# Research on Electricity Price Risk Prediction and Settlement Monitoring Data System Based on Big Data Analysis in the Spot Market Environment

Xiaoqiang Xue[a], Xiaolu Wang[a], Shihua Lu[a], Xinrui Zhong[*b]

Emails: Xiaoqiang Xue:xuexq149@163.com; Xiaolu Wang: vvlulu@163.com;
Shihua Lu:lu_shihua@163.com; Xinrui Zhong[*]: zhongxr@tsintergy.com

[a]Jibei Power Trading Center Co., Ltd., No. 56 Caishikou South Street, Xicheng District, Beijing, China;
[b]Beijing Qingneng Interconnect Technology Co., Ltd., Building 8, Liangcheng, Silicon Valley, No.1 Nongda South Road, Haidian District, Beijing, 5001, China

**Abstract:** Compared with the uncertainty in the production and use of electricity, the fluctuation of electricity spot market price is greater, and it is very difficult to master. Moreover, there is no effective way to predict the price risk, and then the price risk prediction and settlement monitoring have a greater correlation with the relevant participants in the market. At the same time, the reliable electricity price risk prediction and settlement monitoring data system is the basis of spot market operation and not only price risk control but also unbalanced settlement risk . Therefore, the study on electricity price risk prediction and settlement monitoring data system based on big data analysis in the spot market environment, firstly summarizes the method of electricity price risk prediction based on big data analysis, highlighting the main content of the study; secondly, expounds a electricity price risk prediction system based on big data analysis in the spot environment, so as to complete the study.

**Key words:** spot market environment; big data analysis; electricity price risk forecast; settlement monitoring data; system research

## 1. INTRODUCTION

The spot market mainly includes day-day, intra-day and real-time auxiliary service trading market of electrical energy, regulation and reserve, and the spot market, medium and long-term direct trading market and futures power stock market constitute the modern power market system. At present, the positioning of the domestic spot market is that the spot trading "spot trading news topic", as a supplementary part of the market-oriented power balance mechanism, plays the role of discovering the price, improving the trading varieties and forming full competition. Therefore, in the spot market environment, the research of electricity price risk prediction and settlement monitoring data system based on big data analysis has a certain role in promoting the development of economic benefits of China's power industry, and then reflects a certain research value and significance[1].

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

DOI 10.4108/eai.17-11-2023.2342691

## 2. METHOD OF ELECTRICITY PRICE RISK PREDICTION BASED ON BIG DATA ANALYSIS

For the medium-and long-term and short-term prediction methods of electricity price, it is necessary to adopt different prediction models to analyze them, among which the commonly used methods include exponential smoothing method, ARMA model, neural network, etc., and then the relevant structure diagram can be constructed according to the electricity price prediction situation of big data analysis[2-4], As shown in Figure Figure 1.

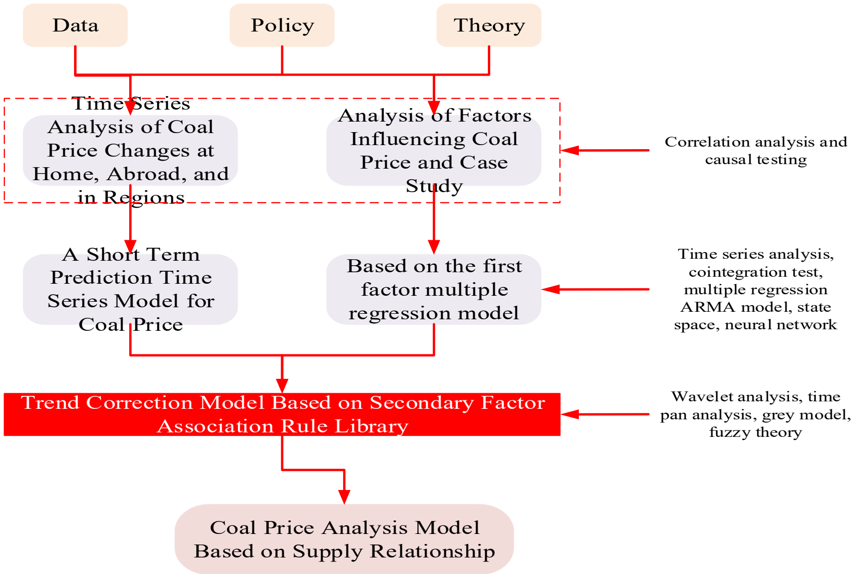


**Figure 1** .Structure diagram of medium-and long-term, day-ahead nodes and weighted average electricity price prediction of the whole network

### 2.1. Short-term prediction method

The short-term prediction method is mainly focused on the short-term electricity price of enterprises. By forecasting it, the commonly used short-term electricity price risk prediction methods mainly include exponential smoothing method and maximum information entropy principle. In addition, the ARMA model is also an important method to study practice sequence data. The short-term prediction methods of this study mainly introduce the first two methods, which are as follows.

1) exponential smoothing method: the method as a conventional and important prediction method of time series, in the actual application process, through the operation of raw data, to get the data information smooth value, according to the calculation content to build the relevant prediction model, finally to predict the future value, judge the risk. Because this method is relatively simple and the learning difficulty is relatively small, the prediction results change relatively greatly, so it is difficult to carry out accurate quantitative prediction, so it can only be used as a tool for preliminary analysis of data.

2) Principle of maximum information entropy: According to the prediction experience of electricity price, most data information is difficult to obtain, and in the offline state, the difficulty is further increased. Therefore, in the absence of some prediction parameters, the principle of maximum information entropy can be used to analyze it by building models. Therefore, the information entropy can be regarded as the probability of the occurrence of a certain information, so that the mathematical significance of the data and the uncertainty of the source are expressed by the formula. The formula is expressed as:

$$H(x) = E[I(x_i)] = E\{\log[1/p(x_i)]\} = -\sum p(x_i)\log[p(x_i)] \tag{1}$$

Where P (xi) is the probability of the i th symbol of the information source; H (x) is the information entropy; i=1,2,3,...,n.

### 2.2. Medium-and long-term prediction methods

The medium and long-term prediction method can be used as a medium and long-term auxiliary power decision, and the effective data basis. Then, this study mainly introduces the state space model and neural network model, and expounds the medium and long-term prediction method as follows.

1) state space model: the model is also known as the dynamic time domain model, is the time as an independent variable of a mathematical model, and through the long-term forecast of prices and economic form, thus in the state space model tube build long-term electricity price change forecast trend graphics, for the enterprise can in the market trading process to provide enough data support.

2) Neural network model: there are relatively many factors affecting the power market, and not all factors are time series. Therefore, when there is a lot of coupling or nonlinear temporal characteristics among the influencing factors, the neural network model can be used to analyze them. For example, the wavelet neural network model based on wavelet analysis has a good function approximation and generalization ability, which can be applied to the electricity price risk prediction.

### 2.3. Precision and range of the prediction methods

The above short-term and medium-and long-term prediction methods have certain advantages and disadvantages, and then different prediction method models also have certain differences in the use range and accuracy, as shown in Table 1 and 2.

**Table 1** .Advantages and disadvantages of short-and medium-term prediction method models

| name | ample | be short of |
|---|---|---|
| exponential smoothing | The smoothing coefficient magnitude is inversely proportional to the action | Actual data analysis is slow to change in response |
| Maximum information entropy principle | Have some of the information | The model algorithm is more complex and is suitable for the short-term |
| state space mode | A wide range of application, which belongs to the internal description | Model building accuracy and other requirements are higher |
| cerebellar model arithmetic computer | Self-study and strong adaptability, fast calculation | A relatively strong dependence on the data |

**Table 2**. Prediction range and progress differences of short-and medium-term prediction methods

| name | Cycle: short-term (-), medium-to-and long-term (+) | Theoretical accuracy:% | the type of transaction Medium-and long-term trading | Spot market recently |
|---|---|---|---|---|
| exponential smoothing | - | More than 90 | × | √ |
| Maximum information entropy principle | - | More than 95 | × | √ |
| state space mode | + | More than 90 | √ | √ |
| cerebellar model arithmetic computer | -、+ | More than 95 | √ | √ |

## 3. AN ELECTRICITY PRICE RISK PREDICTION SYSTEM BASED ON BIG DATA ANALYSIS IN THE SPOT ENVIRONMENT

A price risk prediction system based on big data analysis in the spot environment. The main components include data acquisition module, unit output and transmission line power flow calculation module, sample identification module, the first price calculation module and the second price calculation module[5-6]. The system mainly calculates the key information of the marginal node electricity price by constructing the GP substitution model of the DC optimal power flow model. Then, by solving a set of linear equations, it proposes a risk prediction method of data mixed electricity price, as shown in Figure 2. This method can effectively improve the efficiency of electricity price risk prediction without affecting the electricity price risk prediction.

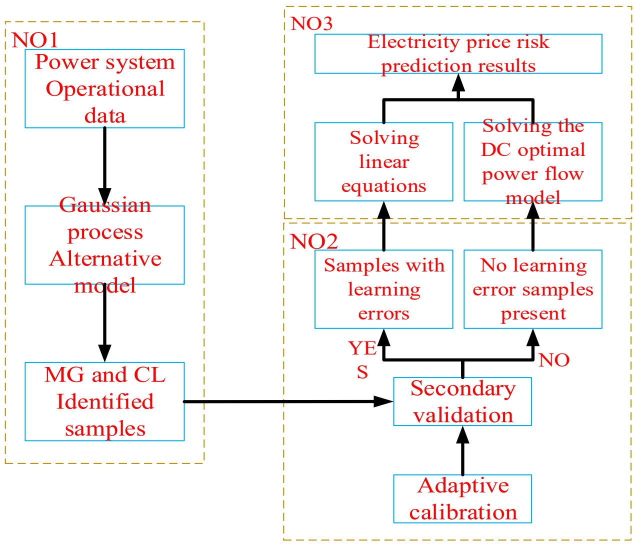


**Figure 2** .Operation diagram of an electricity price risk prediction system based on big data analysis in a spot environment

Combined with the electricity price risk prediction system based on big data analysis, the system is characterized by the optimal DC power-flow Gaussian process substitution model. The formula is expressed as follows:

$$\begin{aligned} &\min \quad c(y) \\ s.t. \quad &Ax + By \geq b, y \in \Omega \end{aligned} \tag{2}$$

Where y is the output, c (·) is the cost function; A is the corresponding matrix of vector x; B is the corresponding matrix of vector y; b is the coefficient; and Ω is the set of outputs.

Output f (x) = [f (x according to formula (2)1),f(x2),...,f(xn)] Follow the distribution, resulting in the following set formula:

$$\begin{bmatrix} f(x_1) \\ \cdots \\ f(x_n) \end{bmatrix} \sim N\left( \begin{bmatrix} m(x_1) \\ \cdots \\ m(x_n) \end{bmatrix}, \begin{bmatrix} C(x_1, x_1) & \cdots & C(x_1, x_n) \\ \vdots & \ddots & \vdots \\ C(x_n, x_1) & \cdots & C(x_n, x_n) \end{bmatrix} \right) \tag{3}$$

Where is x=x1,x2,...,xnFor the output set; m (xn) Is the mean function; C (xn,xn) Is the variance function; thus f (X) | X~N (m (X), C (X, X)), X= (x1,x2,...,xn)TIs a matrix of NP dimension, so get the aboveIn the DC optimal power flow Gaussian process substitution model, the Y obey distribution is expressed as:

$$Y!\ x \sim N\ (m\ (X),\ C\ (X, X) + \sigma^2 In) \tag{4}$$

The I in Equation (4)nFor the unit matrix.

While the parameters y and the initial ytThe joint distribution of! x is expressed as follows:

$$\begin{bmatrix} y \\ y_t \mid x \end{bmatrix} \sim N\left( \begin{bmatrix} m(x) \\ m(x_t) \end{bmatrix}, \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \right) \tag{5}$$

Parameter C in Equation11=C(x,x)+σ²In,C12=C(x,xt),C21=C(xt,x),C22=C(xt,xt)++σ²In; Furthermore, the output yt expected value of the DC optimal tidal current Gaussian process substitution model is expressed as: μ (xt)=m(xt)+C21C11-1(y-m (x)), μ is the expected value.

According to the spot market electricity price risk prediction system based on data analysis in formula (2), the output power characteristics of the boundary unit are expressed as:

$$p_{\min,i} + \varepsilon_{g,i} \leq P_i \leq p_{\max,i} - \varepsilon_{g,i} \tag{6}$$

In the formula, i represents the boundary unit; Pi is the output power of the generator; Pmax, i is the upper limit of power output; Pmin, i is the lower limit of power data; $\varepsilon_{g,i}$ is the relaxation factor of the unit.

According to the spot market electricity price risk prediction system based on big data analysis, the sample identification module verifies the operation data of the power system, unit output and transmission line power flow according to the sample verification criteria. There are two steps. The formula is as follows:

Step 1:

$$N_{MG} + 1 = N_{CL} \tag{7}$$

Step 2: $$\frac{\left|LMP_{\mathrm{i}} - mean\left(LMP_{\mathrm{i}}\right)\right|}{mean\left(LMP_{\mathrm{i}}\right)} \geq P \tag{8}$$

Of the two formulas, NMGFor marginal units; NCLIs the number of blocked transmission lines; mean (LMPi) Is the average node electricity price; p is the proportional threshold; LMPiNonode price; if step 1 is established in the judgment process, it can pass directly without step 2; if step 1 is not valid in the judgment process, step 2 needs to be judged only if the formula of step 2 is satisfied.

According to the spot market price risk prediction system based on big data analysis, the first price calculation model is mainly used for the technology of marginal node price, and the formula is expressed as follows:

$$LMP_i = \partial L / \partial D_i \tag{9}$$

$$\partial L / \partial D_i = \lambda + \sum_{l \in I_L} PTDF_{li}\left(\eta_l^{\min} - \eta_l^{\max}\right) \tag{10}$$

Among the two formulas, L is the Lagrange function of; PTDFli. Power transfer distribution factor from node to line; $\eta_l^{\min}$ is the lower limit of the line flow; $\eta_l^{\max}$ is the upper limit of the line flow; λ Constrained multiplier for power balance; Diis the load demand two; IL is the set of transmission lines. The function of L is represented as:

$$\begin{aligned} L = &\sum_{i \in I_G} c_i P_i - \lambda \sum_{i \in I} \left(P_i - D_i\right) - \sum_{l \in I_L} \eta_l^{\min}\left(\sum_{i \in I} PTDF_{li} \times \left(P_i - D_i\right) + F_l\right) - \sum_{l \in I_L} \eta_l^{\max}\left(-PTDF_{li} \times \left(P_i - D_i\right) + F_l\right) \\ &- \sum_{i \in I_G} \xi_i^{\min}\left(P_i - P_{\min,i}\right) - \sum_{i \in I_G} \xi_i^{\max}\left(P_{\max,i} - P_{\min,i}\right) \end{aligned} \tag{11}$$

In the formula, IG is the set of nodes;is the production cost; Pi is the output power; $\sum_{i \in l} PTDF_{li} \times \left(P_i - D_i\right)$ i is the transmission power of the line; ILis the load demand, and I is the set of system nodes; $\xi_l^{\min}$is the lower limit of the Lagrange multiplier for the generator output constraint; $\xi_l^{\max}$is the upper limit of the Lagrange multiplier for the generator output constraint.

In this system, a GP alternative model is proposed for the DC-OPF problem, which improves the effectiveness of POPF at a lower computational cost. It includes three modules:

Module 1: Training sample generation module. In order to construct the GP alternative model described in (20), for the DC-OPF problem, the training sample set D=(X, Y) can be obtained from the historical operating data of the system operator or by running Monte Carlo simulations. In Monte Carlo simulation, uncertainty vectors ω The DC-OPF output obtained by sampling and calculating for a large number of samples. Here, each row of X is an I-dimensional uncertainty input vector, including the load demand D i of all busbars. The output matrix Y contains columns of IG+IL output variables, which correspond to the unit output P and transmission line power flow PF of each input vector x.

Module 2: Gaussian Process Substitution Model. Based on the training dataset D=[X, Y], we choose the square index SE) covariance kernel function to construct the GP alternative model, with the formula expressed as:

$$C_{SE}(x_k, x_t) = \tau^2 \exp\left( -\frac{(x_k - x_t)^T (x_k - x_t)}{2l^2} \right) \tag{12}$$

By using gradient based optimization algorithms, hyperparameters can be derived while fully establishing the GP substitution model for DC-OPF.

Module 3: Key information identification module for EPRA. For the outputs of GP alternative models (such as P and PF), learning errors are inevitable. In order to identify the key information of EPRA (such as marginal units and blocking lines) and address the impact of learning errors, a relaxation factor was introduced ε。 Then, marginal units and blocked transmission lines can be identified based on the following inequalities.

$$P_{\min,i} + \varepsilon_{g,i} \le P_i \le P_{\max,i} - \varepsilon_{g,i} \tag{13}$$

$$PF_l \ge F_l - \varepsilon_{PF,l} \text{ or } PF_l \le -F_l + \varepsilon_{PF,l} \tag{14}$$

In the formula, i is the marginal unit; Based on the framework proposed in Section 2, EPRA can achieve fast computation without loss of accuracy by blocking transmission lines and obtaining marginal units and blocking transmission lines for each sample.

Example analysis:

In this implementation example, the IEEE30 node system is used for implementation scheme explanation. The following methods will be compared, and the hyperparameter settings of each data-driven method are shown in Table 3. The formula is expressed as:

$$\Delta_{MP} = \frac{\sum_{i \in l_\theta} \sum_{k \in K} \left| \Gamma_{i,k,Mp} - \Gamma_{i,kM0} \right|}{K \cdot p \cdot mean(\Gamma)}, p \in \{1,2,3,4\} \tag{15}$$

In the equation, Γ Output for data-driven methods; P is the dimension of the data-driven method; K is the number of test samples; Mean( Γ) by Γ The mean of.

**Table 3** hyperparameter settings for data-driven methods

| Method (Name) | Set up |
|---|---|
| A1 (based on SAE neural network) | 3 hidden layers, 100 neurons per layer, 200 generations, learning rate=0.0001 |
| A2 (based on SDAE neural network) | 3 hidden layers, 100 neurons per layer, 200 generations, learning rate=0.0001 |
| A3 (based on stacked extreme learning machines) | 500 neurons, 50 reduced hidden nodes, 2 generations selected |
| A4 (based on Gaussian process) | M (x) -0, C (·, ·)=Cse (·, ·), 100 generations selected |

| | |
|---|---|
| A5 (based on this system) | M (x) -0, C (·, ·)=Cse (·, ·), 100 generations selected |

System simulation analysis: Firstly, it has been proven that the evaluation of LMP in the EPRA problem is more complex than the OPF problem, making it difficult to directly use existing data-driven methods for better learning. As shown in Table 4.

**Table 4** Average Error of LMP and DC-OPF in IEEE30 Node System

| method | LMP:% | DC-OPF output error | |
|---|---|---|---|
| | | P:% | PF:% |
| A1 | 6.3 | 2.7 | 5.1 |
| A2 | 6.4 | 2.7 | 5.6 |
| A3 | 6.2 | 1.8 | 2.7 |
| A4 | 7.5 | 2.0 | 3.3 |

From Table 4, it can be seen that due to the discontinuity of LMP, directly learning LMP is a challenge for data-driven methods. In addition, the DC-OPF problem is relatively simple, making the method for learning the output of DC-OPF more reasonable.

Calculation rate comparison: To demonstrate the benefits achieved by the proposed method, we compared the LMP calculation time and error of A1-A5 in the IEEE30 node system, as shown in Table 5.

**Table 5** Average LMP Error of IEEE30 Node System

| Method | Training content | | Test time | LMP error:% |
|---|---|---|---|---|
| | Samples:10000 units | Time | | |
| A1 | 1 | 11.7 | 0.02 | 6.3 |
| A2 | 1 | 21.6 | 0.03 | 6.4 |
| A3 | 1 | 1.0 | 0.34 | 6.2 |
| A4 | 0.02 | 7.5 | 1.38 | 7.5 |
| A5 | 0.02 | 10.4 | 60.9 | 0 |

From Table 5, it can be seen that the relevant device (A5) installed in this system has the highest accuracy, which is exactly the same as the benchmark method. This framework filters out GP learning errors through the identification process and model based adaptive criteria. Compared to the basic standard, the testing time has been reduced by 59.34%. The results show that the system significantly improves the computational efficiency of LMP without losing accuracy. At the same time, for the data-driven learning method of A1-A4, the occurrence of errors is difficult to avoid, even if the number of training samples is increased, the same result will be obtained. Compared to A1-A3, the A5 method has a smaller sample size for training, which meets the requirements of current industrial practice.

## 4. CONCLUSION

In summary, the research on electricity price risk prediction and settlement monitoring data system based on big data analysis in the spot market environment. The developed electricity price risk prediction and settlement monitoring data system is mainly used to handle the uncertainty caused by the full load of the spot market and fluctuations in renewable energy. The probability optimal flow can comprehensively consider various uncertainties in the spot market, Thus, it becomes an effective tool for evaluating LMP in the deregulated electricity market. At the same time, the system operator has solved the optimal economic power generation problem for variable load and renewable energy while meeting power generation and transmission restrictions. In addition, due to the relatively large impact of electricity prices in the spot market, and the emergence of new risk content with market changes, future research of the same type will focus on other new electricity price risk prediction perspectives, and use previous research as a foundation to provide insights into potential areas for future improvement and further research.

## REFERENCES

[1] Wang Junxiu, Dai Zeyu. Research on electricity price prediction and electricity charge risk control analysis based on multi-dimensional Lean Management [J]. Electrical technology, 2021 (16): 135-136,145.

[2] Lu Chengyu, Jiang Ting, Deng Hui, et al. Bidding strategy and income distribution of clean energy power generators in the spot market based on cooperative game [J]. Electric Power Construction, 2020,41 (12): 152-160.

[3] Liao W ,Yang Y ,Wang Q , et al.Interpretable Hybrid Experiment Learning-Based Simulation Analysis of Power System Planning under the Spot Market Environment[J].Energies,2023,16(12):

[4] Irfan M ,Raza A ,2 , et al.Week Ahead Electricity Power and Price Forecasting Using Improved DenseNet-121 Method[J].Computers, Materials Continua,2022,72(3):4249-4265.

[5] Li K,Stephen W,Paul D, et al. Bayesian estimation of electricity price risk with a multi-factor mixture of densities[J]. Quantitative Finance,2022,22(8).

[6] Three rounds of land people, Mori ki ki. Using BSO to learn the LASSO-GRBFN [J]. Journal C (Journal of Electronics, Information and system Department); Journal of Electrical Society. C; Paper esis of Electrical Society. C,2023,143(2):125-132.