Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 26, Issues 2, pp. 837–845, DOI 10.12694/scpe.v26i2.4025

CARBON EMISSION PREDICTION AND SENSITIVITY EVALUATION OF VIRTUAL POWER PLANTS BASED ON BIG DATA AND MULTISCALE ANALYSIS

JIE LI[*], ZHOU YANG[†], WENQIAN JIANG[‡], AND JUNTAO PAN [§]

Abstract. In order to address the issue of increased prediction errors in the peak carbon emissions of virtual power plants due to various influencing factors of electricity carbon emissions, the authors propose a study on the prediction and sensitivity evaluation of virtual power plant carbon emissions based on big data and multi-scale analysis. Firstly, it analyzes the original data sequence and cumulative sequence, use grey BP neural network to construct a carbon emission peak prediction model, then it analyzes the factors affecting electricity carbon emissions, and use recursive calculation method to calculate electricity carbon emissions. Then, it compress the model coefficients to zero through a penalty function and filter out significant variables. Based on the adjacency characteristics of carbon emission flow, the node carbon potential is calculated through finite recursion, and iterative training is carried out within the allowable error range to solve the model and obtain the predicted peak carbon emissions of electricity. The experimental results indicate that the prediction results of the designed method under three scenarios of benchmark setting, low-carbon, and enhanced low-carbon are 40 million tons, 390 million tons, and 40 million tons, respectively, which are consistent with the actual results, indicating that the prediction error of this method is lower and the prediction results are more accurate. The method studied by the authors can provide technical support for carbon emission control and improve prediction accuracy.

Key words: Grey BP neural network, Virtual power plant carbon emissions, Peak prediction, Penalty function

1. Introduction. With the continuous improvement of the electricity market trading system, the trading models have become more diversified, presenting a mixed trading model of bilateral and joint venture transactions coexisting [1]. However, the existing carbon emission accounting methods are not applicable to all trading models. For bilateral trading models, especially green bilateral trading, the allocation method of "nearby power supply, proportional sharing" does not meet the "bilateral" characteristics, and the low-carbon benefits of clean energy power plants cannot be accurately allocated to users who sign bilateral transactions with them. The power industry is an indispensable basic industry in modern society. With the continuous increase in energy demand, the impact of carbon emissions on the environment is becoming increasingly significant. The global power industry is accelerating its transformation towards clean and low-carbon development [2]. As an important component of the power industry, the electricity market coordinates and manages the generation, sales, transmission, and customer relationships in the power system.

The carbon emissions from virtual power plants are influenced by a variety of dynamic factors. Accurately forecasting these emissions is crucial for developing effective carbon reduction strategies. By predicting future carbon dioxide emissions of power plants, it is possible to optimize the power generation mix and implement specific energy-related carbon reduction measures. At present, there is relatively little research on predicting carbon emissions from power plants, and many existing achievements rely on the analysis of influencing factors to predict carbon emissions. Due to the multiple influencing factors of carbon emissions prediction in power plants, traditional small power plants lack advanced sensor equipment, making it difficult to obtain data on various factors, resulting in significant limitations in predicting influencing factors. The author aims to predict and evaluate the sensitivity of carbon emissions from virtual power plants through big data and multi-scale analysis methods [3]. The development of big data technology provides the energy industry with massive data and rich analytical tools, which can effectively capture and analyze various data generated during the

[*]Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China.
[†]Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China (Corresponding author, yangz_gxcsg@163.com)
[‡]Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China.
[§]Metering Centre of Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530023, China.

operation of virtual power plants. The multi-scale analysis method can deeply explore the dynamic changes and influencing factors of carbon emissions from different time and spatial scales [4,5,6].

2. Literature Review. As the share of renewable energy generation rises and flexible resources like distributed energy storage and electric vehicles are increasingly integrated, the discrepancy between the power system's load demand and the spatiotemporal characteristics of power output has become more pronounced. Consequently, the virtual power plant (VPP) has been developed. VPPs leverage advanced communication technologies to efficiently aggregate distributed power sources, controllable loads, and energy storage systems, allowing them to flexibly respond to scheduling directives and participate in electricity market transactions. Therefore, how to accurately predict the peak carbon emissions, formulate carbon reduction strategies based on the prediction results, and control the peak carbon emissions in a short period of time have become urgent problems in the development process of modern society. Currently, scholars are studying methods for predicting peak carbon emissions from different perspectives and theories. Yang et al. introduced a flexible carbon emission mechanism that coordinates energy between electric hydrogen equipment, SMR factories, and gas turbines. Within this new framework and carbon emission mechanism, the VPP makes purchasing and sales decisions in both day-ahead and real-time markets while optimizing the operational strategies of its internal components [7]. Xuejin, W. et al. developed a virtual power plant scheduling method utilizing a low-carbon multi-objective two-stage optimization algorithm. Initially, this method identifies the production and consumption levels of various energy sources within virtual power plants, such as wind power, thermal power, and hydropower. Subsequently, aiming to minimize costs and reduce carbon emissions, multi-objective optimization algorithms are employed to allocate and schedule the energy resources within the virtual power plant [8]. Wu, Y.et al. introduced virtual power plants (VPP) and power-to-gas (P2G) technologies to enhance energy integration. They first proposed a VPP structure connected to P2G and developed a physical output model. Then, they constructed a multi-objective operational optimization model for VPPs, considering electrical interconnection, with the dual goals of reducing carbon emissions and achieving economic operation. They also proposed a solution method for this model. Finally, they validated the contributions of P2G, demand response (DR), and gas storage technology (GST) through case studies [9].

In response to the existing problems of the above methods, the author proposes a research on virtual power plant carbon emission prediction and sensitivity evaluation based on big data and multi-scale analysis. It can flexibly set parameters according to actual situations, reduce prediction errors, and improve prediction accuracy.

3. Method.

3.1. Design of multi-scale state monitoring methods for big data. The multi-scale state monitoring method proposed by the author for big data is mainly based on the big data provided by the information system, constructing corresponding state monitoring functions that meet the conditions for state parameter reconstruction, and monitoring the corresponding state of industrial equipment [10]. At the same time, multi-scale monitoring of equipment status is carried out through benchmark models and residual fusion of multi-step information. The flowchart is shown in Figure 3.1.

3.1.1. Data preprocessing. For big data in device information systems, there is a lot of noise and incompleteness. In order to monitor devices more accurately, it is necessary to preprocess big data, which can greatly improve the quality of big data and the efficiency of monitoring [11]. Data preprocessing mainly consists of four steps, namely data cleaning, data integration, data transformation, and data reduction [12]. Among them, data cleaning refers to smoothing big data, identifying and removing noisy data, removing isolated points, and processing missing values in big data accordingly; Data integration mainly involves integrating big data from different sources into the same data storage system according to the same rules, facilitating the subsequent use of big data; Data transformation refers to the appropriate transformation of data forms through methods such as data generalization, smooth aggregation, normalization, etc., in order to facilitate subsequent use; Data specification refers to the specification representation of the obtained dataset, mainly used to ensure the integrity of the data. The data preprocessing was completed through the above process, providing data support for the construction of the following benchmark models [13,14].

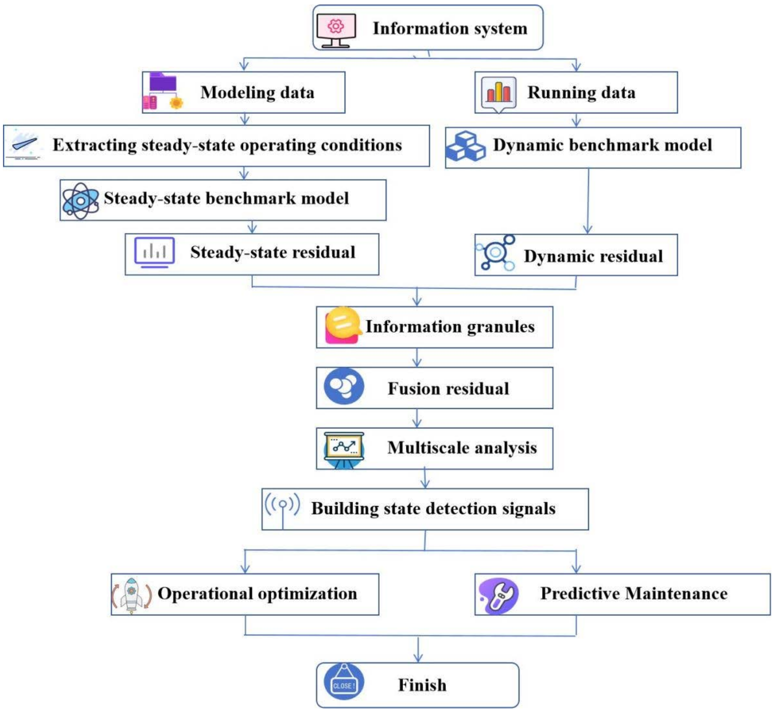


Fig. 3.1: Flow chart of multi-scale state monitoring method for big data

3.1.2. Building benchmark models. Based on the preprocessed data mentioned above, the benchmark model is constructed using deep learning theory. The benchmark model mainly establishes connections between the operating data and design parameters of equipment through certain mechanisms, with the aim of constructing corresponding criteria for establishing industrial equipment parameters and providing reference information for status monitoring. During the operation of the equipment, comparing the reference information provided by the benchmark model with measurable information can construct corresponding residual sequences. If the industrial equipment is working normally, the residual is almost zero; otherwise, it indicates abnormal operation of the equipment. Inconsistencies frequently arise during equipment operation. Consequently, when creating a benchmark model, it is essential to develop both steady-state and dynamic models simultaneously. The steady-state benchmark model is primarily constructed using large datasets of steady-state operating conditions, focusing solely on the parameters during stable equipment operation. In contrast, the dynamic benchmark model is based on extensive data from dynamic operating conditions, reflecting the equipment's performance under varying conditions. During the operation of the equipment, its dynamic working condition data accounts for a large proportion, therefore, this model mainly represents the relationship between this data. To minimize modeling errors resulting from device delays, inertia, and other factors, deep learning techniques are employed to develop the model, thereby enhancing its overall quality [15]. The theoretical model of deep learning is shown in Figure 3.2.

The construction of the benchmark model was completed through the above process, providing reference information for the final state monitoring.

3.2. Construction of prediction model based on grey BP neural network. Carbon emissions are a dynamic process influenced by multiple factors, and artificial neural network models have been proven to have high applicability for predicting carbon emissions. In the construction of artificial neural network carbon emission prediction models, accurately identifying the influencing factors of carbon emissions is the key to improving the accuracy of input layer data in the artificial neural network model, which plays an important role in improving the prediction accuracy of the model. By combining the dynamic variability, nonlinearity, and

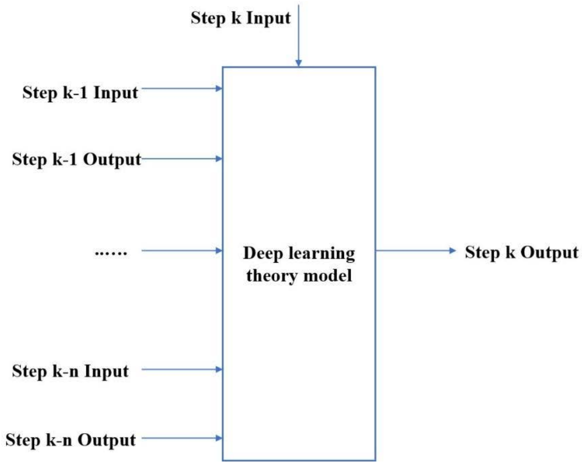


Fig. 3.2: Deep Learning Theory Model

sociality of carbon emissions, optimizing the carbon emission prediction model through parameter optimization, weight initialization, model structure adjustment, learning rate scheduling, and other methods, or constructing a linear nonlinear coupled combination model, the prediction efficiency of the model can be effectively improved. As a data-driven prediction model, the predictive performance of artificial neural network models is affected by initial values, which leads to the network easily falling into local optima and training easily entering overfitting or premature fitting. With the increasing demand for accuracy in carbon emission prediction, further research is needed on the optimization methods of artificial neural network models. The grey BP neural network based on error backpropagation has the advantage of continuously approaching the limit value of the function, which leads to the existence of the optimal solution in the prediction results. In addition, the number of layers, units, and training factors of the model structure can be set according to different environments, making the prediction process more flexible and random, and improving prediction accuracy and efficiency. The grey BP neural network can train a large amount of data and clarify the relationships between the data. By using a fast descent method to continuously adjust the network weights and thresholds through reverse propagation, the error of the network is reduced. The core idea of this method is to introduce a new learning method, which iteratively modifies and trains the network's reverse propagation to ensure that the output vector of the network is consistent with the desired vector [16].

The grey BP neural network model considers random variables as gray variables that vary within a certain interval. After processing, the accumulated sequence shows an exponential growth trend. In this case, the original data sequence and the accumulated sequence are:

$$c^{(0)} = [c^{(0)}(1), c^{(0)}(2), c^{(0)}(3), \cdots, c^{(0)}(n)] \tag{3.1}$$

$$c^{(1)} = [c^{(1)}(1), c^{(1)}(2), c^{(1)}(3), \cdots, c^{(1)}(n)] \tag{3.2}$$

Using the original data sequence and cumulative sequence as input values, a carbon emission peak prediction model based on grey BP neural network is constructed, as shown in Figure 3.3.

Model training is considered complete when the sum of squared errors in the network's output layer falls below a specified threshold. This approach effectively reduces the weighted order deviation within the network [17]. The specific steps of model training are to initialize each node and randomly assign weights and thresholds to each node. After completing the parameter settings, calculate the connection weights and thresholds for the input layer and output layer respectively. Choose the next input method and iterate repeatedly until the network output meets the requirements.

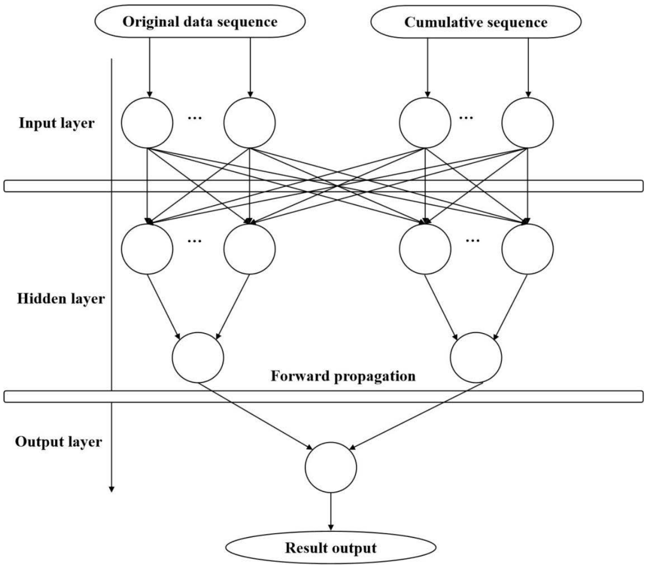


Fig. 3.3: Prediction Model

3.3. Solving the Peak Carbon Emission Prediction Model for Virtual Power Plants. In order to solve the carbon emission peak prediction model for virtual power plants, a recursive algorithm is introduced into the grey BP neural network. This algorithm selects a random carbon emission path, calculates the corresponding impact weights of each factor, and then obtains the corresponding path carbon emission values. The randomly selected new carbon emission path is:

$$L_{i+1} = L_i + \lambda \oplus L'(\xi) \tag{3.3}$$

In the formula, L_i represents the i-th carbon emission path; λ represents the carbon emission coefficient; $L'(\xi)$ represents the carbon emission path extraction function. In the process of calculating the carbon emissions of virtual power plants, the recursive calculation method is used to calculate the carbon emissions of virtual power plants. The formula is:

$$C = \sum_{i=1}^{j} a_i \times f_i \times \lambda_i \tag{3.4}$$

In the formula, a_i represents the i-th electricity consumption mode; f_i represents the i-th power conversion coefficient; λ_i represents the i-th type of electricity carbon emission coefficient; j represents the number of electrical equipment. The recursive algorithm adopts a random selection of carbon emission paths for electricity, mainly by analyzing the influencing factors of carbon emissions on different paths to obtain the maximum and minimum carbon emission cycles. The specific operation process is as follows:

Step 1. Randomly obtain the carbon emissions of each path, and by introducing differential equation parameters and weighting them, a new grey BP neural network model is obtained.

Step 2. Normalization method is used to eliminate non-linear relationships between data and introduce them into the model. The Lasso regression analysis method was used to analyze the impact of various factors on the prediction of peak carbon emissions. Lasso regression analysis compresses the coefficients in the model using a penalty function, turning some factors to 0 to filter out significant variables.

Assuming a as the independent variable and b as the dependent variable, the standard value of the predicted sample obtained after m samples is (a, b), and the kth predicted value of the independent variable a is:

$$x_k = (x_{k1}, x_{k2}, \cdots, x_{km})^T \tag{3.5}$$

In the formula, T represents the prediction period. The regression model of the dependent variable on the independent variable can be expressed as:

$$b_i = \sum_{i=1}^{j} a_i + \epsilon_i \tag{3.6}$$

In the formula, ϵ_i represents a random natural number. If you want to filter out variables that have a significant impact, you need to add a condition to the formula, and the constraint expression is:

$$\begin{cases} arg(t_1, t_2, \cdots, t_t) min||b - ta||^2 \\ s.t. \sum_j \frac{|t|}{\sum t_j^0} \leqslant \phi \end{cases} \tag{3.7}$$

In the formula, t represents the harmonic parameter; ϕ represents the optimal adjustment threshold. Lasso regression is the process of continuously adjusting harmonic parameter values, reducing regression coefficients, compressing variable coefficients until they reach 0, in order to obtain significant variables, namely carbon emission peaks [18].

Step 3. Due to the adjacency of electricity carbon emissions calculation, when calculating the carbon emissions of a node, only the carbon emissions of neighboring nodes need to be obtained, and there is no need to know the carbon emission flow information of that node.

By allocating power to each node, the connections between each node can be obtained. Based on the adjacency of carbon emissions in the power grid, calculate the carbon potential from the initial point to different nodes in sequence. In each iteration, after determining the carbon potential of a certain node, all node carbon potentials can be obtained. Therefore, each iteration can obtain accurate node carbon potential calculation results within any period of time. Finally, the finite recursive method was used to calculate the carbon potential of all nodes in the network. The specific calculation formula is as follows:

$$\sigma_j = \frac{\sum_{i \in \Omega_i} P_i \sigma_i + \sum_{j \in \Omega_j} G_j \sigma_i^G}{\sum_{i \in \Omega_i} P_i + \sum_{j \in \Omega_j} G_j} \tag{3.8}$$

In the formula, P_i represents the active power injected by the node; G_j represents the active power of the power unit branch; Ω_i, Ω_j represents the set of carbon emissions and node injection for the i-th and j-th electricity, respectively. Determine if all nodes have been polled, and if all node carbon potentials have been obtained, complete the recursion.

Step 4. Under the condition of allowable deviation, the prediction model was solved and the peak prediction results of electricity carbon emissions were obtained [19].

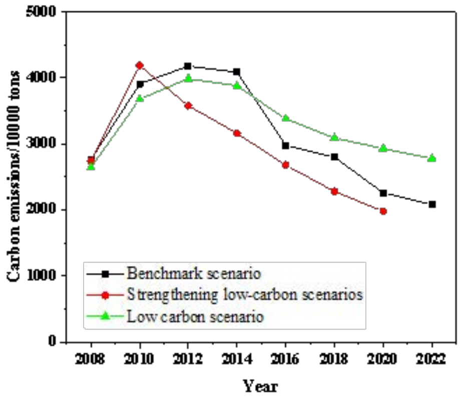
3.4. Experiment. In order to verify the rationality of the author's virtual electric field carbon emission peak prediction method, relevant experiments were designed to verify the feasibility of the method. By analyzing the direction and scale of power exchange between the distribution network and the main network, determine the calculation order of carbon emissions between each major network. If active power is injected into the main grid, it can be used as the main grid power supply, not as the main grid load. If active power is injected into the main grid distribution network, the boundary information between distributed power sources and conventional thermal power plants is used to analyze carbon emissions in the distribution network. Building on this, the carbon potential and injection power of the root node were used as key variables for the main network, and the carbon emissions of the main network were calculated. The distribution of these carbon emissions was then analyzed based on the carbon emission flow within the main network [20].

4. Results and Discussion. The peak of carbon emissions does not mean that the carbon emissions reach their peak in a year, but rather that a stable trend in carbon emissions in a certain region begins from that year. Under the baseline scenario, low-carbon scenario, and enhanced low-carbon scenario, the peak carbon emissions were collected as shown in Table 4.1.

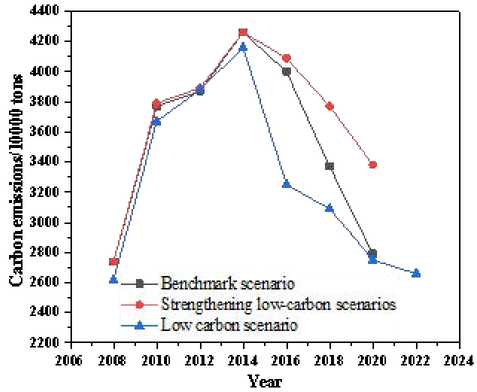
According to Figure 4.1(a), there is a maximum error of 6 million tons between the peak carbon emissions of the STIRPAT model method and the data in Table 4.1 under the baseline scenario; There is a maximum

Table 4.1: Carbon Emission Peak Collection Results

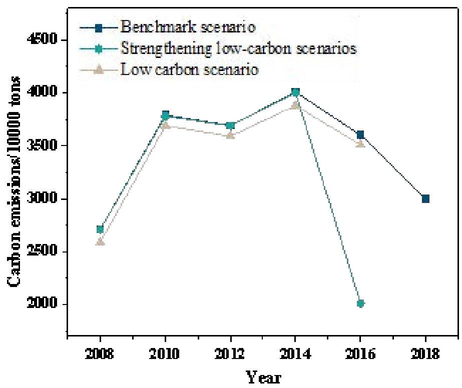
year	Peak carbon emissions/10000 tons		
/year	Benchmark scenario	Low carbon scenario	Strengthening low-carbon scenarios
2008	27001	26001	27001
2010	38001	37001	38001
2012	37001	36001	37001
2014	40001	39001	40001
2016	36001	38001	35001
2018	34001	37001	30001
2020	32001	36001	25001
2022	30001	35001	20001



(a) STIRPAT model method



(b) Double regression prediction model method



(c) Author's research methods

error of 5 million tons between the peak carbon emissions of the STIRPAT model method and the data in Table 4.1 under low-carbon scenarios; There is a maximum error of 4 million tons between the peak carbon emissions of the STIRPAT model method and the data in Table 4.1 under the enhanced low-carbon scenario.

According to Figure 4.1(b), there is a maximum error of 4 million tons between the carbon emission peak of the dual regression prediction model method and the data in Table 1 under the baseline scenario; There is a maximum error of 5 million tons between the carbon emission peak of the dual regression prediction model method in the low-carbon scenario and the data in Table 1; There is a maximum error of 6 million tons between the carbon emission peak of the dual regression prediction model method under the strengthened low-carbon scenario and the data in Table 4.1. According to Figure 4.3(c), it can be seen that the peak carbon emissions under the three scenarios using the research method are consistent with the data in Table 1. Based on the above analysis results, it can be concluded that the peak carbon emissions from electricity using the research method are consistent with actual data, indicating that the prediction results of this method are more accurate.

5. Conclusion. The author proposes a study on virtual power plant carbon emission prediction and sensitivity evaluation based on big data and multi-scale analysis. Due to the non-linear trend of carbon emission changes, the grey BP neural network is used for carbon emission peak prediction. The reason is that the grey BP neural network has good non mapping ability and can flexibly set parameters according to actual situations. The Lasso regression screening method and recursive calculation method are used to solve the model, and the relevant results of electricity carbon emission peak prediction are obtained. The experimental results show that under the baseline scenario, low-carbon scenario, and enhanced low-carbon scenario, the peak carbon emissions of the author's research method are consistent with the actual values, all of which are 40 million tons, 390 million tons, and 40 million tons. The rationality of the author's research method has been verified by the experiment, which can provide technical support for carbon emission control.

6. Acknowledgement. This work was supported by the research project of Guangxi Power Grid Co., Ltd (Project No. GXKJXM20220081)

REFERENCES

[1] Yan, R., Lin, Y., Yu, N., & Wu, Y. (2023). A low-carbon economic dispatch model for electricity market with wind power based on improved ant-lion optimisation algorithm, 8(1), 29-39.
[2] Piao, S., He, Y., Wang, X., & Chen, F. (2022). Estimation of china's terrestrial ecosystem carbon sink: methods, progress and prospects, 65(4), 11.
[3] Sheveleva, G. I. (2022). Corporate governance in generating companies of the russian electric power industry in the context of esg agenda, 5(5), 12.
[4] Leng, K., Li, Z., & Tong, Z. (2022). How will tradable green certificates affect electricity trading markets under renewable portfolio standards?a china perspective, 6(4), 585-598.
[5] Ju, L., Yin, Z., Zhou, Q., Liu, L., Pan, Y., & Tan, Z. (2023). Near-zero carbon stochastic dispatch optimization model for power-to-gas-based virtual power plant considering information gap status theory. International Journal of Climate Change Strategies and Management, 15(2), 105-127.
[6] Su, S., Hu, G., Li, X., Li, X., & Xiong, W. (2023). Electricity-carbon interactive optimal dispatch of multi-virtual power plant considering integrated demand response, 120(10), 2343-2368.
[7] Yang, Z., Li, K., & Chen, J. (2024). Robust scheduling of virtual power plant with power-to-hydrogen considering a flexible carbon emission mechanism. Electric Power Systems Research(Jan.), 6(1), 1-11.
[8] Xuejin, W., Chen, C., Yao, S., & Qiang, C. (2024). Multi-objective two-stage optimization scheduling algorithm for virtual power plants considering low carbon. International Journal of Low-Carbon Technologies, 11(1), 13.
[9] Wu, Y., Wu, J., De, G., & Fan, W. (2022). Research on optimal operation model of virtual electric power plant considering net-zero carbon emission. Sustainability, 47(1), 443-458.
[10] Kuang, Y., Wang, X., Zhao, H., Qian, T., Li, N., & Wang, J., et al. (2023). Model-free demand response scheduling strategy for virtual power plants considering risk attitude of consumers, 9(2), 516-528.
[11] Wang, M., Wang, P., Liang, W. U., Yang, R. P., Feng, X. Z., & Zhao, M. X., et al. (2022). Criteria for assessing carbon emissions peaks at provincial level in china, 13(1), 131-137.
[12] Guo, Q., Xi, X., Yang, S., & Cai, M. (2022). Technology strategies to achieve carbon peak and carbon neutrality for china's metal mines, 29(4), 9.
[13] Liu, E. B., Peng, Y., Peng, S. B., Yu, B., & Chen, Q. K. (2022). Research on low carbon emission optimization operation technology of natural gas pipeline under multi-energy structure, 19(6), 3046-3058.
[14] Abudu, H., Jr, P. K. W., & Lin, B. (2022). Climate pledges versus commitment: are policy actions of middle-east and north african countries consistent with their emissions targets?, 13(4), 612-621.
[15] BAIQinghua, YINXuelian, WANGJing, ZHANGJie, CHUChao, & LIXuejun. (2023). Prediction model of minimum temperature inside solar greenhouse in central hexi corridor based on ridge regression. JOURNAL OF AGRICULTURE, 13(5), 96-100.
[16] FengDONG, ZhaoLI, Ling-HanLI, & Shu-MeiZHANG. (2022). Flow state monitoring of gas-liquid two-phase flow using multiple dynamic kernel principle component analysis. Acta Automatica Sinica, 43(03), 762-773.

[17] Li, J. (2022). Venture financing risk assessment and risk control algorithm for small and medium-sized enterprises in the era of big data. Journal of Intelligent Systems, 31(1), 611-622.
[18] Lee, S. Y., Park, J., & Kim, D. Y. (2023). Context awareness by noise-pattern analysis of a smart factory, 76(8), 1497-1514.
[19] Chaovalitwongse, W. A., Yuan, Y., Zhang, Q., & Liu, J. (2022). Special issue: innovative applications of big data and artificial intelligence, 9(4), 3.
[20] Yuan, X., Deng, H., & Hu, J. (2022). Constructing a ppi network based on deep transfer learning for protein complex detection. IEEJ Transactions on Electrical and Electronic Engineering, 17(3), 436-444.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: May 24, 2024
Accepted: Jul 8, 2024