frontiers | Frontiers in Energy Research

ORIGINAL RESEARCH
published: 15 August 2022
doi: 10.3389/fenrg.2022.956768

Electricity Demand Forecasting With a Modified Extreme-Learning Machine Algorithm

Chen Chen, Chuangang Ou , Mingxiang Liu and Jingtao Zhao*

NARI Technology Co., Ltd., Nanjing, China

To operate the power grid safely and reduce the cost of power production, power-load forecasting has become an urgent issue to be addressed. Although many power load forecasting models have been proposed, most still suffer from poor model training, limitations sensitive to outliers, and overfitting of load forecasts. The limitations of current load-forecasting methods may lead to the generation of additional operating costs for the power system, and even damage the distribution and network security of the related systems. To address this issue, a new load prediction model with mixed loss functions was proposed. The model is based on Pinball–Huber's extreme-learning machine and whale optimization algorithm. In specific, the Pinball–Huber loss, which is insensitive to outliers and largely prevents overfitting, was proposed as the objective function for extreme-learning machine (ELM) training. Based on the Pinball–Huber ELM, the whale optimization algorithm was added to improve it. At last, the effect of the proposed hybrid loss function prediction model was verified using two real power-load datasets (Nanjing and Taixing). Experimental results confirmed that the proposed hybrid loss function load prediction model can achieve satisfactory improvements on both datasets.

Keywords: outliers, whale optimization algorithm, load forecasting, Pinball–Huber regression, extreme-learning machine

OPEN ACCESS

Edited by:
Qihe Shan,
Dalian Maritime University, China

Reviewed by:
Yang Yang,
Nanjing University of Posts and Telecommunications, China
Jinran Wu,
Queensland University of Technology, Australia
Li Xi'An,
Shanghai Jiao Tong University, China

****Correspondence:***
Chen Chen
chenchen3@sgepri.sgcc.com.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research.

Received: *30 May 2022*
Accepted: *17 June 2022*
Published: *15 August 2022*

Citation:
Chen C, Ou C, Liu M and Zhao J (2022) Electricity Demand Forecasting With a Modified Extreme-Learning Machine Algorithm. Front. Energy Res. 10:956768. doi: 10.3389/fenrg.2022.956768

1 INTRODUCTION

As an integrated system that can optimize the allocation of energy resources according to the regional energy structure and energy reserves, integrated energy systems have become an important way to accelerate the global sustainable energy transformation (Wu et al., 2019, 2021). Power-load forecasting is an important part of the power system (Ahmad et al., 2020; Yang et al., 2022b). Accurate power-load forecasting can arrange the start and stop of generator sets more economically and reasonably to maintain power supply and demand balance (Shi et al., 2021), and maintain the safety and stability of power grid operation (Dynge et al., 2021). In addition, it can effectively reduce the cost of power generation, transmission, and distribution; improve economic and social benefits; and ensure the operation of the society (Chu et al., 2021; Lin and Shi, 2022). The existing mainstream power load forecasting methods are mainly divided into two categories: statistical (Rehman et al., 2022) and artificial intelligence methods (Aslam et al., 2021). Factors such as seasons (van der Meer et al., 2018), climates (Alipour et al., 2019), and temperature (Yang et al., 2022d) have a direct impact on power load. Statistical methods are a very effective solution to such systems with trends, seasons, and periodic changes. Many scholars have carried out research on power load forecasting based on these methods such as the auto regressive (AR) (Louzazni et al., 2020), auto regressive moving average (Yan and Chowdhury, 2014), and auto regressive integrated moving average models (Asadi et al., 2012). The statistical methods usually take power load or energy as a single input series, while the artificial intelligence methods consider the relationship between the output and multiple influencing factors. Under the condition of sufficient historical samples, artificial intelligence methods usually have high forecast accuracy and strong generalization ability, such as the support vector machine (Yang et al., 2022a) and neural networks (NNs) (Huang et al., 2002; Oreshkin et al., 2021). Extreme-learning machine (ELM) is an emerging generalized single-hidden-layer feed-forward neural-network-learning algorithm, which can generate hidden variable parameters at random to calculate output weights, and it is widely used in forecast. Liu and Wang (2022) proposed a transfer-learning-based probabilistic wind power forecasting method. Model-based transfer learning is utilized to construct the multilayer extreme-learning machine. The enhanced Crow search algorithm–ELM (ENCSA–ELM) model was proposed to accurately forecast short-term wind power to improve the utilization efficiency of clean energy in Li et al. (2021). ELM is more efficient, has lower computational costs, and has greater generalization than shallow learning systems.

Both time series and artificial intelligence methods usually take the loss function as the training objective. The existing literature mainly uses the mean absolute error (L1 loss), mean absolute percentage error (MAPE), and root mean square error (RMSE) to evaluate the effect of power load forecasting. Furthermore, to develop an algorithmic framework capable of handling data containing outliers, a robust loss function, Huber–Loss (Ge et al., 2019), has been introduced. Compared with other loss functions, this function has different sensitivity to abnormal data and is more tolerant to noise. Furthermore, the loss function can adjust the robustness of the model according to the tuning parameters τ, and it can better suppress the influence of outliers.

However, these improved evaluation indicators are still mostly based on the absolute value criterion, only considering the size of the error, but not the direction of the error. They also do not fully account for the different consequences of positive and negative errors. In fact, the positive and negative errors of the power load forecasting affect the reliability and economy of power differently, so the error evaluation indicators should be differentiated. The improvement of the abovementioned traditional indicators is mainly reflected in the improvement of the mathematical form, the introduction or construction of new statistics, and the construction of a multiindicator evaluation system.

Hybrid algorithmic frameworks have been developed and widely used in power-load forecasting. However, these algorithmic frameworks have hyperparameters that need to be carefully optimized before forecasting. The optimized values of these algorithms determine the performance of the forecast (Yang et al., 2022c). Grid search, gradient descent, and cross validation are commonly used methods for optimizing the parameters of forecast models. The studies have also proposed nature-inspired meta-heuristic optimization algorithms to efficiently optimize these parameters. Geng et al. (2015) proposed a load-forecasting model hybridizing the seasonal SVR and chaotic cloud simulated annealing algorithm to receive more accurate forecasting performance. Xie et al. (2020) proposed a method combined Elman neural network and the particle swarm optimization for the short-term power load forecasting. Heydari et al. (2020) proposed a hybrid model that considers price and load forecasting, including variational mode decomposition, generalized regression NNs, and gravitational search algorithms.

In summary, the current requirement for power load forecasting is increasing from the following perspectives: 1) the forecasting accuracy needs to be improved; 2) a robust loss function is required to develop machine learning framework that can fully account for the different consequences of positive and negative errors and outliers; and 3) more advanced optimization methods are needed to improve model parameters. The contributions of this article are the following: 1) A new hybrid model was proposed to improve the load-forecasting accuracy and prevent overfitting, which combines the Pinball–Huber–ELM with WOA. In specific, in our proposed Pinball–ELM, WOA is employed to search weights and thresholds, which provide good training results for load prediction; and 2) an improved ELM was developed to handle data with outliers. Due to its excellent properties, the Pinball–Huber loss was incorporated into the ELM as the objective function for its training.

The rest of this article is organized as follows. In **Section 2**, we review the basic ELM and propose our powerful ELM. Next, **Section 3** introduces WOA. **Section 4** then illustrates our proposed hybrid loss function load prediction model and presents the model-training process for cross validation of tuning parameters. In **Section 5**, the testing of the proposed hybrid load-forecasting model WOA–Pinball–Huber–ELM using two datasets from Nanjing and Taixing is described. **Section 6** concludes the article.

2 PINBALL–HUBER EXTREME-LEARNING MACHINE

2.1 Extreme-Learning Machine

Unlike traditional NNs, ELM is a single-hidden-layer feed-forward NN that randomly selects its input weights and thresholds. The number of nodes in the input layer, hidden layer, and output layer are *N*, *L*, and *M*, respectively. Under the action of the activation function, the hidden layer output matrix *H* is as follows:

$$H = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & g(\omega_2 \cdot x_1 + b_2) & \cdots & g(\omega_L \cdot x_1 + b_L) \\ g(\omega_1 \cdot x_2 + b_1) & g(\omega_2 \cdot x_2 + b_2) & \cdots & g(\omega_L \cdot x_2 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(\omega_1 \cdot x_N + b_1) & g(\omega_2 \cdot x_N + b_2) & \cdots & g(\omega_L \cdot x_N + b_L) \end{bmatrix}_{N\times L}, \tag{1}$$

where x is the input matrix, ω is the input weight matrix, and b is the threshold in the hidden layer, which are randomly generated in ELM. The output T of the ELM is then

$$T = [t_1, t_2, \ldots, t_N]_{M\times N} = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{Mj} \end{bmatrix}_{M\times L} = \begin{bmatrix} \sum_{i=1}^{t} \beta_{i1} \cdot g\left(\omega_i \cdot x_j + b_i\right) \\ \sum_{i=1}^{t} \beta_{i2} \cdot g\left(\omega_i \cdot x_j + b_i\right) \\ \vdots \\ \sum_{i=1}^{t} \beta_{iM} \cdot g\left(\omega_i \cdot x_j + b_i\right) \end{bmatrix}_{M\times L} \quad (j = 1, 2, \ldots, N), \tag{2}$$

where β is the correlation weight matrix between the hidden and output layers.

ELM mainly uses the randomly generated ω and b, and it selects the least square method to complete the calculation of the β. The algorithm does not need to perform multiple solving operations, which greatly reduces the complexity of the operation.

According to the two theorems in Zhang et al. (2020), when the activation function is differentiable, it is not necessary to adjust all parameters in ELM. At last, the solution of β can be obtained as follows:

$$\beta = H^{-1} T', \tag{3}$$

where H^{-1} is the generalized inverse matrix of H.

2.2 Regression Loss Function

The regression loss function represents the gap between the predicted and actual value. If the gap is larger, the value of the loss function is larger; otherwise, its value is smaller. During the optimization process, through continuous learning and training, the value of the loss function is gradually reduced, so that the performance of the model is continuously improved.

2.2.1 $L2$ Loss

In the training of forecast models, the most commonly used loss function is $L2$ loss (mean squared error), which is defined as

$$L2 = \frac{1}{M} \sum_{i=1}^{M} (y_i - \hat{y}_i)^2, \tag{4}$$

where M is the number of output samples in the training set, y_i represents the expected output of the training set, and $\hat{y}_i$ represents the forecast output of the training set. For models using the $L2$ loss, the convergence is fast when the error is large. However, the $L2$ loss is sensitive to outliers, which affects the performance of the forecast model.

2.2.2 $L1$ Loss

The $L1$ loss (mean absolute error) is more robust to outliers than the $L2$ loss, which is defined as

$$L1 = \frac{1}{M} \sum_{i=1}^{M} |y_i - \hat{y}_i|, \tag{5}$$

where M is the number of output samples in the training set, y_i represents the expected output of the training set, and $\hat{y}_i$ represents the forecast output of the training set. Although the $L1$ loss enhances robustness, it is not smooth and nondifferentiable at zero, and it converges slowly.

2.2.3 Huber Loss

Huber loss is a combination of the $L2$ and $L1$ losses, which includes a parameter δ. δ determines the degree of inclination of the Huber loss on the $L1$ and $L2$ losses; that is, it is used to control the quadratic and linear range of the loss function. The Huber loss combines the advantages of the $L1$ and $L2$ losses, and it is more robust to outliers than the $L2$ loss, while converging faster.

Huber loss reduces the penalty for outliers, so it is a commonly used robust loss function. It is defined as

$$\rho_\delta(r) = \begin{cases} \frac{1}{2} r^2 & |r| \le \delta \\ |r|\delta - \frac{\delta^2}{2} & |r| > \delta, \end{cases} \tag{6}$$

where r represents the absolute value of the difference between the expected output and predicted output. δ represents the tuning parameter, which is used to determine the behavior of the model to deal with outliers.

2.2.4 Pinball Loss

Pinball loss is mostly used in regression analysis problems, which is related to the quantile distance and is not sensitive to outliers. The Pinball loss used is defined as

$$L_\tau = \begin{cases} (y_i - \hat{y}_i)\tau & y_i \ge \hat{y}_i \\ (\hat{y}_i - y_i)(1-\tau) & y_i < \hat{y}_i, \end{cases} \tag{7}$$

where $\tau \in [0, 1]$ is the target quantile to adjust for the positive and negative errors in the forecast. y_i and $\hat{y}_i$ are defined as above. When $\tau = 0.5$, the Pinball loss is the same as the L1 loss, and it can be considered as a generalized L1 loss. When $\tau \ne 0.5$, the Pinball loss has different penalties for positive and negative errors.

2.2.5 Proposed Pinball–Huber Loss

To implement different penalties for positive and negative errors during training, and enhance the robustness of the loss function, thereby improving the accuracy, the proposed improved loss function named the Pinball–Huber loss is

$$V(r) = \begin{cases} \frac{1}{2}\tau r^2 & -\delta \le r < 0 \\ \frac{1}{2}(1-\tau) r^2 & 0 \le r \le \delta \\ \tau\left(|r|\delta - \frac{\delta^2}{2}\right) & r < -\delta \\ (1-\tau)\left(|r|\delta - \frac{\delta^2}{2}\right) & r > \delta, \end{cases} \tag{8}$$

where $\tau \in [0, 1]$ is the target quantile to adjust for the positive and negative errors in the forecast. δ represents the threshold, which is used to determine the behavior of the model to deal with outliers. r represents the absolute value of the difference between the expected output and predicted output. Compared with Huber loss, the Pinball–Huber loss maintains its low sensitivity to outliers in the data and implements different penalties for positive and negative errors, considering the direction of errors. In power load forecasting, since there are often outliers in power load data, and positive and negative errors should be distinguished, our proposed Pinball–Huber is expected to improve the forecasting accuracy and convergence speed.

2.3 Pinball–Huber–Extreme-Learning machine

The presence of data outliers can affect the prediction performance of ELM. The Pinball–Huber loss obtained by introducing the pinball feature based on the Huber loss can effectively distinguish the effects of positive and negative errors and further improve the accuracy of prediction, while maintaining robustness. Therefore, this study introduces the Pinball–Huber loss into the traditional ELM and uses the Pinball–Huber loss as the objective function to train the ω and b of ELM. When dealing with power data with outliers, the model shows quite good robustness, and greatly improves the ELM effect regarding the prediction accuracy. The improved model named the Pinball–Huber–ELM is as follows:

$$\begin{aligned} &\min_{\omega,b} \quad MHL = \frac{1}{M}\sum_{i=1}^{M} V\left(r_i\right) \\ &s.t. \quad V\left(r\right) = \begin{cases} \frac{1}{2}\tau r^2 & -\delta \le r < 0 \\ \frac{1}{2}\left(1-\tau\right) r^2 & 0 \le r \le \delta \\ \tau\left(\left|r\right|\delta - \frac{\delta^2}{2}\right) & r < -\delta \\ \left(1-\tau\right)\left(\left|r\right|\delta - \frac{\delta^2}{2}\right) & r > \delta \end{cases} \\ &\qquad \omega_i = \left[\omega_{i1}, \omega_{i2}, \ldots, \omega_{iN}\right]^T \\ &\qquad b = \left[b_1, b_2, \ldots, b_L\right]^T. \end{aligned} \tag{9}$$

3 WHALE OPTIMIZATION ALGORITHM

The whale optimization algorithm was inspired by the unique bubble net prey method of the whale population (Mirjalili et al., 2016). It searches for the optimal solution through the following three mechanisms: surrounding the prey, searching for the prey, and attacking the prey by the spiral bubble net.

3.1 Surround Prey

1) Whale swimming toward the optimal position

Whale groups can find out the coordinates of their prey and surround them during hunting. In WOA, it is assumed that the position of the optimal individual whale in the current population is the position of the prey, and the optimal whale is surrounded by other whales. The mathematical model is

$$D = \left|C \cdot X^*\left(t\right) - X\left(t\right)\right|, \tag{10}$$

$$X\left(t+1\right) = X^*\left(t\right) - A \cdot D. \tag{11}$$

In **Eqs. 10** and **11**, t is the current iteration; $X*(t)$ is the best-obtained solution in the previous iteration; $X(t)$ is the solution in the current iteration; and $X(t)$ is the solution in the next iteration. The specific formulas of the coefficient vectors A and C are

$$A = 2a \cdot r_2 - a, \tag{12}$$

$$C = 2 \cdot r_1. \tag{13}$$

In **Eqs. 12** and **13**, r_1, r_2 are the two numbers randomly selected in the range of $[-1, 1]$, and a is the convergence factor. As the solution is updated, the value of a decreases linearly from 2 to 0. The formula is

$$a = 2 - 2\left(\frac{t}{t_{\max}}\right), \tag{14}$$

where $t_{\max}$ is the maximum number of iterations.

Equation 11 shows that it can be updated on the basis of the current optimal individual position (X^*, Y^*) and continue to search for the individual position (X, Y). Y^* represents the fitness value obtained from position X^*. Y represents the fitness value obtained from position X. By adjusting the values of the A and C vectors, we can achieve various positions around the best position relative to the current position. Any individual whale is allowed to update its position near the current optimal solution and simulate surrounding the prey.

2) Whales swimming toward random locations

In the process of searching for prey, the method that the vector A changes with the iterative process can be used. In effect, humpback whales randomly explore the solution space based on each other's positions. Therefore, A is used in the global exploration phase to update the whale position so as to stay away from the current individual when the random value is >1 or <1.

In contrast to the local development phase, in the global exploration phase, the positions of individual whales are upgraded based on randomly selected individuals, rather than the best whales found so far. This mechanism focuses on exploration. Thus, when $|A| > 1$, the WOA algorithm performs a global exploration operation. During the prey-hunting phase, the location of the prey is unknown to the whale population. This mechanism focuses on optimization. Thus, when $|A| < 1$, the whales obtain the location of the prey through collective cooperation. Whales use random individual positions in the population as navigation targets to find food, and the mathematical model is described as follows:

$$D = \left|C \cdot X_{rand} - X\right|, \tag{15}$$

$$X\left(t+1\right) = X_{rand} - A \cdot D. \tag{16}$$

In **Eqs. 15** and **16**, X_{rand} represents the position of the whales randomly selected in the current whale population.

The WOA algorithm begins execution with a random set of whale swarm locations. The shrinking envelope is achieved as the convergence factor a decreases. The fluctuation range of the coefficient vector A also decreases as the convergence factor a decreases. That is, when the convergence factor a decreases from 2 to 0 during the iteration, the fluctuation of the coefficient vector A also decreases; its range is $[-a, a]$.

In each iteration, the whale individual updates its position using the randomly selected whale position information or the whale individual position information with the best fitness value obtained so far. As the parameter a decreases from 2 to 0, the transition of the algorithm between the global exploration phase and local development phase is realized. When $|A| > 1$, we randomly select a whale in the population; when $|A| < 1$, we select the current whale with the best fitness value to update the position of the individual whale. Given the value of p, WOA has the ability to swap between helical or circular motion. At last, satisfying a termination condition terminates the WOA algorithm.

3.2 Bubble Net Chase

Two methods were used to build a mathematical model of the predation behavior of whales in bubble nets. The first is the reduction of the wraparound mechanism, which is achieved by reducing the value of a in **Eq. 14**, where the fluctuation range of A also decreases accordingly. In other words, A represents a random value in the interval $[-a, a]$, where a decreases from 2 to 0 during the iteration. Defining a random value of A in [−1, 1], the new position of the individual whale can be defined somewhere between the whale's original position and the current best whale position. The shrinking and wrapping mechanism of whale group predation can be represented by a two-dimensional space $(0 \le A \le 1)$. This space contains all possible positions to transform from (X, Y) to (X^*, Y^*).

The second is the spiral update position mechanism. The method first calculates the distance between the whale located at (X, Y) and the prey located at (X^*, Y^*). Between the location of the whale and prey, the researchers used a spiral equation to mimic the spiral of a humpback whale shape motion. Its mathematical model is described as

$$X(t+1) = D' \cdot e^{bl} \cdot cos(2\pi l) + X^*(t). \tag{17}$$

In **Eq. 17**, $D' = |X^*(t) - X(t)|$ represents the distance between the optimal whale individual and current whale individual in the tth iteration; b represents the constant of the logarithmic spiral equation; l is the random value between [−1, 1] number; and "·" is element-wise multiplication.

Whales follow a spiral path while swimming around their prey in a shortened circle. To obtain a model that simulates this behavior, it is assumed that there is a 50% probability during the optimization process to randomly choose between the encircling mechanism and spiral updating position mechanism to update the positions of individual whales. Its mathematical model is

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D, & p < 0.5 \\ D' \cdot e^{bl} \cdot cos(2\pi l) + X^*(t), & p \le 0.5. \end{cases} \tag{18}$$

In **Eq. 18**, p represents the random number between [0, 1]. After the bubble net chase pattern, the humpback whales begin to randomly search for prey.

4 PROPOSED FORECASTING MODEL

In this section, we propose a hybrid loss function power load forecasting model. This model combines a powerful ELM with an improved Pinball–Huber loss function. To optimize the effect of the improved algorithm, we need to obtain the optimal solution of the tuning parameters of Huber loss through two cross validation in advance.

At last, the implementation flow of our hybrid loss function prediction model is as follows:

1) We obtain the original data and preprocess them. We divide the data into training and test sets appropriately.
2) The training set is then divided into five subsets. Any nonrepetitive part of the five subsets (i.e., any subset) is used as the training set; the remaining four parts of the training set (i.e., the remaining subsets) are used as the test set. We compute MSE_i using the test set and employ a different subset as the test set each time. We use five-fold cross validation for parameters δ and τ. We divide the value range of a into five equal parts. Then we randomly pick a value from each interval and obtain five values, denoted as $[0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 1]$. Experiments were performed five times with different values for each test set used. We use five δ-values and five τ-values to combine 25 experimental data.
3) We first assign empirical parameter values, then apply 2 five-fold cross-validations to average the 25 values of MSE_i to obtain the final average MSE, called CV, $CV = \left(\sum_{i=1}^{k} MSE_i\right)/k$. Twenty-five MSE_i's were compared, and the minimum value of MSE is selected to be substituted into the Pinball–Huber loss regression function.
4) In the training set, we take the minimum value of the Pinball–Huber as the goal, use WOA to solve the optimal parameters ω_i, b_j of ELM, and substitute them into formula to obtain β_{jk}.
5) The input weight of the model, the threshold of intermediate nodes, and the output weight of the model are all brought into the ELM model, and then the input of the test set is substituted into the model to obtain the prediction output of the test set.

5 CASE STUDIES

We performed power-load forecasts for Nanjing and Taixing power data. We recorded Nanjing's power load data (total load power of the grid/MW every 15 min) every half an hour. Nanjing data have 1920 data points (2003.2.18.00:00–2003.3.29.23:30). The training set included 1,152 data points, and the test set included 768 data points. We record Taixing's electricity load data (daily electricity consumption/10,000 kwh) every other day. Taixing data have 1,175 data points (2018.5.13–2021.8.2). The training set includes 705 data points, and the test set includes 470 data points. The specific situation is shown in the **Table 1**.

An evaluation is performed in this subsection; the performance of our proposed WOA–Pinball–Huber–ELM algorithm is evaluated using the power load data of Nanjing and Taixing, as shown in **Tables 2** and **3**. The actual electrical load power and the WOA–Pinball–Huber–ELM-based electrical load forecast result graph for Nanjing's data are shown in **Figure 1**. The

TABLE 1 | Characteristics of experimental data in Nanjing and Taixing.

	Dataset	Size	Min	Max	Median	Mean	Std.
Nanjing	Total data	1920	2,362.152	5,276.500	3,879.217	3,802.732	607.709
	Training data	1,152	2,564.228	5,151.652	3,991.129	3,860.526	635.217
	Testing data	768	2,362.152	5,276.500	3,775.403	3,716.042	553.120
Taixing	Total data	1,175	1,210.872	2,875.318	1,893.728	1,902.926	249.797
	Training data	705	1,210.872	2,516.598	1,799.584	1,795.856	203.690
	Testing data	470	1,424.918	2,875.318	2,040.657	2,063.533	225.466

TABLE 2 | Evaluation index of Nanjing data is obtained from three prediction algorithm experiments.

	Train			Test		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
WOA-ELM	191.52	178.92	0.04	202.42	187.02	0.04
WOA-L1-ELM	168.93	120.63	0.04	194.06	155.35	0.04
WOA-L2-ELM	161.20	120.82	0.04	193.94	152.58	0.04
WOA-Pinball-ELM	182.92	141.66	0.04	192.49	150.48	0.04
WOA-HUBER-ELM	172.50	136.45	0.04	188.21	149.17	0.04
WOA-Pinball-Huber-ELM	153.69	115.11	0.03	186.05	146.58	0.04

TABLE 3 | Evaluation index of Taixing data is obtained from three prediction algorithm experiments.

	Train			Test		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
WOA-ELM	58.56	43.78	0.02	74.04	60.25	0.02
WOA-L1-ELM	59.50	41.47	0.02	69.67	58.04	0.02
WOA-L2-ELM	57.71	41.68	0.02	73.91	57.45	0.02
WOA-Pinball-ELM	58.49	42.00	0.02	70.79	51.62	0.02
WOA-HUBER-ELM	57.44	41.30	0.02	73.36	54.72	0.03
WOA-Pinball-Huber-ELM	57.62	41.58	0.02	69.57	51.48	0.02

actual electrical load power and the WOA–Pinball–Huber–ELM-based electrical load forecast result graph for Taixing data are shown in **Figure 2**.

As far as the power load data in Nanjing are concerned, we use the proposed algorithm (WOA–Pinball–Huber–ELM) and the compared algorithms (WOA–ELM, WOA–L1–ELM, WOA–L2–ELM, WOA–Pinball–ELM, WOA–Huber–ELM) for experiments, and the experimental results are shown in **Table 2**. From the experimental results of the test set in **Table 2**, the four (improved based on the basic WOA–ELM) algorithms and the basic WOA–electric power data predicted by ELM algorithm. The above four improved algorithms corresponding to the calculated three evaluation indicators (RMSE, MAE, and MAPE) data are mostly better than the three calculated by the basic WOA–ELM algorithm.

The Pinball–Huber loss function is obtained by combining the Pinball and Huber loss. We improved the WOA–Pinball–Huber–ELM, WOA–Pinball–ELM, WOA–Huber–ELM, WOA–L1–ELM, and WOA–L2–ELM algorithms from the Pinball–Huber, Pinball loss, Huber loss, and L1 and L2, respectively. Furthermore, we used these algorithms to predict the electric power data and calculate the corresponding three evaluation indicators. The three evaluation indexes calculated by the WOA–Pinball–Huber–ELM algorithm were better than the three evaluation indexes calculated by the improved WOA–Pinball–ELM and WOA–Huber–ELM algorithms. In addition, the three evaluation indexes calculated by the WOA–Pinball–Huber–ELM algorithm were better than the three evaluation indexes calculated by the improved WOA–L1–ELM and WOA–L2–ELM algorithms.

Considering the power load data of Taixing, the experimental results are presented in **Table 3**. From the experimental results of the test set in **Table 3**, it can be seen that the four improved algorithms and the basic WOA–ELM algorithm predict the power load prediction power data. The power data obtained by the above four improved algorithms correspond to the calculated

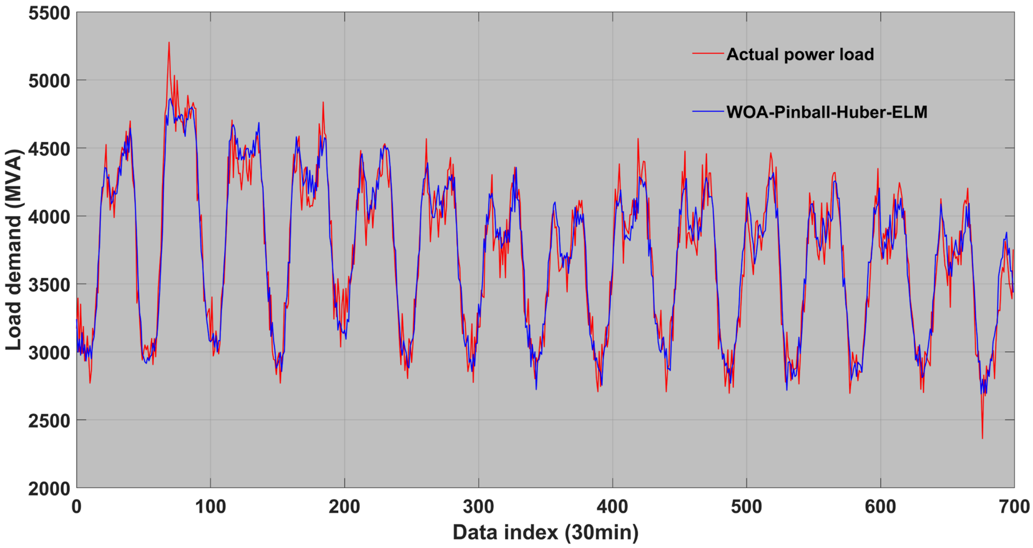

FIGURE 1 | Actual power and prediction result graph based on WOA–Pinball–Huber–extreme-learning machine (ELM) in Taixing.

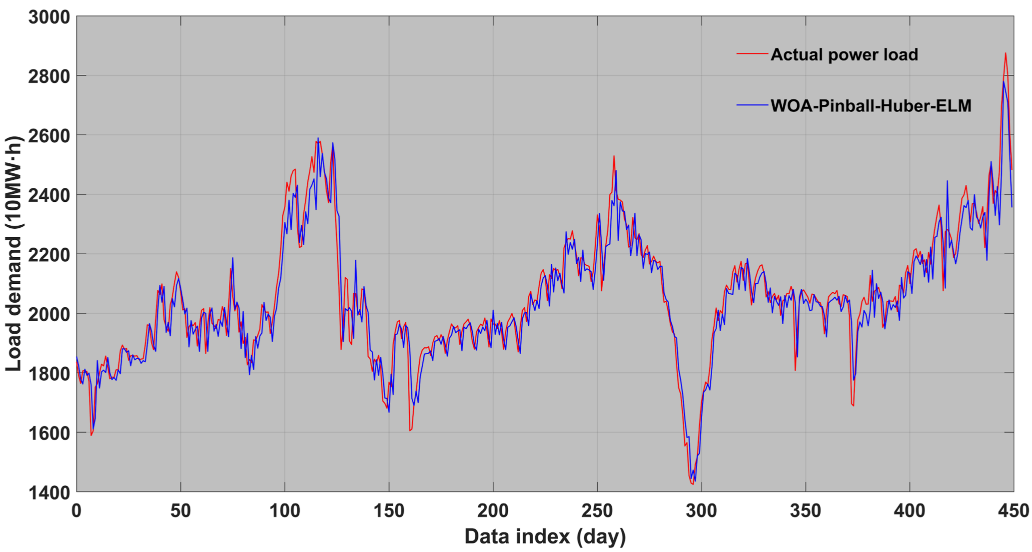

FIGURE 2 | Actual power and prediction result graph based on WOA–Pinball–Huber–ELM in Taixing.

three evaluation indicators RMSE, MAE, and MAPE data than the three calculated using the basic WOA–ELM algorithm. The evaluation index data are small, and the effect is better.

It is obvious from **Table 3** that the three evaluation indexes calculated using the WOA–Pinball–Huber–ELM algorithm were better than the three evaluation indexes calculated by the improved WOA–Pinball–ELM and WOA–Huber–ELM algorithms. At last, the three evaluation metrics calculated by the WOA–Pinball–Huber–ELM algorithm are better than those calculated by the improved WOA–L1–ELM and WOA–L2–ELM algorithms.

6 CONCLUSION

To ensure the safe operation of the grid, we must confirm that the power-load forecast is accurate and effective. However, the complexity of the grid structure brings many difficulties to future power-load forecasting, and the current popular forecasting methods cannot handle all the difficulties. To address this challenge, this study proposed a new hybrid loss function load prediction model, the WOA–Pinball–Huber–ELM. It is a combination of the Pinball–Huber ELM and whale optimization algorithm. The Pinball–Huber loss, which is insensitive to outliers and largely prevents overfitting, is treated as the objective function for our optimized ELM training. Based on two real power load forecasting datasets in Nanjing and Taixing and comparative experiments with two improved algorithms, our WOA–Pinball–Huber–ELM model shows great advantages in handling outliers and improving forecasting accuracy.

In future work, our proposed framework can be employed for other forecasting problems in environmental science (Zhang et al., 2021, 2022) and bioinformatics (Miao et al., 2022).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

CC: supervision, investigation, project administration; CO: software, visualization, formal analysis, writing–original draft; ML: writing–review and editing; JZ: formal analysis, writing–review and editing

FUNDING

This research was funded by the National Key Research and Development Program of China under Grant No. 2021YFB2401303.

REFERENCES

Ahmad, T., Zhang, H., and Yan, B. (2020). A Review on Renewable Energy and Electricity Requirement Forecasting Models for Smart Grid and Buildings. *Sustain. Cities Soc.* 55, 102052. doi:10.1016/j.scs.2020.102052

Alipour, P., Mukherjee, S., and Nateghi, R. (2019). Assessing Climate Sensitivity of Peak Electricity Load for Resilient Power Systems Planning and Operation: A Study Applied to the texas Region. *Energy* 185, 1143–1153. doi:10.1016/j.energy.2019.07.074

Asadi, S., Tavakoli, A., and Hejazi, S. R. (2012). A New Hybrid for Improvement of Auto-Regressive Integrated Moving Average Models Applying Particle Swarm Optimization. *Expert Syst. Appl.* 39, 5332–5337. doi:10.1016/j.eswa.2011.11.002

Aslam, S., Herodotou, H., Mohsin, S. M., Javaid, N., Ashraf, N., and Aslam, S. (2021). A Survey on Deep Learning Methods for Power Load and Renewable Energy Forecasting in Smart Microgrids. *Renew. Sustain. Energy Rev.* 144, 110992. doi:10.1016/j.rser.2021.110992

Chu, W., He, W., Jiang, Q., Zhang, S., Hu, Z., Xu, G., et al. (2021). Optimization of Operation Strategy for a Grid Interactive Regional Energy System. *Energy Build.* 250, 111294. doi:10.1016/j.enbuild.2021.111294

Dynge, M. F., Crespo del Granado, P., Hashemipour, N., and Korpås, M. (2021). Impact of Local Electricity Markets and Peer-To-Peer Trading on Low-Voltage Grid Operations. *Appl. Energy* 301, 117404. doi:10.1016/j.apenergy.2021.117404

Ge, J., Li, H., Wang, H., Dong, H., Liu, H., Wang, W., et al. (2019). Aeromagnetic Compensation Algorithm Robust to Outliers of Magnetic Sensor Based on Huber Loss Method. *IEEE Sensors J. PP* 1–1. doi:10.1109/jsen.2019.2907398

Geng, J., Huang, M.-L., Li, M.-W., and Hong, W.-C. (2015). Hybridization of Seasonal Chaotic Cloud Simulated Annealing Algorithm in a Svr-Based Load Forecasting Model. *Neurocomputing* 151, 1362–1373. doi:10.1016/j.neucom.2014.10.055

Heydari, A., Majidi Nezhad, M., Pirshayan, E., Astiaso Garcia, D., Keynia, F., and De Santoli, L. (2020). Short-term Electricity Price and Load Forecasting in Isolated Power Grids Based on Composite Neural Network and Gravitational Search Optimization Algorithm. *Appl. Energy* 277, 115503. doi:10.1016/j.apenergy.2020.115503

Huang, H.-C., Hwang, R.-C., and Hsieh, J.-G. (2002). A New Artificial Intelligent Peak Power Load Forecaster Based on Non-fixed Neural Networks. *Int. J. Electr. Power & Energy Syst.* 24, 245–250. doi:10.1016/s0142-0615(01)00026-6

Li, L.-L., Liu, Z.-F., Tseng, M.-L., Jantarakolica, K., and Lim, M. K. (2021). Using Enhanced Crow Search Algorithm Optimization-Extreme Learning Machine Model to Forecast Short-Term Wind Power. *Expert Syst. Appl.* 184, 115579. doi:10.1016/j.eswa.2021.115579

Lin, B., and Shi, L. (2022). New Understanding of Power Generation Structure Transformation, Based on a Machine Learning Predictive Model. *Sustain. Energy Technol. Assessments* 51, 101962. doi:10.1016/j.seta.2022.101962

Liu, Y., and Wang, J. (2022). Transfer Learning Based Multi-Layer Extreme Learning Machine for Probabilistic Wind Power Forecasting. *Appl. Energy* 312, 118729. doi:10.1016/j.apenergy.2022.118729

Louzazni, M., Mosalam, H., Khouya, A., and Amechnoue, K. (2020). A Non-linear Auto-Regressive Exogenous Method to Forecast the Photovoltaic Power Output. *Sustain. Energy Technol. Assessments* 38, 100670. doi:10.1016/j.seta.2020.100670

Miao, M., Wu, J., Cai, F., and Wang, Y.-G. (2022). A Modified Memetic Algorithm with an Application to Gene Selection in a Sheep Body Weight Study. *Animals* 12, 201. doi:10.3390/ani12020201

Mirjalili, S., Lewis, A., and Andrew, S. (2016). The Whale Optimization Algorithm. *Adv. Eng. Softw.* 95, 51–67. doi:10.1016/j.advengsoft.2016.01.008

Oreshkin, B. N., Dudek, G., Pełka, P., and Turkina, E. (2021). N-Beats Neural Network for Mid-term Electricity Load Forecasting. *Appl. Energy* 293, 116918. doi:10.1016/j.apenergy.2021.116918

Rehman, A., Zhu, J.-J., Segovia, J., and Anderson, P. R. (2022). Assessment of Deep Learning and Classical Statistical Methods on Forecasting Hourly Natural Gas Demand at Multiple Sites in spain. *Energy* 244, 122562. doi:10.1016/j.energy.2021.122562

Shi, Y., Li, Y., Zhou, Y., Xu, R., Feng, D., Yan, Z., et al. (2021). Optimal Scheduling for Power System Peak Load Regulation Considering Short-Time Startup and Shutdown Operations of Thermal Power Unit. *Int. J. Electr. Power. Energy Syst.* 131, 107012. doi:10.1016/j.ijepes.2021.107012

van der Meer, D. W., Munkhammar, J., and Widén, J. (2018). Probabilistic Forecasting of Solar Power, Electricity Consumption and Net Load: Investigating the Effect of Seasons, Aggregation and Penetration on Prediction Intervals. *Sol. Energy* 171, 397–413. doi:10.1016/j.solener.2018.06.103

Wu, J., Cui, Z., Chen, Y., Kong, D., and Wang, Y.-G. (2019). A New Hybrid Model to Predict the Electrical Load in Five States of australia. *Energy* 166, 598–609. doi:10.1016/j.energy.2018.10.076

Wu, J., Wang, Y.-G., Tian, Y.-C., Burrage, K., and Cao, T. (2021). Support Vector Regression with Asymmetric Loss for Optimal Electric Load Forecasting. *Energy* 223, 119969. doi:10.1016/j.energy.2021.119969

Xie, K., Yi, H., Hu, G., Li, L., and Fan, Z. (2020). Short-term Power Load Forecasting Based on Elman Neural Network with Particle Swarm Optimization. *Neurocomputing* 416, 136–142. doi:10.1016/j.neucom.2019.02.063

Yan, X., and Chowdhury, N. A. (2014). Mid-term Electricity Market Clearing Price Forecasting Utilizing Hybrid Support Vector Machine and Auto-Regressive Moving Average with External Input. *Int. J. Electr. Power & Energy Syst.* 63, 64–70. doi:10.1016/j.ijepes.2014.05.037

Yang, Y., Tao, Z., Qian, C., Gao, Y., Zhou, H., Ding, Z., et al. (2022a). A Hybrid Robust System Considering Outliers for Electric Load Series Forecasting. *Appl. Intell.* 52, 1630–1652. doi:10.1007/s10489-021-02473-5

Yang, Y., Wang, Z., Gao, Y., Wu, J., Zhao, S., and Ding, Z. (2022b). An Effective Dimensionality Reduction Approach for Short-Term Load Forecasting. *Electr. Power Syst. Res.* 210, 108150.

Yang, Y., Zhou, H., Wu, J., Ding, Z., and Wang, Y.-G. (2022c). Robustified Extreme Learning Machine Regression with Applications in Outlier-Blended Wind-Speed Forecasting. *Appl. Soft Comput.* 122, 108814. doi:10.1016/j.asoc.2022.108814

Yang, Y., Zhou, H., Wu, J., Liu, C.-J., and Wang, Y.-G. (2022d). A Novel Decompose-Cluster-Feedback Algorithm for Load Forecasting with Hierarchical Structure. *Int. J. Electr. Power & Energy Syst.* 142, 108249. doi:10.1016/j.ijepes.2022.108249

Zhang, B., Tan, R., and Lin, C. J. (2020). Forecasting of E-Commerce Transaction Volume Using a Hybrid of Extreme Learning Machine and Improved Moth-Flame Optimization Algorithm. *Appl. Intell.* 1–14, 1840. doi:10.1007/s10489-020-01840-y

Zhang, S., Wu, J., Jia, Y., Wang, Y.-G., Zhang, Y., and Duan, Q. (2021). A Temporal Lasso Regression Model for the Emergency Forecasting of the Suspended Sediment Concentrations in Coastal Oceans: Accuracy and Interpretability. *Eng. Appl. Artif. Intell.* 100, 104206. doi:10.1016/j.engappai.2021.104206

Zhang, S., Wu, J., Wang, Y.-G., Jeng, D.-S., and Li, G. (2022). A Physics-Informed Statistical Learning Framework for Forecasting Local Suspended Sediment Concentrations in Marine Environment. *Water Res.* 218, 118518. doi:10.1016/j.watres.2022.118518

Conflict of Interest: Authors CC, CO, ML, and JZ were employed by the company NARI Technology Co., Ltd.