alkhaled, L. and Sivakumar, V. (2023) 'Real-time human detection and counting system using deep learning computer vision techniques', *Artificial Intelligence and Applications*, Vol. 1, No. 4, pp.221–229.
- Panda, S.K. and Ray, P. (2022) 'Analysis and evaluation of two short-term load forecasting techniques', *International Journal of Emerging Electric Power Systems*, Vol. 23, No. 2, pp.183–196.
- Rajesh, M., Sri, M.N.S., Jeyakrishnan, S., Anusha, P., Manikanta, J.E., Sateesh, N. and Vijayakumar, S. (2024) 'Optimization parameters for electro discharge machining on Nimonic 80A alloy using grey relational analysis', *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Vol. 18, No. 3, pp.1429–1442.
- Samuel, B.O., Sumaila, M. and Dan-Asabe, B. (2024) 'Multi-objective optimization and modeling of a natural fiber hybrid reinforced composite (PxGyEz) for wind turbine blade development using grey relational analysis and regression analysis', *Mechanics of Advanced Materials and Structures*, Vol. 31, No. 3, pp.640–658.

- Sun, C., Ning, Y., Shen, D. and Nie, T. (2024) 'Graph neural network-based short term load forecasting with temporal convolution', *Data Science and Engineering*, Vol. 9, No. 2, pp.113–132.
- Tao, R., Ding, P., Peng, R. and Qiao, J. (2023) 'Prediction of asphalt pavement performance based on DEPSO-BP neural network', *Canadian Journal of Civil Engineering*, Vol. 50, No. 8, pp.709–720.
- Tong, C., Zhang, L., Li, H. and Ding, Y. (2022) 'Temporal inception convolutional network based on multi-head attention for ultra-short-term load forecasting', *IET Generation, Transmission & Distribution*, Vol. 16, No. 8, pp.1680–1696.
- Veeramsetty, V., Reddy, K.R., Santhosh, M., Mohnot, A. and Singal, G. (2022) 'Short-term electric power load forecasting using random forest and gated recurrent unit', *Electrical Engineering*, Vol. 104, No. 1, pp.307–329.
- Yi, S., Liu, H., Chen, T., Zhang, J. and Fan, Y. (2023) 'A deep LSTM-CNN based on self-attention mechanism with input data reduction for short-term load forecasting', *IET Generation, Transmission & Distribution*, Vol. 17, No. 7, pp.1538–1552.
- Zhang, R., Zhang, C. and Yu, M. (2022) 'A similar day based short-term load forecasting method using wavelet transform and LSTM', *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 17, No. 4, pp.506–513.