Power Spot Market Supply and Demand Forecasting and Indicator Anomaly Detection Method Integrating Autoformer and Bayesian Optimization

Yin Ma[1], Jian Cui[1], Miaojing Zhang[1], Dandan Liu[1], Jie Han[1], Kaili Han[2]

[1]State Grid Gansu Electric Power Company, Lanzhou 730000, Gansu, China
[2]State Grid Gansu Electric Power Research Institute, Lanzhou 730000, Gansu, China

Corresponding author: Yin Ma (e-mail: mayinlanzhou@163.com)

ABSTRACT Accurate forecasting of electricity spot market supply and demand and timely anomaly detection are essential for intelligent energy management and communication-assisted monitoring systems operating over distributed electromagnetic information networks. To improve prediction accuracy under high-frequency non-stationary fluctuations and accelerate hyperparameter convergence in multi-source feature spaces, this study proposes a power market forecasting and indicator anomaly detection framework integrating Autoformer with particle swarm optimization and Bayesian optimization (PSO-BO). The Autoformer model exploits trend–seasonal decomposition to capture long- and short-term temporal dependencies, while PSO performs global exploration and Bayesian optimization refines local hyperparameters for efficient model tuning. A Bayesian probabilistic anomaly detection model based on a normal-inverse-gamma prior is further established to quantify residual uncertainty and achieve adaptive online warning. Experimental results on a 24-hour rolling forecasting task demonstrate MAE/RMSE values of 1.23/1.75 GW for load prediction and 1.85/2.53 GW for generation prediction, with anomaly detection accuracy reaching 93.87%. The proposed framework exhibits strong sensitivity to rapid market fluctuations and provides an effective solution for realtime intelligent scheduling and secure monitoring. Moreover, its distributed forecasting and adaptive decision mechanisms offer valuable support for communication-enabled energy systems and electromagnetic sensing infrastructures requiring reliable information transmission and operational awareness.

INDEX TERMS electricity spot market, Autoformer, Bayesian optimization, anomaly detection, intelligent energy systems

I. INTRODUCTION

WITH the rapid expansion of renewable energy capacity and the continuous improvement of the electricity spot market mechanism, the uncertainty of power grid operation and load volatility has increased significantly; similarly, the global industry faces heightened volatility driven by the rapid, unpredictable shifts in fashion trends and the fluctuating costs of raw materials, both demanding more robust predictive capabilities [1]. The electricity spot market relies on accurate short-term supply and demand forecasts to support market clearing and dispatch decisions [2,3]. However, the superposition of non-stationary, highfrequency noise and multi-level seasonality makes it difficult for traditional forecasting methods to achieve both accuracy and real-time performance [4,5]. At the same time, anomaly detection for monitoring indicators, such as real-time load, also faces challenges with distribution drift and insufficient threshold adaptability. Delayed early warning responses can lead to amplified scheduling risks and market volatility. Exploring a comprehensive framework that can efficiently capture minute-to-hour fluctuations while also enabling online adaptive tuning and anomaly identification has significant engineering and academic value. The primary goal of this study is to construct an end-to-end framework that can efficiently capture highfrequency minute-to-hour disturbances in the electricity spot market while also enabling online adaptive hyperparameter optimization and accurate anomaly detection. This robust, adaptive capability is equally crucial , which requires real-time sensitivity to high-frequency fluctuations caused by fast-moving consumer preferences and volatile input costs, necessitating immediate, accurate operational adjustments.

This study incorporates multi-source power load, power generation output, and auxiliary meteorological and calendar features into an improved Autoformer model. Trend-seasonal decomposition and a sparse attention mechanism are used to simultaneously extract features at different time scales. Subsequently, a two-stage hybrid strategy of particle swarm optimization (PSO) and Bayesian optimization (BO) is employed to perform a global coarse search and local fine-tuning of key hyperparameters, such as the number of encoding layers, decoding layers, and learning rate. Finally, a normal-inverse gamma Bayesian model is constructed based on historical forecast residuals. Real-time risk quantification is applied to map the forecast residuals into risk scores. The posterior distribution is updated instantly via a sliding window, and risk alerts are generated with a 95% confidence interval threshold. This framework enables dynamic threshold adaptive detection of abnormal fluctuations. Compatible with existing power grid dispatch systems, SCADA (Supervisory Control and Data Acquisition)/EMS (Energy Management System), and the IEC 61850 communication protocol, it can be seamlessly embedded in actual operational environments, ensuring the algorithm's deployability and stability on industrial-grade platforms. Experimental results demonstrate that this method significantly improves load and output forecast accuracy (MAE (Mean Absolute Error) is reduced to 1.23 GW and 1.85 GW, respectively) using two years of electricity spot market data from the Jiangsu region. This represents reductions of 2.92 GW and 3.16 GW compared to the traditional BiGRU. The method achieves stability in high dimensions within approximately 71 epochs. It improves the F1 score for anomaly detection to 84.97%, demonstrating its significant impact on regulating supply and demand in the electricity spot market.

Contributions of this paper: (1) Combining Autoformer with PSO-BO achieves efficient hyperparameter adaptive tuning under high-frequency non-stationary fluctuations in the electricity spot market. (2) Designing a Bayesian probabilistic anomaly detection model based on normal-inverse gamma priors, dynamically updating thresholds to achieve accurate online anomaly warnings. (3) Integrating multi-source highdimensional features and conducting tests in multiple regions, the paper improves the time series feature capture capability and prediction accuracy of supply and demand forecasting through trend-seasonal decomposition and sparse attention mechanism.

II. RELATED WORK

In the field of short-term supply and demand forecasting in the electricity spot market, various methods have attempted to cope with the high volatility and nonlinear characteristics of load and renewable energy output. Jain R proposed the WT-ARIMA (Wavelet Transform-Autoregressive Integrated Moving Average) model to address the intermittent nature of renewable energy. It uses a wavelet transform to separate trends and seasonality to predict the day-ahead market load. Experiments on bus systems have shown that it outperforms the classic ARIMA in terms of MAPE (Mean Absolute Percentage Error). However, it still relies on linear assumptions and is difficult to handle the interaction effects of multi-source high-dimensional features [6]. Wang K et al. combined LSTM (Long Short-Term Memory) with Informer to build an integrated learning framework to capture both short-term correlation and long-term dependency, significantly improving the MSE (Mean Squared Error) and MAE of long-term load forecasting. However, the model structure is complex, and the computational overhead is high [7]. Waheed W et al. used LSTM-RNN to maintain an hourly MAPE of about 1.5% in a noisy environment, indicating that RNN (Recurrent Neural Network) has a certain degree of robustness in load forecasting, but still needs more exploration of hyperparameter tuning and multi-source data fusion [8]. While these approaches have made progress in addressing linear decomposition, multi-layer dependencies, data scarcity, or local correlation, they lack adaptive decomposition of multi-periodic components and fail to efficiently optimize hyperparameters. Deep time series models, combined with automated tuning techniques, are urgently needed to improve the responsiveness and accuracy of short-term forecasts.

In the field of short-term supply and demand forecasting in the electricity spot market, researchers have applied autoformer structures and various optimization algorithms to improve the response to highfrequency fluctuations and accelerate the convergence of hyperparameters. Zhao L et al. proposed an improved autotransformer method for model fusion by integrating the prediction results of multivariate linear regression and LightGBM (Light Gradient Boosting Machine) with Autoformer, which achieved significant improvements in MAE and MSE [9]. The Autoformer decomposition module has difficulty accurately distinguishing between sudden noise and long-term trends under extreme spike events, affecting short-term prediction accuracy. The model itself lacks a built-in hyperparameter adaptation mechanism and requires reliance on experience or additional searches to obtain the optimal configuration, resulting in long deployment costs and tuning cycles. When fusing multi-source heterogeneous features, simple attention splicing cannot fully capture the physical constraint associations between different sources. It is urgent to combine domain knowledge or physical constraints to improve the stability and interpretability of the model. In the field of power indicator anomaly detection, researchers mainly use deep autoencoders and signal processing methods to improve recognition and positioning capabilities. Lin R et al. built an anomaly detection service based on DVAE (Discrete Variational Autoencoder). They improved the robustness of the model through a sliding window update strategy, but there are problems such as high retraining overhead [10]. Constructing a Bayesian probabilistic anomaly detection model based on historical residuals is of great significance for achieving high-precision and low-latency early warning of power load indicators.

III. PSO-BO-AUTOFORMER MODEL CONSTRUCTION AND INDICATOR ANOMALY DETECTION METHOD

A. MULTI-SOURCE INPUT CONSTRUCTION AND FEATURE FUSION

1) Missing Value Interpolation and Outlier Removal

The study samples each data source at a unified hourly scale and indexes load, thermal power, hydropower, wind power, photovoltaic power, meteorological variables, calendar indicators, and electricity prices into a unified index T.

For any sequence x(t) with a missing value at T, linear interpolation is performed using adjacent time-scale points. In outlier processing, the global mean and standard deviation of each sequence are first calculated. If |x(ti)-μ|>3σ, the point is considered an outlier and replaced using the interpolation method.

2) Numerical Normalization and Time Coding

For each feature sequence, Min–Max normalization is used, and the calculation formula is shown in Equation (1).

$$x'(t) = \frac{x(t) - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

In formula (1), $x'(t)$ represents the normalized data, and $x_{\min} = \min_k x(t_k)$ and $x_{\max} = \max_k x(t_k)$. In calendar feature processing, sine-cosine timestamp encoding is used, as shown in formula (2).

$$PE_{2i}(t) = \sin\left(\frac{t}{10000^{2i/d}}\right), PE_{2i+1}(t) = \cos\left(\frac{t}{10000^{2i/d}}\right) \tag{2}$$

a

In formula (2), d represents the encoding dimension, and periodic information is applied in $PE(t)$.

3) Multi-Source Feature Matrix Construction

The preprocessed time series vectors are concatenated at each time t, as shown in formula (3). The entire sequence T forms the input tensor X.

$$\begin{aligned}\hat{x}(t) = \big[&x_{load}(t), P_{th}(t), P_{hyd}(t), P_{wind}(t), P_{pv}(t),\\ &M'(t), E'(t), PE(t)\big]^T\end{aligned} \tag{3}$$

In formula (3), $x_{load}(t)$ represents load; $P_{th}(t)$ represents thermal power; $P_{hyd}(t)$ represents hydropower; $P_{wind}(t)$ represents wind power; $P_{pv}(t)$ represents photovoltaic power; $M'(t)$ represents meteorological variables; $E'(t)$ represents electricity price; and $PE(t)$ represents calendar characteristics.

4) Feature Fusion Layer Design

First transforming the dimension of X, as shown in (4).

$$H^{(0)} = \text{LayerNorm}(XW^{(0)} + b^{(0)}) \tag{4}$$

In formula (4), $W^{(0)}$ represents the weight matrix; $b^{(0)}$ represents the bias; and V represents the output after dimension transformation.

Multi-head attention fusion is applied to $H^{(0)}$ to simultaneously extract temporal and cross-feature dependencies, and each head is concatenated and projected, as shown in formula (5).

$$\begin{cases} head_i = \text{Attention}\left(H^{(0)}W_i^Q, H^{(0)}W_i^K, H^{(0)}W_i^V\right) \\ \text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \\ H^{(att)} = [head_1; \ldots; head_h]W^O \end{cases} \tag{5}$$

In formula (5), W_i^Q, W_i^K, and W_i^V represent the weight matrices for computing queries, keys, and values, respectively. Q, K, and V represent the query matrix, key matrix, and value matrix, respectively. d_k represents the dimension of the key matrix K. $H^{(att)}$ represents the final output of the multi-head attention mechanism, and W^O represents the weight matrix used to fuse the outputs of multiple heads.

After multi-head attention processing, residual connection and layer normalization are performed, and enhanced with the feedforward network, as shown in (6).

$$\begin{cases} H^{(1)} = \text{LayerNorm}\left(H^{(0)} + H^{(att)}\right) \\ H^{(2)} = \text{LayerNorm}\left(H^{(1)} + \text{FFN}(H^{(1)})\right) \end{cases} \tag{6}$$

In formula (6), FFN represents the feedforward network, nd $H^{(2)}$ represents the final feature input. This paper does not introduce any external time series decomposition preprocessing methods before inputting data into Autoformer. Instead, it relies entirely on Autoformer's internal sequence decomposition mechanism to perform end-to-end learning and modeling of trend and fluctuation (seasonal) components. After handling missing values, removing outliers, normalizing, and time encoding, all original multi-source time series are directly constructed into a unified multi-source feature tensor input model. In Autoformer's sequence decomposition module, the trend and seasonal components are adaptively separated for each channel using a moving average-based differentiable decomposition operator. Compared with external decomposition methods such as EMD/VMD, this strategy avoids problems such as decomposition scale selection, mode aliasing, and non-backpropagation of errors, enabling the decomposition process to be co-optimized with the prediction objective function, thereby better serving the improvement of prediction accuracy. The embedded decomposition effectively enhances the modeling ability for the non-stationarity and multi-time scale characteristics of electricity spot market time series while maintaining the simplicity of the model structure. Introducing external decomposition as preprocessing may lead to information redundancy or error accumulation, with limited improvement in the final prediction performance. Therefore, this paper chooses the trend-seasonal joint decomposition mechanism inside Autoformer as the only sequence decomposition method, and models and reconstructs it separately in the subsequent encoding-decoding process.

B. AUTOFORMER MODEL CONSTRUCTION

1) Sequence Decomposition Module

This paper applies an Autoformer model to perform trend-seasonal decomposition on the original power load and power output time series, modeling them separately and reconstructing them in the decoding stage to enhance the ability to capture non-stationary fluctuations [11,12]. For the input fusion feature tensor, a sliding average is applied independently to each channel according to the time window length L to separate the trend term and the seasonal term, as shown in equations (7) and (8).

$$T_{i,j} = \frac{1}{L} \sum_{k=i-\lfloor L/2 \rfloor}^{i+\lfloor L/2 \rfloor} H_{k,j}^{(2)} \tag{7}$$

$$S_{i,j} = H_{i,j}^{(2)} - T_{i,j} \tag{8}$$

In formulas (7) and (8), L represents the window length; $T_{i,j}$ represents the trend component of the jth channel at the i-th moment; and $S_{i,j}$ represents the seasonal component.

2) Encoder

The study conducts parallel processing of the trend matrix T and the seasonal matrix S. First, the query, key, and value are calculated for the decomposed T, as shown in (9).

$$Q = TW^Q, \quad K = TW^K, \quad V = TW^V \tag{9}$$

Sparse attention mask is used to preserve local temporal dependencies, as shown in (10).

$$\mathrm{MaskAttention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + \alpha M\right) V \tag{10}$$

Formula (10) M represents the sparse attention mask, and α represents the scaling factor of the mask matrix.

Executing sparse attention heads in parallel and processing the output concatenation and linear mapping as shown in (11).

$$\begin{cases} head_l = \mathrm{MaskAttention}(Q_l, K_l, V_l), \quad l = 1, \ldots, h \\ SA(T) = [head_1; \ldots; head_h]W^O \end{cases} \tag{11}$$

In formula (11), headl represents the output of the l-th attention head; SA(T) represents the final output of the self-attention mechanism; and WO represents the weight matrix used to fuse the outputs of multiple heads.

Now, performing residual and normalization processing, as shown in formula (12).

$$\tilde{T} = \mathrm{LayerNorm}(T + SA(T)) \tag{12}$$

In formula (12), $\tilde{T}$ represents the result after residual and normalization processing.

The feedforward network consists of two fully connected layers, and after residual and normalization, the formula is shown in (13).

$$\begin{cases} \mathrm{FFN}(x) = W_2\sigma(W_1x + b_1) + b_2 \\ E_T = \mathrm{LayerNorm}\left(\tilde{T} + \mathrm{FFN}(\tilde{T})\right) \end{cases} \tag{13}$$

In formula (13), σ represents ReLU, and ET represents the processed output of the trend term. The same sparse multi-head self-attention and feedforward network are applied to the seasonal term S to obtain ES. The output of the final encoder layer is shown in formula (14).

$$E(N_e) = [E_T \| E_S] \tag{14}$$

In formula (14), Ne represents the number of encoder layers; E(Ne) represents the output of the last encoder layer; and $\|$ represents channel concatenation.

The attention feature map at the encoder stage is shown in Figure 1.

In Figure 1, Figure 1(a) reflects the distribution of trend attention, showing the response intensity of each attention head to the trend component in the time series, showing a concentrated and smooth peak, indicating that the model can effectively capture the long-term change trend of power load; Figure 1(b) shows the seasonal attention characteristics, with more frequent and diverse fluctuations, reflecting the model's sensitivity to periodic and short-term fluctuations. Overall, the two attention maps focus on features at different time scales, which helps the Autoformer to accurately model and predict non-stationary highfrequency fluctuations in the electricity spot market.

3) Decoder

In the decoder, the future sequence is generated in an autoregressive manner, and the encoder features are applied by combining cross-attention [13]. The decoded input sequence is subjected to the same decomposition and mapping as the pre-encoder, obtaining D_T and D_S.

The study uses a self-attention mask to ensure that the decoder only accesses the current and previous moments, as shown in (15).

$$\mathrm{SelfAttnDec}(D_T) = \mathrm{MaskAttention}(D_T, D_T, D_T) \tag{15}$$

In formula (15), DT represents the input sequence of the decoder.

Now, the decoded representation is interacted with the encoder output, as shown in formula (16).

$$\begin{gathered} C_T = \mathrm{softmax}\left(\frac{D_T E_T^T}{\sqrt{d_k}}\right) E_T, \\ C_S = \mathrm{softmax}\left(\frac{D_S E_S^T}{\sqrt{d_k}}\right) E_S \end{gathered} \tag{16}$$

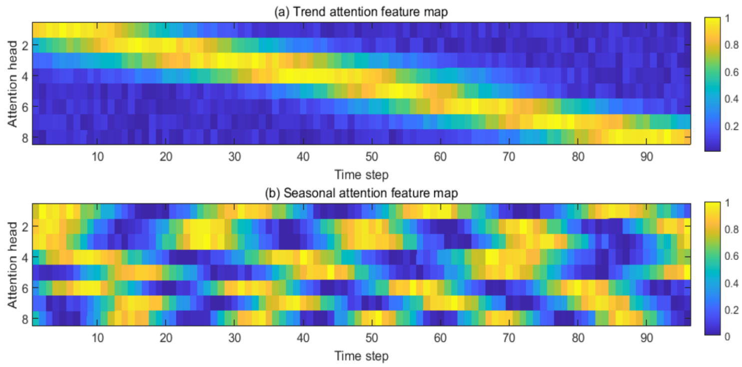

FIGURE 1. (a) Trend Attention Map; Figure 1 (b) Seasonal Attention Map Attention feature map at the encoder dtage

The output of the decoder layer l is shown in formula (17).

$$\begin{cases} \hat{T}^{(l)} = \mathrm{LayerNorm}\left(D_T + \mathrm{SelfAttnDec}(D_T) + C_T\right) \\ \hat{S}^{(l)} = \mathrm{LayerNorm}\left(D_S + \mathrm{SelfAttnDec}(D_S) + C_S\right) \end{cases} \tag{17}$$

In formula (17), $\hat{T}^{(l)}$ and $\hat{S}^{(l)}$ represent the trend and season processing results.

Using the same encoder, applying a feedforward network and normalizing the results, then concatenating the trend and season, as shown in formula (18).

$$\begin{aligned} D(N_d) = \Big[\tilde{T}^{(N_d)} + \mathrm{FFN}\left(\tilde{T}^{(N_d)}\right); \\ \tilde{S}^{(N_d)} + \mathrm{FFN}\left(\tilde{S}^{(N_d)}\right)\Big] \end{aligned} \tag{18}$$

In formula (18), D(Nd) represents the final output of the decoder, and Nd represents the number of decoder layers.

This paper finally maps the final output of the decoder back to the target prediction dimension c through a linear layer, as shown in formula (19).

$$\hat{Y} = D(N_d)W_P + b_P \tag{19}$$

In formula (19), Y represents the power load and power output forecast for the future time. WP represents the weight matrix used for prediction, and bP represents the bias vector used for prediction.

C. AUTOMATIC OPTIMIZATION OF HYPERPARAMETERS OF PSO-BO ALGORITHM

In this paper, the PSO-BO algorithm is used to jointly optimize the hyperparameters of the Autoformer model end-to-end, minimizing the validation set MAE as the fitness function. Specifically, a two-stage hybrid strategy is adopted, first using PSO for a global coarse search, then using BO for a refined local search. The hyperparameter vector is shown in Equation (20).

$$\tilde{x} = [h, N_e, N_d, L_q, \eta, B] \tag{20}$$

In formula (20), $\tilde{x}$ represents the hyperparameter vector; Ne represents the number of encoding layers; Nd represents the number of decoding layers; Lq represents the degree of decoding; η represents the learning rate; and B represents the batch size.

The PSO-BO-Autoformer model structure is shown in Figure 2.

In Figure 2, the PSO-BO algorithm is shown at the top, and the Autoformer model is shown at the bottom. This framework takes multi-source data as input, uses a feature fusion layer to achieve a unified representation of heterogeneous features, and then uses the Autoformer encoder-decoder structure to perform trend-seasonal decomposition and modeling of power load and power output, achieving accurate forecasting of future time series. On this basis, a two-stage joint hyperparameter optimization mechanism is applied. First, a global search is performed through PSO to obtain a high-quality initial hyperparameter set, and then a local fine search is performed with BO. Finally, the optimal hyperparameter configuration is output and fed back to model training to achieve adaptive optimization of end-to-end prediction performance.

1) PSO Global Rough Search

(1) Particle Swarm Initialization Constructing P particles, and initializing the position and velocity of each particle as shown in formula (21) [14,15].

$$\tilde{x}_i^{(0)} \sim U(\tilde{X}), \quad v_i^{(0)} \sim U(-v_{\max}, v_{\max}), \quad i = 1, \ldots, P \tag{21}$$

In formula (21), vmax represents the velocity upper limit vector, and UU represents the uniform distribution.

(2) Fitness Evaluation For each particle, the model is trained on the validation set and the fitness is calculated, as shown in formula (22).

$$f(\tilde{x}_i^{(t)}) = \frac{1}{Mc} \sum_{k=1}^{M} \sum_{j=1}^{c} \left| \hat{Y}_{k,j} - Y_{k,j} \right| \tag{22}$$

In formula (22), M represents the number of verification samples, and $f(\tilde{x}_i^{(t)})$ represents the fitness calculation result.

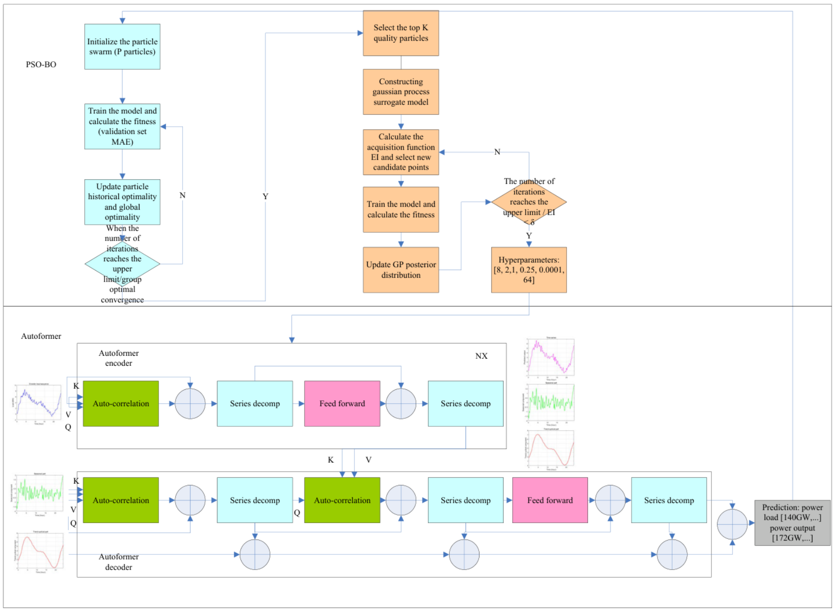

FIGURE 2. PSO-BO-Autoformer model structure

(3) Individual and group updates Recording the historical optimal position and global optimal position of each particle, and the calculation is shown in formula (23).

$$\begin{cases} p_i = \arg\min_{s \le t} f\left(\tilde{x}_i^{(s)}\right) \\ g = \arg\min_i f(p_i) \end{cases} \tag{23}$$

In formula (23), pi represents the historical optimal position, and g represents the global optimal position.

(4) Velocity and Position Iteration For each iteration, the particle velocity and position are updated as shown in formula (24).

$$\begin{cases} v_i^{(t+1)} = \omega v_i^{(t)} + c_1 r_1 \left(p_i - \tilde{x}_i^{(t)}\right) + c_2 r_2 \left(g - \tilde{x}_i^{(t)}\right) \\ \tilde{x}_i^{(t+1)} = \tilde{x}_i^{(t)} + v_i^{(t+1)} \end{cases} \tag{24}$$

In formula (24), ω represents the inertia weight; c1 and c2 represent learning factors; r1 and r2 represent random numbers. $v_i^{(t+1)}$ represents the velocity vector of the i-th particle at the t+1th iteration, and $\tilde{x}_i^{(t+1)}$ represents the position vector of the i-th particle at the t+1th iteration.

2) BO Local Refined Search

(1) Gaussian Process Proxy Model For the initial sample set $D_0 = \{(\tilde{x}, y) | y = f(\tilde{x}), \tilde{x} \in S_0\}$, a Gaussian Process (GP) prior is constructed as shown in

(25) [16].

$$\begin{cases} f(\tilde{x}) \sim GP\left(m(\tilde{x}), k(\tilde{x}, \tilde{x}')\right) \\ k(\tilde{x}, \tilde{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}r}{l}\right) \\ r = \|\tilde{x} - \tilde{x}'\|_2 \end{cases} \tag{25}$$

In formula (25), $m(\tilde{x})$ represents the mean function, which is 0, and the kernel function is Matern 5/2. The hyperparameters σ_f^2 and l are determined by maximum likelihood estimation.

(2) Posterior distribution update For any candidate point, its predicted distribution is shown in formula (26).

$$\begin{cases} f(\tilde{x}^*) | D_t \sim N\left(\mu_t(\tilde{x}^*), \sigma_t^2(\tilde{x}^*)\right) \\ \mu_t(\tilde{x}^*) = k_*^\top \left(K + \sigma_n^2 I\right)^{-1} y \\ \sigma_t^2(\tilde{x}^*) = k(\tilde{x}^*, \tilde{x}^*) - k_*^\top \left(K + \sigma_n^2 I\right)^{-1} k_* \end{cases} \tag{26}$$

In formula (26), σ_t^2 represents the observation noise variance, and $\tilde{x}^*$ represents any candidate point. k_* represents the covariance vector, and K represents the covariance matrix. $k_* = [k(\tilde{x}^*, \tilde{x}_i)]_i$ and $K = [k(\tilde{x}_i, \tilde{x}_j)]_{i,j}$.

(3) Acquisition Function—Expected Improvement (EI)

In the definition of the objective minimization problem, the EI (Expected Improvement) acquisition function is shown in formula (27).

$$\begin{aligned} \alpha_{EI}(\tilde{x}) &= E[\max(y_{\min} - f(\tilde{x}), 0)] \\ &= (y_{\min} - \mu_t(\tilde{x}))\Phi\left(\frac{y_{\min} - \mu_t(\tilde{x})}{\sigma_t(\tilde{x})}\right) \\ &\quad + \sigma_t(\tilde{x})\phi\left(\frac{y_{\min} - \mu_t(\tilde{x})}{\sigma_t(\tilde{x})}\right) \end{aligned} \tag{27}$$

In formula (27), Φ and ϕ represent the cumulative distribution function and probability density function of the standard normal distribution, respectively. $y_{\min}$ represents the current optimal observation value.

(4) BO Iteration For the t-th iteration, performing candidate optimization, solving for $\tilde{x}_{t+1} = \arg\max_{\tilde{x}' \in \tilde{X}} \alpha_{EI}(\tilde{x}')$, evaluating and expanding the dataset, calculating $y_{t+1} = f(\tilde{x}_{t+1})$, and updating $D_{t+1} =$

$D_t \cup \{(\tilde{x}_{t+1}, y_{t+1})\}$. Retraining the GP, reestimating the kernel hyperparameters using D_{t+1}, and updating the posterior. It stops when the number of BO iterations reaches the upper limit or the EI value falls below the threshold δ.

After the PSO phase is completed, the top K optimal particle sets are retained and used D0 as the starting point for the BO phase. Hyperparameters are iteratively refined, and the output x*= arg

$\min_{(x,y)\in D_T^{BO}} y$ is used

as the global optimal hyperparameter configuration for the Autoformer.

D. BAYESIAN PROBABILISTIC ANOMALY DETECTION MODEL

This paper combines time series correlation to model the residual sequence. The definition of the residual vector r(t) is shown in Equation (28).

$$r(t) = Y(t) - \hat{Y}(t) = [r_1(t), r_2(t), \dots, r_c(t)]^T \tag{28}$$

In formula (28), Y(t) represents the true observation value, and Y(t) represents the model output prediction value.

This paper uses the normal-inverse gamma (NIG) conjugate before performing Bayesian modeling of the residual distribution parameters [17,18]. For each dimension, the hypothesis and prior density formula are shown in (29).

$$\begin{cases} r_i|\mu,\sigma^2 \sim N(\mu,\sigma^2), \quad \mu,\sigma^2 \sim NIG(\mu_0,\kappa_0,a_0,b_0) \\ p(\mu,\sigma^2) = N\left(\mu|\mu_0, \dfrac{\sigma^2}{\kappa_0}\right) \times \text{Inv-}\Gamma(\sigma^2|a_0,b_0) \end{cases} \tag{29}$$

In formula (29), μ_0 represents the residual mean prior; κ_0 represents the mean confidence; and a0 and b0 represent the shape and scale parameters of the noise variance, respectively. p(μ,σ2) represents the

prior density. In the prior setting selection, μ_0 is 0, κ_0 is 1, a0 is 2, and $b_0 = \frac{1}{2N_0}\sum_{i=1}^{N_0}(r_i - \bar{r})^2$; and N0 represents the initial warm start residual window size. In residual modeling, this paper chooses the normal-inverse gamma prior as the joint prior for the residual mean and variance, mainly based on three considerations: theoretical consistency, inference resolvability, and robustness in anomaly detection. Assuming the conditional distribution of the residuals is normal, NIG is the standard conjugate prior in the case of unknown mean and variance. Its posterior maintains the same distribution family form, allowing parameter updates to obtain closed-form solutions, thus avoiding instability caused by numerical approximation or sampling, making it suitable for rapid updates in online or sliding window scenarios. Secondly, compared to simpler modeling methods that only set a normal prior for the mean and a fixed constant or independent inverse gamma prior for the variance, NIG can explicitly characterize the coupling relationship between mean uncertainty and variance uncertainty. This is particularly important for residual sequences in the electricity spot market that exhibit heteroscedasticity due to weather, policy, and load fluctuations. The predicted distribution obtained by marginalizing the NIG prior follows a Student's t-distribution. Its heavy-tailed nature naturally improves robustness to extreme residuals and makes it less prone to misclassifying incidental noise as anomalies compared to the pure normal assumption. Furthermore, it offers more reasonable coverage in constructing confidence intervals. Therefore, while maintaining model simplicity and computational efficiency, the normal-inverse gamma conjugate prior provides a more theoretically sound and practically advantageous option for residual uncertainty modeling and anomaly threshold inference.

Under the observed residual data, the parameter posterior is NIG (Inverse Normal Gamma) distributed, and the hyperparameter update formula is shown in (30).

$$\begin{cases} \kappa_n = \kappa_0 + n \\ \mu_n = \dfrac{\kappa_0\mu_0 + n\bar{r}_n}{\kappa_n} \\ a_n = a_0 + \dfrac{n}{2} \\ b_n = b_0 + \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}(r_i - \bar{r}_n)^2 + \dfrac{\kappa_0 n(\bar{r}_n - \mu_0)^2}{2\kappa_n} \end{cases} \tag{30}$$

Formula (30) rn represents the current residual mean.

For the new residual at a certain moment, its predicted distribution is the Student-t distribution. Taking the two-sided critical value corresponding to the confidence level α (95%), the Student-t distribution and the upper and lower thresholds are shown in Formula (31).

$$\begin{cases} r_{n+1}|D_n \sim t_{2a}^{\;n}\left(\text{location} = \mu_n, \text{scale} = \sqrt{\dfrac{b_n(1+\kappa_n^{-1})}{a_n}}\right) \\ L_n = \mu_n - t_{\nu,\alpha/2}\sqrt{\dfrac{b_n(1+\kappa_n^{-1})}{a_n}} \\ U_n = \mu_n + t_{\nu,\alpha/2}\sqrt{\dfrac{b_n(1+\kappa_n^{-1})}{a_n}} \end{cases} \tag{31}$$

(a) Trend Attention Map; Figure 1 (b) Seasonal Attention Map

In formula (31), a represents the confidence level, and $t_{\nu,\alpha/2}$ represents the critical value. If $r_{n+1} \notin [L_n, U_n]$, it is considered an anomaly and an alarm signal is generated.

Under the c prediction target dimensions, each hyperparameter set is maintained and the threshold is calculated independently. Vectorized detection is shown in formula (32).

$$\begin{aligned} d(t_{n+1}) &= [d_1, \dots, d_c]^T, \\ d_j &= \begin{cases} 1, & r_j(t_{n+1}) < L_n^{(j)} \\ & \text{or} \quad r_j(t_{n+1}) > U_n^{(j)} \\ 0, & \text{other} \end{cases} \end{aligned} \tag{32}$$

In formula (32), dj is 1, which means the j-th indicator is abnormal. $L_n^{(j)}$ and $U_n^{(j)}$ represent the upper and lower thresholds, respectively.

The Bayesian residual detection module in this paper first uses the model prediction residuals as direct input, and uses the Student's t prediction interval based on NIG prior to determine "significant residuals" (i.e., large residuals) and trigger an initial alarm. However, it clearly distinguishes between "residual anomalies" (indicating inconsistencies between model predictions and observations) and "abnormal real indicators/market events" (such as abnormal price spikes, transmission line failures, or generator disconnection). To this end, the system performs secondary correlation verification after triggering a residual alarm: it retrieves several exogenous evidences within a ± 1 hour window (sudden changes in time-of-use electricity prices, price jumps in the trading center, generator output/commissioning/retirement alarms, equipment alarm records from SCADA/EMS, and extreme weather reports from the meteorological station, etc.), and classifies the alarms according to a rule-based criterion (if a residual alarm is accompanied by a relative change in time-of-use electricity prices >5% or a sudden decrease in regional net output exceeding X% of planned output/several generator disconnection alarms, it is labeled as "market/operational event"; otherwise, it is labeled as "model/data anomaly". This two-step process (residual detection $\rightarrow$ auxiliary evidence association) retains the advantages of the residual method in terms of real-time performance and sensitivity, while significantly reducing the false alarm rate caused by model limitations or data quality issues through cross-validation of exogenous signals. This allows the output to be used as a rapid early warning for market events, as well as a diagnostic signal for model failure or data problems, triggering operation/schedule or model retraining processes respectively.

E. MODEL TRAINING AND OPTIMIZATION

1) Loss Function Design

The experiment calculates the mean square error for the trend term and the seasonal term separately, and weights them according to the weights, and adds weight decay terms to all trainable parameters, as shown in (33).

$$\begin{cases} \mathcal{L}_{pred} = \dfrac{1}{BMc}\displaystyle\sum_{k=1}^{B}\sum_{i=1}^{M}\sum_{j=1}^{c}\Big[\alpha(\hat{T}_{k,i,j}-T_{k,i,j})^2 \\ \qquad\qquad + \beta(\hat{S}_{k,i,j}-S_{k,i,j})^2\Big] \\ \mathcal{L}_{reg} = \dfrac{\lambda}{2}\left\|\theta\right\|_2^2 \end{cases} \tag{33}$$

In formula (33), α and β represent weight coefficients, Tk,i,j and Sk,i,j represent the decomposed true trend and seasonal components. θ represents the trainable parameter, and λ represents the weight coefficient. The final loss function L is Lpred+Lreg.

2) Optimizer and Gradient Updates

The AdamW optimizer is used for optimization, and the parameter update formula is shown in (34) [19-21]. A linear warmup + cosine annealing strategy is used to optimize the learning rate, and early stopping is performed when the MAE does not decrease for 10 consecutive epochs [22,23].

$$\theta_{t+1} = \theta_t - \eta_t\left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\epsilon} + \lambda\theta_t\right) \tag{34}$$

In formula (34), η_t represents the current learning rate and the value of ϵ is 10^{-8}.

The hyperparameters are shown in Table 1. In Table 1, the parameters in PSO-BO are determined based on experimental tuning.

TABLE 1. Hyperparameters

Hyperparameters	Search space range	Parameter value
Number of attention heads	4, 6, 8, 10	8
Number of encoding layers	1, 2, 3, 4	2
Number of decoding layers	1, 2, 3	1
Decoding degree	0.1, 0.25, 0.5	0.25
Learning rate	10-4 ~ 10-2	0.0001
Batch size	32, 64, 128	64
Number of PSO particles	20 ~ 50	30
Maximum number of PSO iterations	50 ~ 100	80
PSO learning factor c1	1.5 ~ 2.5	2.0
PSO learning factor c2	1.5 ~ 2.5	2.0
PSO inertia weight	0.4 ~ 0.9	0.7
Number of Bayesian optimization iterations	20 ~ 40	30

IV. ELECTRICITY SPOT MARKET SUPPLY AND DEMAND FORECASTING AND INDICATOR ANOMALY DETECTION EXPERIMENT

A. EXPERIMENTAL DATA

The data for this study comes from the State Grid Corporation of China and the Jiangsu Electric Power Dispatching and Control Center, covering electricity spot market operations from January 2023 to December 2024. The data includes power load data and power generation output data (including various power generation methods such as thermal power, hydropower, wind power, and photovoltaic power) within Jiangsu Province, as well as supporting meteorological data, electricity price information, and calendar characteristics. Data is sampled hourly, with a total sample size of 17,520 records (8,760 hours/year $\times$ 2 years). Load and power generation data are sourced from the Jiangsu Electric Power Dispatching Center, covering the total regional load and the actual output of each major generator. Meteorological data is sourced from the Jiangsu Meteorological Observatory of the China Meteorological Administration, covering wind speed, temperature, humidity, radiation, and other indicators. Electricity price data is provided by the Jiangsu Electricity Exchange Center, including time-of-use spot market prices. Calendar features include weekend, holiday, and weekday indicators.

The data used in this paper comes from the State Grid Corporation of China and the Jiangsu Provincial Power Dispatch and Control Center (load and generation output), the Jiangsu Provincial Meteorological Observatory of the China Meteorological Administration (meteorological variables), and the Jiangsu Provincial Power Trading Center (time-of-use pricing). The data covers the period from January 1, 2023 to December 31, 2024, with hourly sampling granularity, for a total sample size of 17,520 hours (8,760 hours/year × 2 years). To ensure temporal consistency and reproducibility, the data is divided chronologically: the first 70% (12,264 hours) of the training set is used for model training, the middle 15% (2,628 hours) of the validation set is used for hyperparameter selection and early shutdown detection, and the last 15% (2,628 hours) of the test set is used for final evaluation. The evaluation of the 24-hour rolling prediction was performed on the test set with a sliding window of 1 hour (each time using the fixed model parameters learned during the training/validation phase and rolling the prediction for the next 24 hours), so the number of independent 24-hour prediction samples used to report performance was 2605.

B. EVALUATION METRICS

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{35}$$

In formula (35), yi represents the true value and $\hat{y}_i$ represents the predicted value.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{36}$$

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{37}$$

C. EXPERIMENTAL PLAN

To comprehensively evaluate the performance of this PSO-BO-Autoformer model in the electricity spot market supply and demand forecasting task, the experiment sets up several representative forecasting benchmark models for comparison, including: SWO (Spider Wasp Optimizer)-Informer, PSO-VMD-TCN- Attention, Informer-LSTM, CNN-LSTM, Autoformer, FEDformer (Frequency Enhanced Decomposed Transformer) and BiGRU (Bidirectional Gated Recurrent Unit). Under the same data set and input length settings, all models uniformly use MSE, MAE, and MAPE as evaluation indicators to perform rolling forecast evaluation of power load and output in the next 24 hours.

In the indicator anomaly detection experiment, a Bayesian probabilistic anomaly detection model based on historical residuals is used as the main model, and compared with the following mainstream detection methods: fixed threshold judgment, Variational Auto Encoder (VAE), One-Class SVM (One-Class Support Vector Machine) and Gaussian Mixture Module (GMM). Each model uses the same prediction residual as input and uniformly evaluates its anomaly recognition ability in load and output sequences using precision, recall, F1-score, and accuracy to reflect the model's adaptability to different anomaly forms.

To verify the model's stability and generalization capabilities, the experiment also includes multi-region migration prediction and multiple missing rate robustness tests. In the migration experiment, the trained model is migrated to non-training provinces for zero-shot prediction. In the missing rate test, feature matrices with different missing rates (1%, 5%, 10%, etc.) are constructed to observe the changes in prediction accuracy of each model under incomplete input information, demonstrating the practicality and robustness of the proposed method from multiple dimensions.

V. SUPPLY AND DEMAND FORECAST AND INDICATOR ANOMALY DETECTION RESULTS

A. SUPPLY AND DEMAND FORECAST CURVE ANALYSIS

The demand-side forecast curve is shown in Figure 3.

In Figure 3, the prediction curves of the PSO-BO-Autoformer model closely match the actual load, accurately capturing peak-to-valley trends with minimal error. During the peak hours (8:00-10:00 AM), the actual load reached approximately 133.66 GW at 8:00 AM and 133.03 GW at 9:00 AM, while the PSO-BO-Autoformer model predicted loads ranging from 133.34 GW to 135.51 GW, respectively, with minimal error. Other models, such as the BiGRU model, exhibited significant overestimation, with the peak prediction at 9:00 AM reaching 152.86 GW. This trend is also evident during the off-peak hours (10:00 PM-12:00 AM), where the PSO-BO- Autoformer predictions (85.93 GW to 90.03 GW) are closer to the actual load (85.95 to 89.87 GW), demonstrating superior robustness.

The supply-side forecast curve is shown in Figure 4.

In Figure 4, the prediction curves of the PSO-BO-Autoformer model closely match the actual output. At 7:00 AM and 6:00 PM, the PSO-BO-Autoformer model's predicted values are 157.49 GW and 128.60 GW, respectively, very close to the actual values of 157.00 GW and 127.48 GW, with relatively small errors. The BiGRU model's predictions at the same time points are 160.93 GW and 136.38 GW, respectively, showing large deviations, reflecting its lower prediction accuracy, especially during periods of high fluctuation.

B. COMPARISON OF SUPPLY AND DEMAND FORECAST ACCURACY

To comprehensively evaluate the performance of our PSO-BO-Autoformer model in electricity spot market supply and demand forecasting, this paper selects several representative models for comparative analysis, measuring the load and generation capacity forecast accuracy using three metrics:

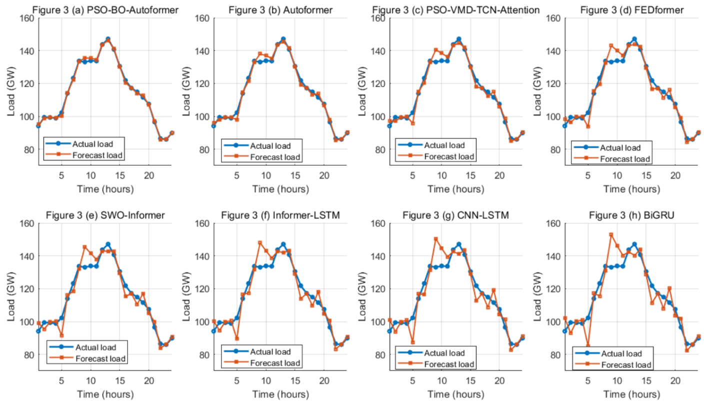

Figure 3. (a) PSO-BO-Autoformer prediction and actual load; Figure 3 (b) Autoformer prediction and actual load
Figure 3 (c) PSO-VMD-TCN-Attention prediction and actual load; Figure 3 (d) FEDformer prediction and actual load
Figure 3 (e) SWO-informer prediction and actual load; Figure 3 (f) Informer-LSTM prediction and actual load
Figure 3 (g) CNN-LSTM prediction and actual load; Figure 3 (h) BiGRU prediction and actual load

FIGURE 3. Demand-side prediction curve

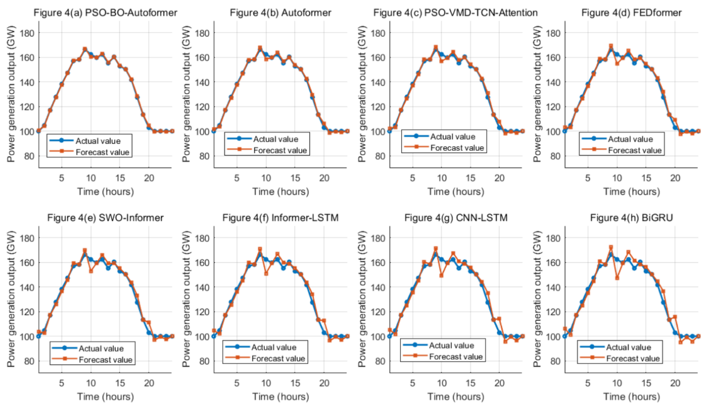

Figure 4. (a) PSO-BO-Autoformer prediction and actual power generation; Figure 4 (b) Autoformer prediction and actual power generation
Figure 4 (c) PSO-VMD-TCN-Attention prediction and actual power generation; Figure 4 (d) FEDformer prediction and actual power generation
Figure 4 (e) SWO-informer prediction and actual power generation; Figure 4 (f) Informer-LSTM prediction and actual power generation
Figure 4 (g) CNN-LSTM prediction and actual power generation; Figure 4 (h) BiGRU prediction and actual power generation

FIGURE 4. Supply-side prediction curve

MAE, RMSE, and MAPE. The results are shown in Figure 5.

In power load forecasting (Figure 5a), PSO-BO-Autoformer is far ahead of other models with a MAE of 1.23 GW, RMSE of 1.75 GW, and MAPE of 0.98%. The Autoformer model also performs well with an MAE of 1.58 GW, an RMSE of 2.12 GW, and an MAPE of 1.27%, but still has gaps of approximately 0.35 GW, 0.37 GW, and 0.29% from PSO-BO-Autoformer, respectively. PSO-VMD-TCN-Attention achieves a MAE of 2.10 GW, an RMSE of 2.85 GW, and a MAPE of 1.68%, surpassing the Autoformer by 0.52 GW and 0.41% respectively. The weakest BiGRU, with a MAE of 4.15 GW, an RMSE of 5.27 GW, and a MAPE of 3.45%, demonstrates its inability to capture spikes and high-frequency fluctuations.

In power generation forecasting (Figure 5b), the overall error of each model is slightly higher than that of load forecasting. PSO-BO-Autoformer achieves an MAE of 1.85 GW, an RMSE of 2.53 GW, and a MAPE of 1.12%. Autoformer achieves an MAE of 2.34 GW, an RMSE of 3.10 GW, and a MAPE of 1.45%, which are worse than PSO-BO-Autoformer by 0.49 GW, 0.57 GW, and 0.33%. PSO-VMD-TCN-Attention achieves a MAE of 2.78 GW, an RMSE of 3.65 GW, and a MAPE of 1.83%. BiGRU performs the worst, with a MAE of 5.01 GW, an RMSE of 6.12 GW, and a MAPE of 3.67%. This indicates that the high volatility of power generation output and the coupling of multiple sources slightly increase the errors of all models, but PSO-BO-Autoformer maintains its lead.

The error magnitudes reported in this article are in gigawatts (GW) (e.g., MAE = 1.23 GW, RMSE = 1.75 GW), equivalent to 1230 MW and 1750 MW respectively; the corresponding power generation prediction errors (e.g., 1.85 GW/2.53 GW) can also be converted to 1850 MW/2530 MW respectively. The reason for reporting in GW is that this article focuses on the total system load and overall

AEM JOURNAL.ORG

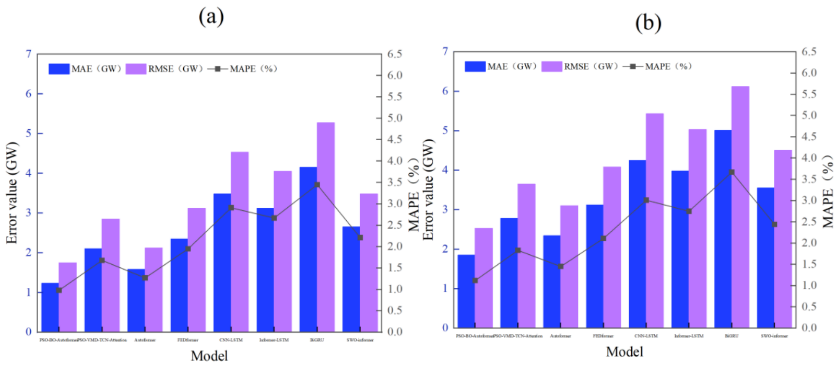

Figure 5. (a) Power load forecast accuracy; Figure 5 (b) Power generation forecast accuracy

FIGURE 5. Comparison of supply and demand forecast accuracy

output of Jiangsu Province (peak values in the ~130 GW range). Using GW at this system scale facilitates intuitive comparison and industry communication. More importantly, looking at absolute quantities (GW) alone can lead to misunderstandings; a more appropriate evaluation should be based on relative errors and the operational context. In this paper, the mean absolute percentage error (MAPE) corresponding to MAE = 1.23 GW is only 0.98%, indicating that the relative magnitude of the prediction deviation on the total system load is less than 1%. This is usually lower than or within the order of magnitude of the dispatch/reserve capacity ratio commonly used in provincial dispatch (usually in the range of "a few percentage points"). Therefore, it is considered practical and acceptable in applications such as system-level dispatch and day-ahead/real-time market clearing.

C. HYPERPARAMETER CONVERGENCE CURVE

The hyperparameter convergence curve is shown in Figure 6.

In the low-dimensional space (Figure 6(a)), PSO quickly takes the lead with an MAE of 0.2243 after 0 iterations, which drops to 0.0979 after 10 iterations. PSO-BO, which has the best accuracy, has the lowest initial MAE (0.1912), but converges slightly slower, reaching 0.0641 after about 20 iterations, and finally stabilizes at around 0.0501 after 80 iterations. Bacterial Foraging Optimization (BFO) has an initial MAE of 0.2977, which only drops to 0.0546 after 80 iterations, consistently lagging behind the other algorithms. GA-PSO and SWO are in the middle, with initial MAEs of 0.2416 and 0.2673, respectively, and ultimately converge to 0.0503 and 0.0502 after 80 iterations.

In the medium dimension (Figure 6(b)), the final MAE values of all algorithms further decrease. PSO-BO converges to approximately 0.0401 after 80 iterations, followed closely by PSO, which reaches 0.04015. GA- PSO starts from 0.2094 and finally reaches 0.0406 after 80 iterations. SWO, WMA (Bacterial Foraging Optimization), and BFO converge to approximately 0.0402, 0.0425, and 0.0460, respectively. In the highdimensional space (Figure 6(c)), the initial MAE of PSO-BO is 0.1203, which drops to 0.0301 after 80 iterations. PSO approaches 0.0302 after 80 iterations, reaching a stable MAE of around 65 epochs. Meanwhile, BFO and WMA maintain MAEs of 0.0416 and 0.0346 after 80 iterations. Overall, PSO-BO achieves optimal results in all dimensions, reaching stability in around 71 epochs. Its final MAE in both medium- and high-dimensional feature spaces approaches the floor.

D. ACCURACY OF INDICATOR ANOMALY DETECTION

The experiment compares the performance of various anomaly detection methods on power spot market load and power output data to evaluate their recognition effects in practical applications. The results are shown in Figure 7.

As shown in Figure 7, the Bayesian probability-based anomaly detection model outperforms other methods across all metrics, achieving 93.87% accuracy, 88.45% precision, 81.76% recall, and 84.97% F1 score. The VAE achieves 90.23% accuracy, 82.56% precision, 72.14% recall, and 76.94% F1 score. The GMM achieves 89.58% accuracy, 78.64% precision, 70.12% recall, and 74.07% F1 score. The fixed threshold method and One-Class SVM achieve F1 scores of 73.05% and 72.53%, respectively, with lower recall scores of only 67.89% and 69.47%. Overall, the Bayesian probability-based anomaly detection model achieves the highest detection accuracy, demonstrating a significant advantage in balancing the reduction of false positives with the improvement of the anomaly detection rate.

E. ANALYSIS OF HIGH-FREQUENCY FLUCTUATION RESPONSE CAPABILITY

The frequent sudden load fluctuations and short-term anomalies in the electricity spot market place extremely high demands on the model's response capability. The experiment evaluates the timeliness of each model in responding to high-frequency non-stationary disturbances from two di-

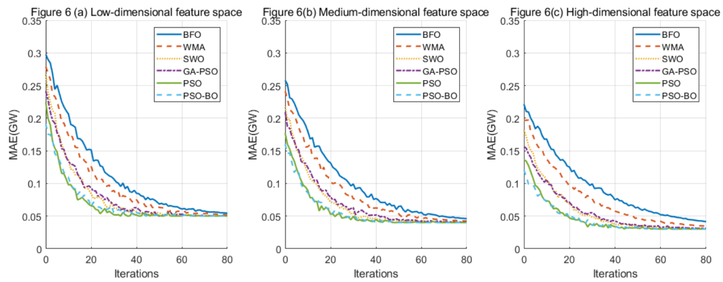

Figure 6. (a) Hyperparameter convergence curve for low-dimensional feature space; Figure 6 (b) Hyperparameter convergence curve for medium-dimensional feature space

Figure 6 (c) Hyperparameter convergence curve for high-dimensional feature space

FIGURE 6. Hyperparameter convergence curve

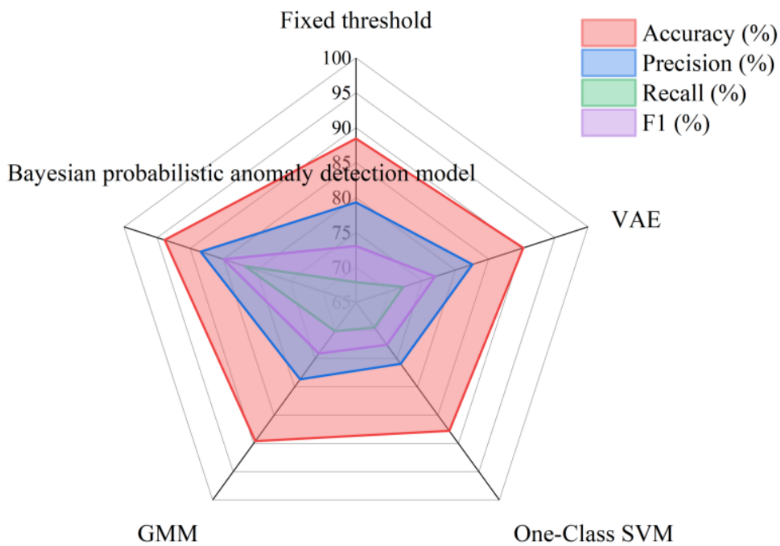

FIGURE 7. Accuracy of indicator anomaly detection

mensions: fluctuation response delay and anomaly detection response time. The results are shown in Figure 8.

In terms of fluctuation response latency shown in Figure 8, PSO-BO-Autoformer leads all models with the lowest latency of 0.46s, while Autoformer reaches 0.59s. Both models can respond to sudden fluctuations in about half a second. Other models, PSO-VMD-TCN-Attention (0.72s), FEDformer (0.78s), SWO-informer (0.85s), and Informer-LSTM (0.91s), all have latency between 0.72s and 0.91s, meeting general scheduling requirements. However, CNN-LSTM (1.08s) and BiGRU (1.23s) have latencies exceeding 1s, indicating a relatively slow response and are slightly insufficient for scenarios with rapid fluctuations at the minute level. In terms of anomaly detection response time, PSO-BO-Autoformer also performed best, issuing an alarm in just 0.31s. Autoformer and PSO-VMD-TCN-Attention achieved delays of 0.38s and 0.49s, respectively, with delays under 0.5s. FEDformer (0.52s), SWO-informer (0.57s), and Informer-LSTM (0.61s) are in the middle. However, CNN-LSTM (0.74s) and BiGRU (0.81s) had detection delays approaching or exceeding 0.7s due to the sequential operations of convolutional or recurrent units in their model structures.

F. REAL-TIME PERFORMANCE AND RESOURCE CONSUMPTION ANALYSIS

The experiment analyzes the model's inference time, memory usage, FLOPs (Floating Point Operations Per Second), parameter size, and power consumption at different input dimensions (20, 40, 60, 80, and 100). The focus is on comparing the resource consumption of each model with a 100-dimensional input, providing a reference for actual deployment. The real-time performance and resource consumption analysis are shown in Table 2.

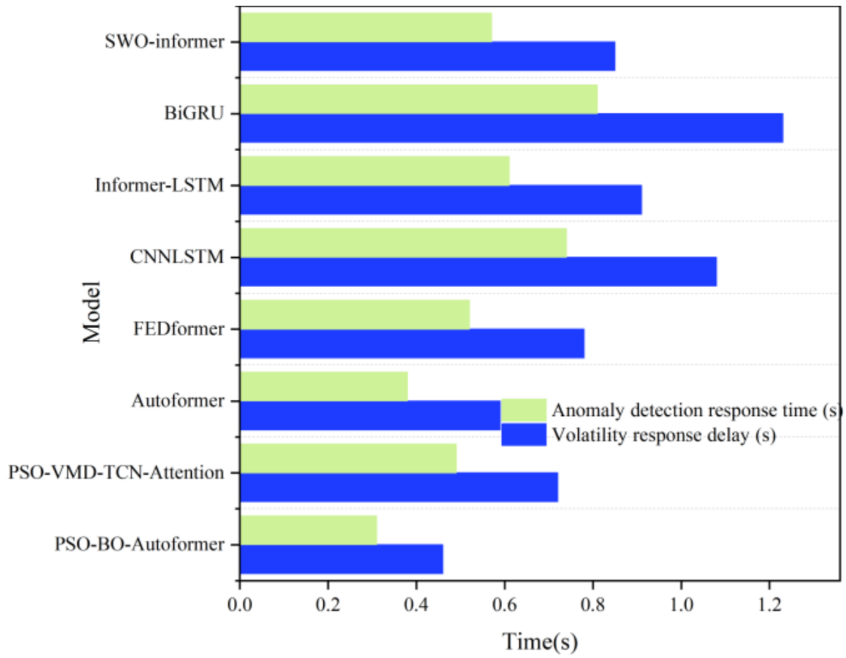

FIGURE 8. High-frequency fluctuation response capability results

TABLE 2. Real-time performance and resource consumption analysis

Dimensions/Model	Inference time (s)	Memory usage (MB)	FLOPs (G)	Parameters (M)	Power consumption (W)
20	0.12	350	12.35	6.85	18.23
40	0.24	420	24.82	6.85	19.04
60	0.37	490	37.65	6.85	19.72
80	0.50	565	49.80	6.85	20.16
100	0.63	630	62.15	6.85	20.58
PSO-BO-Autoformer	0.63	630	62.15	6.85	20.58
SWO-informer	0.57	580	58.12	5.90	19.40
PSO-VMD-TCN-Attention	0.69	700	70.45	7.10	21.85
Informer-LSTM	0.74	715	75.30	7.25	22.30
CNN-LSTM	0.81	745	80.12	7.55	23.10
Autoformer	0.60	600	60.50	6.40	20.10
FEDformer	0.68	675	68.25	6.95	21.50
BiGRU	0.85	760	82.70	7.80	23.75

Table 2 shows a clear linear growth trend in the model's real-time performance and resource usage under different input dimensions. When the dimension increases from 20 to 100, the model's inference time increases from 0.12s to 0.63s, memory usage rises from 350MB to 630MB, and FLOPs climb from 12.35GB to 62.15GB, while the number of parameters remains unchanged (6.85M). Power consumption increases from 18.23W to 20.58W. The trend shows that as the input feature dimension increases, the computational complexity and memory pressure increase nearly linearly, but the number of parameters and basic power consumption change little, reflecting that the fixed scale of the model's main structure is more closely related to the input dimension.

G. SIGNIFICANCE TEST

The experiment conducts a one-way two-sample T-test on the fluctuation response delay and hyperparameter optimization time data, using the PSO-BO-Autoformer and PSO-BO algorithms as the benchmark models. The test steps are as follows:

(1) Null hypothesis H0: There is no significant difference in the mean between the model to be compared and the baseline model (the difference is zero). Alternative hypothesis H1: There is a significant difference in the mean between the model to be compared and the baseline model (the difference is non-zero).

(2) A two-sided T-test is used with a significance level of $\alpha = 0.05$.

(3) If the P value is < 0.05, the null hypothesis is rejected and a significant difference is considered; otherwise, the null hypothesis is accepted.

The T-test results (fluctuation response delay, hyperparameter optimization time) are shown in Table 3.

TABLE 3. T-test results (fluctuation response delay, hyperparameter optimization time)

Model	Mean ± SD (s)	t value	P value	Significance
PSO-BO-Autoformer	0.46 ± 0.04	—	—	—
Autoformer	0.59 ± 0.05	6.28	0.00005	Significant
PSO-VMD-TCN-Attention	0.72 ± 0.07	10.43	<0.00001	Significant
FEDformer	0.78 ± 0.06	11.56	<0.00001	Significant
SWO-informer	0.85 ± 0.08	12.98	<0.00001	Significant
Informer-LSTM	0.91 ± 0.09	14.12	<0.00001	Significant
CNN-LSTM	1.08 ± 0.11	16.75	<0.00001	Significant
BiGRU	1.23 ± 0.12	18.34	<0.00001	Significant
Methods	Mean ± SD (min)	t value	P value	Significance
PSO	18.5 ± 1.6	5.27	0.0004	Significant
PSO-BO	12.3 ± 1.2	—	—	—
SWO	21.6 ± 1.8	7.04	0.0001	Significant
GA-PSO	19.8 ± 1.7	5.81	0.0003	Significant
WMA	24.2 ± 2.1	8.56	0.00001	Significant
BFO	22.7 ± 1.9	7.88	0.00001	Significant

In the significance test of the fluctuation response delay in Table 3, all models except the baseline PSO-BO-

Autoformer (0.46±0.04s) showed highly significant differences (P<0.0001). The average delay of the Autoformer is 0.59±0.05s, with a t value of 6.28 (P=0.00005). Other models, including PSO-VMD-TCN- Attention (0.72±0.07s, t=10.43), FEDformer (0.78±0.06s, t=11.56), and SWO-informer (0.85±0.08s, t=12.98), also significantly outperforms the baseline with P<0.00001. The slowest BiGRU (1.23±0.12s) exhibits the largest response time lag, with t=18.34 and P<0.00001. Overall, the fluctuation responses of all compared models significantly lag behind those of the PSO-BO-Autoformer, demonstrating its superior response efficiency in high-frequency perturbation scenarios.

In a significance test of hyperparameter optimization time, PSO-BO, used as a baseline (12.3 ± 1.2 minutes), also demonstrates significant acceleration compared to other methods. PSO takes 18.5 ± 1.6 minutes, with a t-value of 5.27 (P = 0.0004) compared to the baseline. GA-PSO and BFO took 19.8 ± 1.7 minutes (t = 5.81, P = 0.0003) and 22.7 ± 1.9 minutes (t = 7.88, P = 0.00001), respectively. WMA, the most time-consuming, reaches a whopping 24.2 ± 2.1 minutes (t = 8.56, P = 0.00001). This demonstrates that PSO-BO significantly outperforms other mainstream algorithms in accelerating hyperparameter search.

H. COMPARATIVE EXPERIMENT OF HYPERPARAMETER OPTIMIZATION METHODS

To directly compare with standard methods, this paper designs comparative experiments to quantify the advantages of the PSO-BO hybrid strategy in terms of convergence time and final prediction performance.

The experimental steps are as follows: Under the same dataset and preprocessing procedures as the other experiments above (hourly samples from Jiangsu Province in 2023–2024), hyperparameter search is performed on the high-dimensional configuration (input dimension = 100). The comparison methods are the proposed PSO-BO (two-stage: PSO global + BO local), pure BO (direct search with GP+EI), pure PSO (same stage settings as the PSO in this paper), and dense random search. To ensure comparability, all methods are run under the same candidate evaluation budget (each method is allowed a maximum of 80 model training/validation evaluations). Each method is independently repeated 5 times to evaluate stability, and the following are recorded: total hyperparameter optimization time (minutes), final load prediction MAE (GW), average number of iterations to reach convergence, and standard deviation of MAE in 5 trials. Table 4 presents the comparison results of hyperparameter optimization methods.

TABLE 4. Comparison Results of Hyperparameter Optimization Methods

Methods	Optimization time (min)	Load MAE (GW, mean ± SD)	Average convergence iterations
PSO-BO (This paper)	12.3 ± 1.2	1.23 ± 0.02	71 ± 4
Pure BO (GP+EI)	15.8 ± 1.5	1.31 ± 0.03	80 ± 6
Pure PSO	18.5 ± 1.6	1.28 ± 0.025	65 ± 5
Dense Random Search	22.0 ± 2.0	1.36 ± 0.04	100 ± 8

As shown in Table 4, under the same candidate evaluation budget (maximum 80 model training/validation runs) and high-dimensional input configuration (input dimension 100), different hyperparameter optimization methods exhibit significant differences in convergence efficiency and prediction accuracy. The proposed PSO-BO hybrid optimization strategy outperforms other methods in both optimization time and final prediction performance, with an average optimization time of only 12.3 ± 1.2 minutes, significantly lower than that of pure BO (15.8 ± 1.5 minutes), pure PSO (18.5 ± 1.6 minutes), and dense random search (22.0 ± 2.0 minutes). This demonstrates that the two-stage search mechanism can effectively reduce ineffective evaluation overhead. In terms of prediction accuracy, PSO-BO achieved a load prediction MAE of 1.23 ± 0.02 GW, while pure BO achieved 1.31 ± 0.03 GW and dense random search achieved 1.36 ± 0.04 GW, also outperforming pure PSO's 1.28 ± 0.025 GW. Regarding convergence behavior, pure PSO reached a stable range relatively quickly after 65 ± 5 iterations, but its final accuracy was slightly lower than PSO-BO. Pure BO and random search required 80 ± 6 and 100 ± 8 iterations respectively to converge, demonstrating lower search efficiency in high-dimensional hyperparameter spaces. Overall, PSO-BO, through a synergistic mechanism of "PSO global coarse search + BO local fine search," achieved a better balance between convergence speed, stability, and final prediction accuracy, effectively supporting the rationality and advantage of its 12.3-minute optimization time in high-dimensional scenarios.

I. MULTI-SOURCE FEATURE CONTRIBUTION ANALYSIS AND ABLATION EXPERIMENT

A controlled ablation experiment was designed to quantify the impact of various features on prediction accuracy and high-dimensional hyperparameter optimization behavior. Under the same data preprocessing and time division (hourly data from Jiangsu in 2023–2024, training/validation/test division) and model/search budget as the experimental scheme, five sets of input features were constructed: (A) only historical load and lag terms (Load-only); (B) Load + weather (Temperature/Wind/Radiation, etc.); (C) Load + power generation composition (thermal power/hydropower/wind power/photovoltaic output); (D) Load + price and calendar (time-of-use electricity price, holiday/working day coding); (E) All features (Full, multi-source fusion, the main setting in this paper, input dimension approximately 100). For each feature configuration, the PSO-BO hyperparameter search method (with the same maximum number of candidate evaluations, up to 80 times) was repeated 5 times to evaluate stability. The total time spent on hyperparameter optimization, the final predicted load MAE/RMSE/MAPE (GW or %), and the average number of convergence iterations were recorded. The experimental results of multi-source feature contribution ablation are shown in Table 5.

TABLE 5. Experimental Results of Multi-Source Feature Contribution Ablation

Feature set (description)	Input dimension (≈)	Optimization time (min)	MAE (GW ± SD)	RMSE (GW)	MAPE (%)	Average convergence iterations
A. Load-only (Historical Load/Lag)	5	6.5 ± 0.6	1.85 ± 0.04	2.50	1.45	50 ± 3
B. Load + Weather	15	8.4 ± 0.8	1.45 ± 0.03	1.95	1.15	58 ± 4
C. Load + Generation Mix	30	10.2 ± 0.9	1.33 ± 0.03	1.85	1.05	64 ± 4
D. Load + Price + Calendar	40	11.0 ± 1.0	1.30 ± 0.025	1.80	1.02	66 ± 5
E. Full (All sources, this article)	100	12.3 ± 1.2	1.23 ± 0.02	1.75	0.98	71 ± 4

As shown in Table 5, different feature combinations have a significant impact on prediction performance and high-dimensional optimization behavior. When only historical load and its lag terms (feature set A, input dimension approximately 5) are used, the model has the largest prediction error, with a MAE of 1.85 ± 0.04 GW and a MAPE of 1.45%. Although this model has the shortest optimization time (6.5 ± 0.6 min) and converges in an average of 50 ± 3 iterations, it is insufficient to fully characterize the external driving factors of electricity load. After introducing meteorological features (feature set B, 15 dimensions), the MAE significantly decreases to 1.45 ± 0.03 GW, indicating that variables such as temperature, wind speed, and radiation have significant explanatory power for load fluctuations. Further incorporating power generation composition information (feature set C, 30 dimensions) reduces the MAE to 1.33 ± 0.03 GW and the RMSE to 1.85 GW, showing that supply-side structural information helps improve prediction stability. The feature set D (40 dimensions), incorporating price and calendar factors, further improves performance, achieving MAE and MAPE of 1.30 ± 0.025 GW and 1.02%, respectively, reflecting the complementary effects of time-of-use pricing and holiday effects on load behavior. The complete model using all multi-source features (feature set E, approximately 100 dimensions) achieves optimal results, further reducing MAE to 1.23 ± 0.02 GW, RMSE to 1.75 GW, and MAPE to 0.98%. However, the corresponding hyperparameter optimization time increases to 12.3 ± 1.2 min, with an average of 71 ± 4 convergence iterations. With increasing feature dimensions, prediction accuracy shows a continuous upward trend, while computational cost and the number of convergence iterations also rise. PSO-BO can still achieve stable convergence within an acceptable time in a high-dimensional multi-source feature space, verifying the effectiveness of multi-source feature fusion in improving prediction accuracy and its feasibility in high-dimensional optimization.

VI. CONCLUSION

This paper adopts a method for electricity spot market supply and demand forecasting and indicator anomaly detection that integrates Autoformer and Bayesian optimization, offering a powerful, generalized framework whose focus on non-stationary, multi-source data is also highly beneficial for tackling volatility , such as optimizing raw material procurement against rapid price changes. Multi-scale features are extracted through trend-seasonal decomposition and a sparse attention mechanism. Combined with the PSO- BO algorithm two-stage optimization strategy, the forecasting performance and convergence efficiency are significantly improved. The constructed Bayesian anomaly detection model performs well in terms of accuracy and real-time performance. Experiments show that the model achieves accuracies of 1.23GW/1.75GW and 1.85GW/2.53GW in load and output forecasts, respectively. The convergence times to reach stability in high dimensions are around 71 epochs, and hyperparameter optimization takes only 12.3 minutes. This method achieves good results in terms of accuracy and responsiveness, but there is still room for improvement in terms of generalization across multiple regions and identification of extreme anomalies. In the future, the framework can further apply multimodal perception and transfer learning mechanisms to enhance the model's generalization and robustness, and serve real-time intelligent control of power systems in a wider range of scenarios; additionally, these enhancements will allow the framework to be readily transferred to the industry to optimize dynamic production planning by integrating market data, visual quality control information, and raw material logistics.

AUTHOR CONTRIBUTIONS

Conceptualization – Yin Ma, Jian Cui, Miaojing Zhang, Dandan Liu, Jie Han and Kaili Han; methodology – Yin Ma, Jian Cui, Miaojing Zhang, Dandan Liu, Jie Han and Kaili Han; investigation – Yin Ma, Jian Cui, Miaojing Zhang, Dandan Liu, Jie Han and Kaili Han; writing-original draft preparation – Yin Ma, Jian Cui, Miaojing Zhang, Dandan Liu, Jie Han and Kaili Han. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by,

1. Research and Application of Key Technologies for Monitoring Provincial Bilateral Spot Markets Adapting to High Proportion of New Energy project number#: 522722230062

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] M. Pavlík, F. Kurimsky, and K. Sevc, "Renewable energy and price stability: An analysis of volatility and market shifts in the European electricity sector (2015–2025)," Applied Sciences, vol. 15, no. 12, pp. 6397-6433, 2025, doi: 10.3390/app15126397.

[2] A. Bara, S. V. Oprea, and A. C. Băroiu, "Forecasting the Spot Market Electricity Price with a Long Short-Term Memory Model Architecture in a Disruptive Economic and Geopolitical Context," International Journal of Computational Intelligence Systems, vol. 16, no. 1, pp. 130-151, 2023, doi: 10.1007/s44196-023-00309-3.

[3] B. F. Hobbs, V. Krishnan, J. Zhang, H. F. Hamann, C. Siebenschuh, R. Zhang, et al., "How can probabilistic solar power forecasts be used to lower costs and improve reliability in power spot markets? A review and application to flexiramp requirements," IEEE Open Access Journal of Power and Energy, vol. 9, no. 1, pp. 437-450, 2022, doi: 10.1109/OAJPE.2022.3217909.

[4] S. Deng, J. Inekwe, V. Smirnov, A. Wait, and C. Wang, "Seasonality in deep learning forecasts of electricity imbalance prices," Energy Economics, vol. 137, no. 1, pp. 1-10, 2024, doi: 10.1016/j.eneco.2024.107770.

[5] H. Qiu, R. Hu, J. Chen, and Z. Yuan, "Short-Term Electricity Load Forecasting Based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Improved Sparrow Search Algorithm–Convolutional Neural Network–Bidirectional Long Short-Term Memory Model," Mathematics, vol. 13, no. 5, pp. 813-845, 2025, doi: 10.3390/math13050813.

[6] R. Jain and V. Mahajan, "Load forecasting and risk assessment for energy market with renewable based distributed generation," Renewable Energy Focus, vol. 42, no. 1, pp. 190-205, 2022, doi: 10.1016/j.ref.2022.06.007.

[7] K. Wang, J. Zhang, X. Li, and Y. Zhang, "Long-term power load forecasting using LSTM-informer with ensemble learning," Electronics, vol. 12, no. 10, pp. 2175-2193, 2023, doi: 10.3390/electronics12102175.

[8] W. Waheed, Q. Xu, M. Aurangzeb, S. Iqbal, S. H. Dar, and Z. M. S. Elbarbary, "Empowering data-driven load forecasting by leveraging long short-term memory recurrent neural networks," Heliyon, vol. 10, no. 24, pp. 1-15, 2024, doi: 10.1016/j.heliyon.2024.e40934.

[9] L. Zhao, Z. Wu, and M. Du, "Improved Autuoformer Electricity load forecasting based on model fusion," Advances in Engineering Technology Research, vol. 11, no. 1, pp. 8, 2024, doi: 10.56028/aetr.11.1.8.2024.

[10] R. Lin, S. Chen, Z. He, B. Wu, H. Zou, X. Zhao, et al., "Electricity behavior modeling and anomaly detection services based on a deep variational autoencoder network," Energies, vol. 17, no. 16, pp. 3904-3923, 2024, doi: 10.3390/en17163904.

[11] Z. Wang, Z. Chen, Y. Yang, C. Liu, X. Li, and J. Wu, "A hybrid autoformer framework for electricity demand forecasting," Energy Reports, vol. 9, no. 1, pp. 3800-3812, 2023, doi: 10.1016/j.egyr.2023.02.083.

[12] M. J. Walczewski and H. Wohrle, "Prediction of electricity generation using onshore wind and solar energy in Germany," Energies, vol. 17, no. 4, pp. 844-870, 2024, doi: 10.3390/en17040844.

[13] D. Liu and H. Wang, "Time series analysis model for forecasting unsteady electric load in buildings," Energy and Built Environment, vol. 5, no. 6, pp. 900-910, 2024, doi: 10.1016/j.enbenv.2023.07.003.

[14] L. P. Dai, "Performance analysis of deep learning-based electric load forecasting model with particle swarm optimization," Heliyon, vol. 10, no. 16, pp. 1-20, 2024, doi: 10.1016/j.heliyon.2024.e35273.

[15] Y. Huang, Z. Huang, J. H. Yu, X. Dai, and Y. Li, "Short-term load forecasting based on IPSO-DBiLSTM network with variational mode decomposition and attention mechanism," Applied Intelligence, vol. 53, no. 10, pp. 12701-12718, 2023, doi: 10.1007/s10489-022-04174-z.

[16] S. Zeng, C. Liu, H. Zhang, B. Zhang, and Y. Zhao, "Short-term load forecasting in power systems based on the Prophet–BO–XGBoost model," Energies, vol. 18, no. 2, pp. 227-241, 2025, doi: 10.3390/en18020227.

[17] A. K. M. Nor, S. R. Pedapati, M. Muhammad, and V. Leiva, "Abnormality detection and failure prediction using explainable Bayesian deep learning: Methodology and case study with industrial data," Mathematics, vol. 10, no. 4, pp. 554-590, 2022, doi: 10.3390/math10040554.

[18] X. Zhao, X. Wang, and M. W. Golay, "Bayesian network–based fault diagnostic system for nuclear power plant assets," Nuclear Technology, vol. 209, no. 3, pp. 401-418, 2023, doi: 10.1080/00295450.2022.2142445.

[19] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," Neural Computing and Applications, vol. 35, no. 23, pp. 17095-17112, 2023, doi: 10.1007/s00521-023-08568-z.

[20] S. Sun and H. Gao, "Meta-AdaM: An meta-learned adaptive optimizer with momentum for few-shot learning," Advances in Neural Information Processing Systems, vol. 36, no. 1, pp. 65441-65455, 2023, [Online]. Available: https://doi.org/10.52202/075280-2855.

[21] Y. Hong and J. Lin, "On convergence of adam for stochastic optimization under relaxed assumptions," Advances in Neural Information Processing Systems, vol. 37, no. 1, pp. 10827-10877, 2024, doi: 10.52202/079017-0346.

[22] C. Zhang, Y. Shao, H. Sun, L. Xing, Q. Zhao, and L. Zhang, "The WuC-Adam algorithm based on joint improvement of Warmup and cosine annealing algorithms," Math. Biosci. Eng, vol. 21, no. 1, pp. 1270-1285, 2024, doi: 10.3934/mbe.2024054.

[23] O. V. Johnson, C. Xinying, K. W. Khaw, and M. H. Lee, "ps-CALR: periodic-shift cosine annealing learning rate for deep neural networks," IEEE Access, vol. 11, no. 1, pp. 139171-139186, 2023, doi: 10.1109/ACCESS.2023.3340719.