



Short-term integrated forecasting method for wind power, solar power, and system load based on variable attention mechanism and multi-task learning

Han Wang^a, Jie Yan^{a,*}, Jiawei Zhang^a, Shihua Liu^a, Yongqian Liu^a, Shuang Han^a, Tonghui Qu^b

^a State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (NCEPU), School of New Energy, North China Electric Power University, Beijing, 102206, China

^b Powerchina Jilin Electric Power Survey and Design Institute Corporation Limited, Changchun Jilin province, 130022, China

ARTICLE INFO

Handling editor: Henrik Lund

Keywords:

Wind and solar power
System load
Integrated forecasting
Variable attention mechanism
Multi-task learning
Coupling relationship

ABSTRACT

Improving the forecasting accuracy of wind power, solar power, and system load to support the source-load cooperative dispatch is an important direction to reduce the uncertainty at source and load sides. The current research mainly focuses on a single object, ignoring the interactive coupling relationship among them, which limits the improvement of forecasting accuracy. Therefore, this paper proposes a short-term integrated forecasting method of wind-solar-load. Firstly, a feature extraction module of linkage characteristics of wind-solar-load is built based on variable attention mechanism. Secondly, a multi-task learning model that can automatically calculate the optimal loss weights for different forecasting tasks is constructed to simultaneously accomplish the wind and solar power forecasting tasks through Fully Connected Neural Network. Finally, a load forecasting model which fuses historical load and power forecasting information is established based on Long Short-Term Memory. The operation data of 8 wind farms and 6 solar plants, and the load data of a nearby city are used for instance analysis. The results show that the power forecasting error (root mean square error) of each wind farm, solar plant, and system load is reduced by 4.84 %, 1.86 %, and 3.02 % on average, respectively, compared with the corresponding traditional methods.

1. Introduction

1.1. Motivation

The energy systems in most countries are in a period of clean, low-carbon, and smart transition to achieve the goal of carbon neutrality; the structure of the power system is changing dramatically. On the source side, a high proportion of renewable energy will become the key feature of the power system, and wind and solar power with randomness and volatility will become the primary power source [1]. On the load side, electricity consumption will become diversified, intelligent, and active [2]. The core problem of power system operation is coping with the increasing double uncertainty at both source and load sides [3].

Accurate and reliable forecasting results of wind power, solar power, and system load can effectively reduce the adverse impact of their uncertainty, providing critical information to support the safe and economic operation of the power system [4–6]. However, the increasing proportion of wind and solar power on the source side and the increasing

amount of smart electricity consumption on the load side not only put forward higher requirements on the forecasting accuracy of wind and solar power but also bring new challenges for load forecasting [7,8].

As wind power, solar power, and load are closely related to meteorological factors such as wind speed, temperature, irradiance, and relative humidity; a certain interactive coupling relationship exists among the three objects under different operating scenarios of the power system. The forecasting accuracy can be effectively improved if this relationship is considered during the forecasting. However, the dynamic coupling relationship among wind-solar-load has not adequately considered in existing research, making it challenging to address the impact of the increasing uncertainty at both source and load sides of the power system. It is urgent to explore new power forecasting methods to achieve high forecasting accuracy of wind power, solar power, and system load simultaneously.

* Corresponding author.

E-mail address: yanjie_freda@163.com (J. Yan).

Table 1

Summary of different wind power/solar power forecasting methods.

Forecasting methods	Applicable scenarios	Advantages	Disadvantages	Main models
Physical method [9,10]	New power station; Inside the station	Not need much historical data	Need a significant amount of time and computational resources	Computational fluid dynamics model/Solar radiation transmission equation
Statistical method [11–13]	Power station that has been built for some time; Single or regional stations	Less affect by spatial scale	Rely on a large number of historical data; Model's interpretability is poor	ARMA, SVM, ANN, RNN, LSTM et al.

1.2. Literature survey

1.2.1. Short-term wind power/solar power/load forecasting

1) Wind power/solar power forecasting

Short-term wind/solar power forecasting is usually based on Numerical Weather Prediction (NWP) results. The forecasting methods can be divided into physical method and statistical method, according to the modeling principle.

The basic steps of wind power forecasting by physical method are as follows. Firstly, various factors such as terrain and altitude are considered, and a computational fluid dynamics model is established to depict the flow conditions within the wind farm. Then, NWP data is used as input to calculate the wind conditions at the hub height of each wind turbine. Finally, the wind speed is converted into power based on the wind speed-power curve of each wind turbine, and the individual power is added to obtain the total power of the wind farm [9].

The basic steps of solar power forecasting by physical method are as follows. Firstly, NWP data is used as input to calculate the solar irradiance at each photovoltaic array through the solar radiation transmission equation at first. Then, taking the components parameters of the solar cell, such as rated power, working temperature range, and efficiency into account, the photovoltaic component operating equation is established to calculate the power of each photovoltaic array. Finally, the individual power is added to obtain the total power of the solar power station [10].

The statistical method usually establishes the mapping relationship between NWP data and actual power data to obtain forecasting results [11–13]. The wind and solar power forecasting accuracy is directly affected by the accuracy of NWP [14]. The commonly used methods include autoregressive moving average model, support vector machine, artificial neural network, deep learning, combination methods, etc. [15, 16]. Yang et al. proposed a short-term wind power forecasting method based on dynamic and static feature fusion mining, three new statistical features, namely wind speed change rate, dynamic migration and transfer feature are constructed from NWP as model inputs. Bidirectional gate recurrent unit and temporal and spatial attention mechanism are incorporated to improve the forecasting accuracy [17]. Cui et al. presented an improved hybrid model that utilizes long short-term memory by considering wind power ramp events. The proposed model can capture the temporal dynamics of wind power and has better performance than a few existing models [18]. Dewangan et al. explored different combined-forecast methods such as mean, median, linear regression and non-linear regressions for day-ahead solar power forecasting models to improve forecasting accuracy and reduce the computational burden of models [19].

The physical method does not require a lot of historical data and is

Table 2

Summary of different load forecasting methods.

Forecasting methods	Applicable scenarios	Advantages	Disadvantages	Main models
Mathematical equation method [22–24]	Small samples Big data	Not need much historical data; Fast prediction speed	Need to specify the model type before forecasting; Difficult to fit nonlinear varying load	Regression analysis, exponential smoothing, load derivation et al.
Artificial intelligence method [25–27]		Not need to specify mapping relationship between input variables and load in advance	Require a large amount of historical data; Model's interpretability is poor	ANN, CVM, fuzzy prediction et al.

suitable for newly constructed wind/solar power plants. However, model solving requires a significant amount of time and computational resources, its application is usually limited within the power plant. The statistical method relies on a large amount of historical data to obtain the statistical patterns for power forecasting. The forecasting time is less affected by spatial scale and can be applied to both single power plants and regional power plants. The summary of different wind power/solar power forecasting methods is listed in Table 1.

2) Load forecasting

Short-term load forecasting methods are usually based on historical load data and the related meteorological data, then establish the mapping relationship between the drive data and load to obtain forecasting results [20,21]. The forecasting methods can be divided into mathematical equation forecasting and artificial intelligence forecasting, according to whether the forecasting model is a black box model.

Load forecasting based on mathematical equations using non-black box models, the commonly used methods including regression analysis, exponential smoothing, load derivation, etc. [22]. Yin et al. presented a new fractional multivariate grey Bernoulli model for short-term load forecasting, which introduced the nonlinear function and Levy flight to enhance local exploitation and global exploration ability [23]. He et al. proposed a short-term power load probability density forecasting method based on Yeo-Johnson transformation quantile regression using Gaussian kernel function, which can efficiently avoid the quantile crossing problem and obtain more accurate forecasting results [24].

Load forecasting based on artificial intelligence using black box models, the commonly used methods including artificial neural network, correlation vector machine, fuzzy prediction, etc. [25]. Zhang et al. proposed an improved hybrid model for short-term power load forecasting, which used a seasonal autoregressive integrated moving average model to predict conventional subsequences and a deep confidence network to predict random subsequences [26]. Wan et al. combined convolutional neural networks, long short-term memory, and attention mechanisms for short-term power load forecasting, which can address the issue of information loss due to excessively long input time series data, and can be potentially useful for energy companies and policymakers in making informed decisions [27].

Mathematical equation forecasting needs to specify the model type before forecasting. Artificial intelligence forecasting does not need to specify the mapping relationship between input variables and load in advance and can adjust the forecasting model according to the results, which is also a commonly used load forecasting method at present. The summary of different load forecasting methods is listed in Table 2.

Table 3
Summary of main attention mechanisms.

Attention type	Description
Self-attention [30]	Reflects the similarity of elements in the input and output sequences, and assigns different weights to different elements in the learning process
Multi-head self-Attention [28]	Attention weights are calculated separately and their results are concatenated
Spatial attention [32]	Learn the attention weights of different regions adaptively
Temporal attention [33]	Generate attention mask in time and use it to select key frames
Variable attention [34]	Input data is dynamically weighted

As wind power, solar power, and system load are affected by multiple factors, there are many input variables during the forecasting. In order to reduce the complexity, improve the efficiency and performance of the model, some scholars introduce the attention mechanisms in the modeling process. In Ref. [28], an attention temporal convolutional network, which is built on stacked dilated causal convolutional networks and attention mechanisms, is proposed to perform the ultra-short-term spatiotemporal forecasting of renewable resources. Zhang et al. proposed input attention mechanism and hidden connection mechanism to greatly enhance the accuracy and efficiency of RNN-based load forecasting models [29]. Jiang et al. proposed and validated a wind power prediction model, which is a linear network architecture featuring a cosine-related cross attention mechanism [30]. Hou et al. proposed an improved LSTM model based on CNN and attention mechanism to improve the prediction accuracy of solar irradiance [31]. The summary of main attention mechanisms used in wind power, solar power, and system load forecasting is listed in Table 3.

1.2.2. Forecasting methods of new energy and system load

In summary, current research on the short-term forecasting of wind power, solar power, and system load is mainly focused on a single object. The interactive coupling relationship among wind-solar-load is useful to improve the forecasting accuracy. Therefore, some scholars began to study integrated forecasting methods at both source and load sides. The current research can be mainly divided into the following two categories.

- Wind and solar power are regarded as “negative load”, and predict the net load directly [35,36]. Alipour et al. predicted the net load (system load minus wind and solar power) in the short and medium term based on an unsupervised autoencoder and supervised cascade neural network [37]. Van et al. used the cross-validation method to obtain an appropriate covariance function and then made the probabilistic net load forecasting (system load minus solar power) based on the dynamic Gaussian process [38].

This method cannot obtain the forecasting results of different

objects, and the application scenarios are concentrated in the microgrid.

- The multi-input and multi-output model is used to predict the new energy and load simultaneously, and the forecasting results of different objects can be obtained [39]. Laouafi et al. established a backpropagation neural network in each season to predict wind power, solar power, and system load 1 h ahead based on historical data [40]. Quan et al. proposed a lower upper bound estimation method to construct prediction intervals aiming at various problems of traditional methods for prediction interval construction. Then they established neural network-based prediction intervals for load and wind power forecasting [41].

This type of forecasting method is time series forecasting based on historical data. However, compared with the system load, the randomness and volatility of wind and solar power are much stronger, resulting in low forecasting accuracy when only the historical data are used. It is not suitable for short-term forecasting.

The comparisons of existing forecasting methods and the proposed integrated forecasting method are listed in Table 4.

1.3. Our contributions

In order to achieve high-precision forecasting of wind power, solar power, and system load simultaneously under short-term time scale, an integrated forecasting method based on variable attention mechanism and multi-task learning (VAM-MTL) is proposed in this paper. The main contributions of this paper are as follows.

- To reduce the model's complexity due to the increase of forecasting objects, a key inputs extraction module of wind power, solar power, and system load forecasting based on variable attention mechanism is established at first.
- The wind and solar power forecasting tasks are carried out simultaneously based on multi-task learning, and the homoscedasticity uncertainty is introduced to realize the automatic optimization of loss weights for different forecasting tasks, which solves the problem of manual adjustment.
- A short-term load forecasting model considering the interactive coupling relationship between new energy and load is established based on Long Short-Term Memory (LSTM). Besides the historical load data, wind and solar power forecasting results on the predicted day are also taken into the inputs of the load forecasting model to improve accuracy.

1.4. Organization of this paper

The rest of the manuscript is organized as follows. Section 2 proposes an integrated forecasting method of wind-solar-load based on VAM-MTL. Section 3 introduces the evaluation indexes for deterministic

Table 4
Comparisons of existing forecasting methods and the proposed integrated forecasting method.

	Existing forecasting methods				Proposed integrated forecasting method
	Short-term wind/solar power forecasting	Short-term system load forecasting [22,25]	Net load forecasting [32,33] ¹	Multi-input and multi-output forecasting [36]	
Forecasting objects	Single object, wind power or solar power	Single object, system load	Single object, net load	Different objects	Different objects
Application scenario	Source side of the power system (wind farms or solar plants)	Load side of the power system (system load)	Microgrid	Power system	Power system
Forecasting technique	Utilizing weather prediction information	Time series forecasting	Time series forecasting	Time series forecasting	Utilizing weather prediction information
Applicable time scale	Short-term	Short-term	Ultra-short-term, short and medium term	Ultra-short-term, medium term	Short-term
Development approach of wind and solar power	Centrally developed	/	Distributed developed	Centrally developed	Centrally developed

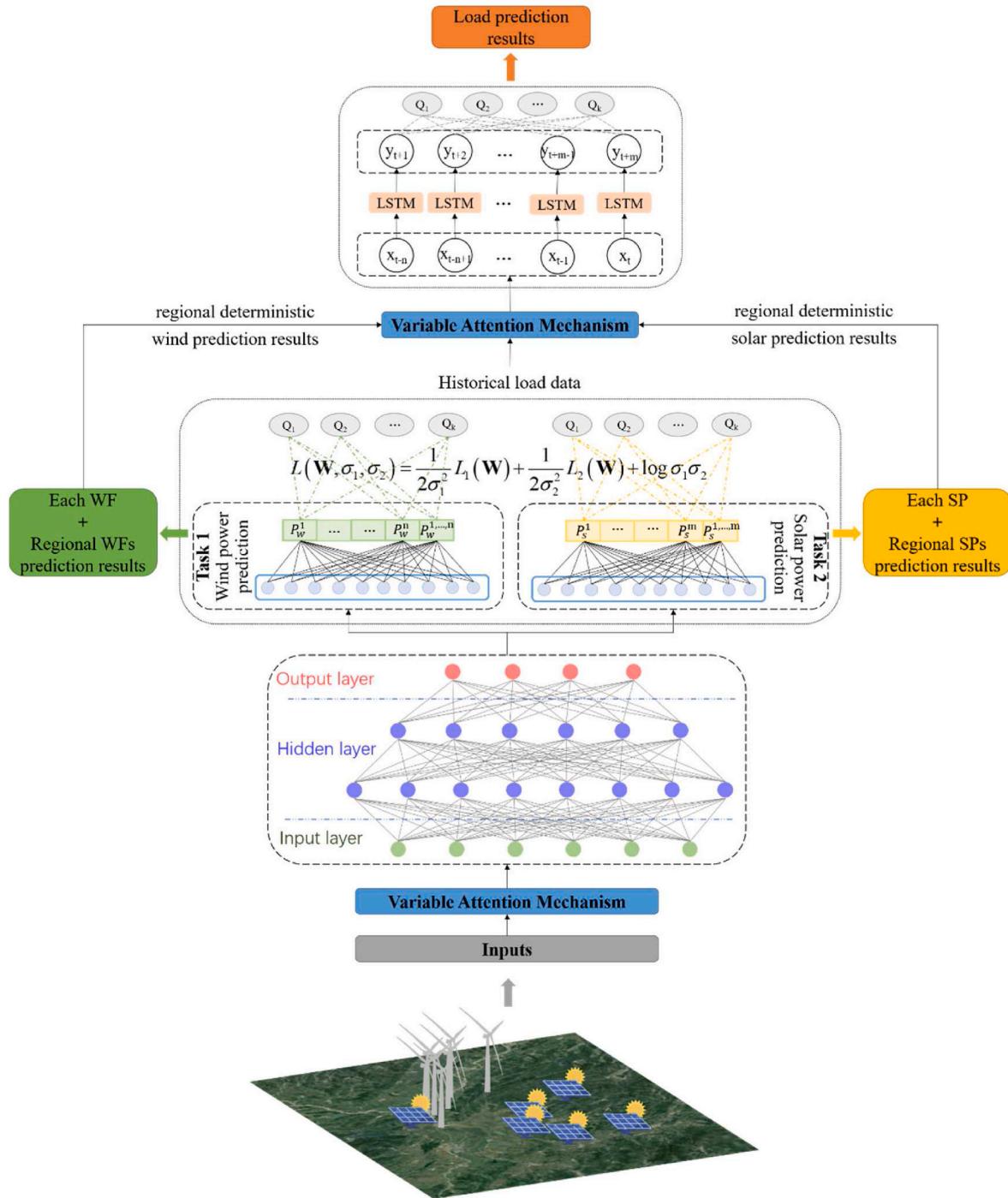


Fig. 1. Structure of the integrated short-term forecasting method of wind power, solar power, and system load.

forecasting and uncertainty forecasting. The operation data of 8 wind farms and 6 solar plants, and the load data of a city near the above new energy stations are used for instance analysis, and the robustness of the proposed method is modified in Section 4. Section 5 concludes the manuscript.

2. Integrated forecasting method based on variable attention mechanism and multi-task learning

An integrated short-term forecasting method of wind power, solar power, and system load based on VAM-MTL is proposed in this section. The model structure is shown in Fig. 1, and the main modeling steps are

as follows.

Step 1: To construct the inputs and outputs of wind and solar power forecasting. The inputs set consists of NWP data for each wind farm (WF) and each solar plant (SP), including wind speed, irradiance, and temperature; the outputs set consists of the power of each WF and each SP, and the total power of regional WFs and regional SPs.

Step 2: To deal with inputs of the forecasting model based on the variable attention mechanism, extracting the valuable input information and reducing the complexity of the model. The details are shown in Section 2.1.

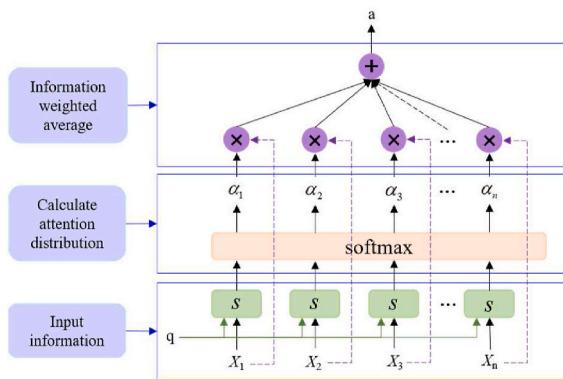


Fig. 2. Process of variable attention mechanism.

Step 3: To establish an integrated forecasting model for wind and solar power based on multi-task learning.

Step 3.1: To set up the shared hidden layer of wind and solar power forecasting using Fully Connected Neural Network (FCNN).

Step 3.2: To set up the output layer of the wind power forecasting task and solar power forecasting task using FCNN based on multi-task learning. Different from the traditional multi-task learning method, where the loss weights of each task are determined manually based on experience or experiments, the loss weights of wind and solar power forecasting tasks are calculated automatically according to homoskedasticity uncertainty in the proposed integrated forecasting method, the details are shown in Section 2.2. The results are more scientific and reliable.

Step 3.3: To set up the uncertainty forecasting module of wind and solar power by quantile regression algorithm.

Step 4: To predict the power of each WF, each SP, regional WFs, and regional SPs using the integrated forecasting model constructed in Step 3.

Step 5: To construct the inputs and outputs of system load forecasting. In addition to the historical load data, the input set also contains the regional WFs power and the regional SPs power on the date to be predicted; the output set consists of the load data on the date to be predicted.

Step 6: To extract the critical information of the inputs of the forecasting model based on the variable attention mechanism.

Step 7: To establish the load forecasting model by applying LSTM, which considers the future information of wind and solar power. Furthermore, the quantile regression algorithm is used to achieve uncertainty forecasting. LSTM is an improved recurrent neural network (RNN), which can solve the gradient disappearance problem of RNN in the long sequence training process.

Step 8: To predict the system load using the forecasting model constructed in Step 7. In the load forecasting process, the inputs are historical load data, and the deterministic forecasting results of regional WFs and regional SPs obtained in Step 4.

2.1. Key inputs extraction based on variable attention mechanism

The inputs and outputs of the forecasting model are increased obviously when wind power, solar power, and system load are predicted simultaneously compared to predicting only one variable, which increases the complexity of the forecasting model and makes the forecasting task more difficult. Therefore, the variable attention mechanism is introduced in this paper to reduce the complexity of the forecasting model, i.e., not all input information is incorporated into the neural network for calculation; only the task-related information is selected and input into the subsequent neural network, as shown in Fig. 2.

The key inputs extraction process of wind power, solar power, and system load forecasting based on variable attention mechanism are as follows.

- 1) Input information, $X = [X_1, X_2, \dots, X_n]$;
 - 2) Calculate the attention distribution α_i , as shown in Eq. (1). α_i represents the attention level of the i th information in context query q .
- $$\alpha_i = \text{softmax}(s(key_i, q)) = \text{softmax}(s(X_i, q)) \quad (1)$$

where $s(X_i, q)$ is the scoring function of attention.

The proposed integrated wind-solar-load forecasting method applied the additive model for scoring:, which is calculated as shown in Eq. (2).

$$s(x_i, q) = v^T \tanh(Wx_i + Uq) \quad (2)$$

- 3) Information weighted average, the soft attention mechanism is used to encode the input information X , as shown in Eq. (3).

$$\text{att}(q, X) = \sum_{i=1}^n \alpha_i X_i \quad (3)$$

2.2. Loss function for wind and solar power forecasting based on homoscedastic uncertainty

Wind power forecasting and solar power forecasting belong to different forecasting tasks, and there are differences between the two forecasting models. If the wind and solar data is directly input into a multiple to multiple forecasting model, it is impossible to build a targeted forecasting model for wind or solar power, making it difficult to achieve high forecasting accuracy. However, the complementarity of wind and solar power in the same region makes their forecasting models relevant. The forecasting accuracy can be improved if wind and solar power are integrated predict. Therefore, a multi-task learning method based on a shared representation can be adapted to wind power forecasting and solar power forecasting. The information sharing between two tasks can promote the learning outcome of each sub-task, thus improving the forecasting accuracy, learning efficiency, and generalization ability of the forecasting model [42,43].

How to realize multi-objective optimization is the key to multi-task learning. At present, the commonly used method is to obtain the total loss by weighting the losses of each task, as shown in Eq. (4), and then optimize the total loss.

$$L_{\text{total}} = \sum_i \omega_i L_i \quad (4)$$

where L_{total} is the total loss function; ω_i is the weight of task i ; L_i is the loss function of task i .

This type of method can convert the multi-objective optimization problem into a single-objective optimization problem, but the learning effect of the model is susceptible to the weights of each task, and the weights are usually set by human experience or experiments based on the values of loss values and convergence speed of different tasks. This way of manually adjusting the weights of each task is time-consuming, laborious, and difficult to obtain weights with good effects for each task. Therefore, the Bayesian modeling thought is introduced in the proposed integrated forecasting method to automatically obtain the optimal weights for each task.

There are two types of uncertainty in Bayesian modeling: 1) Epistemic uncertainty caused by the lack of training data, which can be reduced by increasing the amount of training data. 2) Aleatoric uncertainty caused by the unexplained information in the training data. Specifically, the aleatoric uncertainty can be further divided into data-dependent (heteroskedastic) uncertainty and task-dependent (homoscedastic) uncertainty. Heteroskedastic uncertainty depends on the input data and is reflected in the model's outputs; homoscedastic

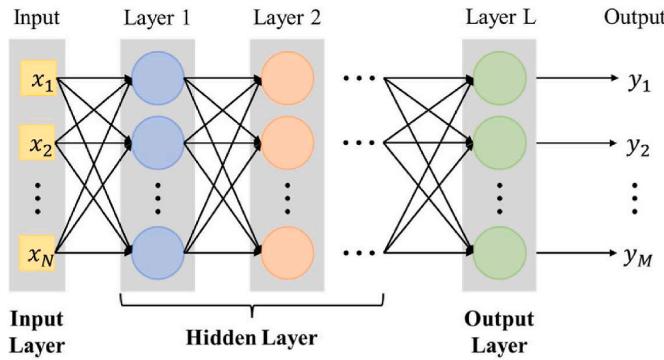


Fig. 3. Structure diagram of FCNN.

uncertainty varies only with the task. The homoscedastic uncertainty can represent the relative difficulty between different tasks and can be used to automatically find the optimal weight of various task losses in multi-task learning [44]. Therefore, we use the homoscedastic uncertainty as a basis for weighting losses in this study.

The process of loss function construction for wind power forecasting task and solar power forecasting task based on homoscedastic uncertainty is as follows.

First, the model's likelihood function is defined in Eq. (5), and the likelihood function for multi-task learning can be obtained, as shown in Eq. (6).

$$p(\mathbf{y}|\mathbf{f}^W(\mathbf{x})) = N(\mathbf{f}^W(\mathbf{x}), \sigma^2) \quad (5)$$

$$p(\mathbf{y}_1, \dots, \mathbf{y}_k | \mathbf{f}^W(\mathbf{x})) = \prod_{i=1}^k p(\mathbf{y}_i | \mathbf{f}^W(\mathbf{x})) \quad (6)$$

where \mathbf{x} is inputs of the model; $\mathbf{f}^W(\mathbf{x})$ is outputs of the model; \mathbf{W} is weights of tasks; σ is the observation noise parameters of the model; $\mathbf{y}_1, \dots, \mathbf{y}_k$ is the outputs of task 1, ..., task k .

Then, the maximum likelihood estimation is transformed into minimizing negative log-likelihood, as shown in Eq. (7).

$$\log p(\mathbf{y}|\mathbf{f}^W(\mathbf{x})) \propto -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{f}^W(\mathbf{x})\|^2 - \log \sigma \quad (7)$$

The outputs of the wind power forecasting task and solar power forecasting task are represented by \mathbf{y}_1 and \mathbf{y}_2 , respectively, and the likelihood function of the two tasks is shown in Eq. (8).

$$\begin{aligned} p(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{f}^W(\mathbf{x})) &= p(\mathbf{y}_1 | \mathbf{f}^W(\mathbf{x})) \cdot p(\mathbf{y}_2 | \mathbf{f}^W(\mathbf{x})) \\ &= N(\mathbf{y}_1; \mathbf{f}^W(\mathbf{x}), \sigma_1^2) \cdot N(\mathbf{y}_2; \mathbf{f}^W(\mathbf{x}), \sigma_2^2) \end{aligned} \quad (8)$$

The loss function of the optimization model $L(\mathbf{W}, \sigma_1, \sigma_2)$ is changed to minimize negative log-likelihood by taking the logarithm of the likelihood function for the two tasks, as shown in Eq. (9).

$$\begin{aligned} L(\mathbf{W}, \sigma_1, \sigma_2) \\ = -\log p(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{f}^W(\mathbf{x})) \propto \frac{1}{2\sigma_1^2} \|\mathbf{y}_1 - \mathbf{f}^W(\mathbf{x})\|^2 + \frac{1}{2\sigma_2^2} \|\mathbf{y}_2 - \mathbf{f}^W(\mathbf{x})\|^2 + \log \sigma_1 \sigma_2 \\ = \frac{1}{2\sigma_1^2} L_1(\mathbf{W}) + \frac{1}{2\sigma_2^2} L_2(\mathbf{W}) + \log \sigma_1 \sigma_2 \end{aligned} \quad (9)$$

where $L_1(\mathbf{W})$ is the loss of wind power forecasting, $L_1(\mathbf{W}) = \|\mathbf{y}_1 - \mathbf{f}^W(\mathbf{x})\|^2$; $L_2(\mathbf{W})$ is the loss of solar power forecasting, $L_2(\mathbf{W}) = \|\mathbf{y}_2 - \mathbf{f}^W(\mathbf{x})\|^2$.

We interpret minimizing this last objective with respect to σ_1 and σ_2 as learning the relative weight of the losses $L_1(\mathbf{W})$ and $L_2(\mathbf{W})$ adaptively, based on the data. As σ_1 (the noise parameter for the variable \mathbf{y}_1) increases, the weight of $L_1(\mathbf{W})$ decreases. On the other hand, as the noise decreases, the weight of the respective objective increases. The noise is

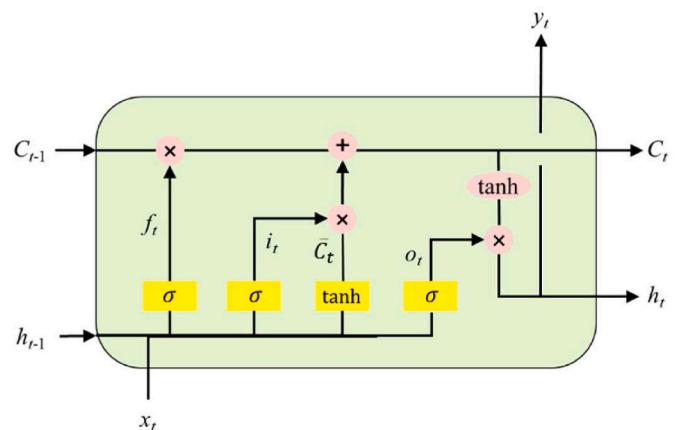


Fig. 4. Structure diagram of LSTM.

discouraged from increasing too much (effectively ignoring the data) by the last term in the objective, which acts as a regularizer for the noise terms.

2.3. FCNN

The structure diagram of FCNN is shown in Fig. 3, including the input layer, hidden layer and output layer. It has a flexible structure that allows the number of neurons to be adjusted as required.

The calculation process of FCNN can be divided into two parts, one is forward propagation of calculated values, the other is reverse propagation of errors [45].

The calculation process of forward propagation is as follows.

$$H = f(W_j X + b_j) \quad (10)$$

$$Y = W_k H + b_k \quad (11)$$

where H, Y is the output of hidden layer and model; W_j, W_k is the weight of the hidden layer and the output layer, and b_j, b_k is the bias.

Error back propagation is the most direct and effective algorithm for updating model parameters, which refers to the process of transferring errors from the output layer to the input layer one by one. First, calculating the error between the model output and the observation values.

$$E = \sum_{j=1}^l (\hat{y}_j - y_j)^2 \quad (12)$$

where \hat{y}_j is the prediction of neural network, y_j is the observation data.

The parameter update in error back propagation is an iterative process, and the parameter W update estimation equation is as follows:

$$W_k = W_k + dW_k \quad (13)$$

$$W_j = W_j + dW_j \quad (14)$$

The changed value in the weights is determined according to the stochastic gradient descent strategy, and for a given learning rate θ , can be calculated as:

$$dW = -\theta \frac{\partial E}{\partial W} \quad (15)$$

According to the chain derivative rule, the weights of output layer and hidden layer satisfy the following equations:

$$\frac{\partial E}{\partial W_k} = \frac{\partial E}{\partial Y} \frac{\partial Y}{\partial W_k} \quad (16)$$

$$\frac{\partial E}{\partial W_j} = \frac{\partial E}{\partial Y} \frac{\partial Y}{\partial H} \frac{\partial H}{\partial f} \frac{\partial f}{\partial W_j} \quad (17)$$

2.4. LSTM

LSTM is an optimized variant of RNN, which adds memory cell and introduces “gate” mechanism to regulate information flow to avoid gradient explosion and gradient disappearance in RNN network [46]. It is mainly composed of input gate, forget gate, output gate and a status update layer, the basic structure is shown in Fig. 4.

The calculation process of LSTM is as follows. Firstly, the values of the forget gate, input gate, output gate and candidate state are calculated according to the external state h_{t-1} at the previous time and the input at the current time. Secondly, the internal state C_{t-1} of the previous moment and the last step is used to calculate the values of forgetting gate, input gate and the candidate state, and update the internal state C_t . Finally, information is passed to the external state h_t through the current internal state and the output gate.

The algorithm of each node in the LSTM is shown as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (18)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (19)$$

$$C_t = f_t \otimes C_{t-1} \oplus i_t \otimes \tanh(W_C \cdot (h_{t-1}, x_t) + b_C) \quad (20)$$

$$o_t = \sigma(W_o \cdot (h_{t-1}, x_t) + b_o) \tanh C_t \quad (21)$$

where, i_t is the output of the input gate at time t ; f_t is the output of the forget gate at time t ; σ is the activation function, $\sigma = 1 / (1 + e^{-x})$; \tanh is a hyperbolic tangent function; o_t is the output of the output gate at time t ; C_t and C_{t-1} are the states at time t and time $t - 1$, respectively.

3. Evaluation indexes

The root mean square error (RMSE) [47] and mean absolute error (MAE) [48] are used as the evaluation index of wind power, solar power, and system load deterministic forecasting. The calculation formulas are shown in Eq. (22) and Eq. (23).

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (P^f(t) - P^m(t))^2}{N}} \quad (22)$$

$$MAE = \frac{\sum_{t=1}^N |P^f(t) - P^m(t)|}{N} \quad (23)$$

where $P^f(t)$ is the normalized value of predicted power at time t ; $P^m(t)$ is the normalized value of actual power at time t .

The skill score is used to evaluate the uncertainty forecasting model, which can assess the reliability and sharpness of the model simultaneously. The skill score value represents the comprehensive performance of the model; the larger value, the stronger performance [49].

The calculation formulas for the skill score $S_c^{(1-\beta)}$ under $(1-\beta)$ confidence level are shown in Eq. (24) and Eq. (25).

$$\xi^{(\alpha)}(t) = \begin{cases} 1, & \text{if } P^m(t) < \hat{q}^{(\alpha)}(t) \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

$$S_c^{(1-\beta)} = \frac{1}{N} \sum_{t=1}^N (\xi^{(\alpha_1)}(t) - \alpha_1) (P^m(t) - \hat{q}^{(\alpha_1)}(t)) + (\xi^{(\alpha_2)}(t) - \alpha_2) (P^m(t) - \hat{q}^{(\alpha_2)}(t))$$

$$\alpha_1 = \frac{\beta}{2}, \alpha_2 = 1 - \frac{\beta}{2} \quad (25)$$

where $\hat{q}^{(\alpha)}(t)$ is the normalized value of predicted power under α -quantile at time t ; $\xi^{(\alpha)}(t)$ is a binary variable, presenting whether the actual power is larger than the predicted power under α -quantile at time t .

4. Case study

4.1. Data description

The operation data of 8 wind farms and 6 solar plants centrally developed in southern China, and the load data of a city near the above stations are used to verify the effectiveness of the proposed integrated short-term forecasting method. The length of data is 1 year, and the time resolution is 15 min. The operation data of wind farms and solar plants include the NWP data, such as wind speed, irradiance, temperature, and the actual power data of stations. The data in the first 18 days per month is used as the training sample, the data in the last 7 days per month is used as the testing sample, and the remaining data per month is used as the verification sample. It is day-ahead power forecasting in this study.

4.2. Correlation analysis

Firstly, the correlation between the actual power of each WF, each SP, and the system load is quantified by using Pearson linear correlation coefficient [50], Spearman rank correlation coefficient [51], and Kendall rank correlation coefficient [52], respectively. The basic principles and calculation methods of these correlation coefficients are as follows.

1) Pearson linear correlation coefficient

Pearson correlation coefficient does not change with the location and scale of variables, and is often used to measure the linear correlation between variables. It is suitable for the data sets that are uniformly sized, continuous, and normally distributed. The value of Pearson linear correlation coefficient is between -1 and 1. The calculation formula is as shown in Eq. (26):

$$r = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (26)$$

where, $\text{cov}(X, Y)$ is covariance between X and Y , σ_X (σ_Y) is standard deviation of X (Y), x_i (y_i) is the data of variable X (Y) at the i th observation.

2) Spearman rank correlation coefficient

Spearman correlation coefficient is a kind of rank correlation coefficient, which measures the correlation direction and closeness of the linear correlation between two variables. The calculation formula is as shown in Eq. (27):

$$r_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)} \quad (27)$$

where, $d_i = R(x_i) - R(y_i)$, $R(x)$ and $R(y)$ are the rank of X and Y , respectively, N is the sample size.

Spearman rank correlation coefficient is less strict on data conditions. As long as the observed values of the two variables are pairs of rating data, or the rating data obtained from continuous variable observation data, regardless of the overall distribution of the two variables and the size of the sample size, Spearman ranking correlation coefficient can be used to study.

3) Kendall rank correlation coefficient

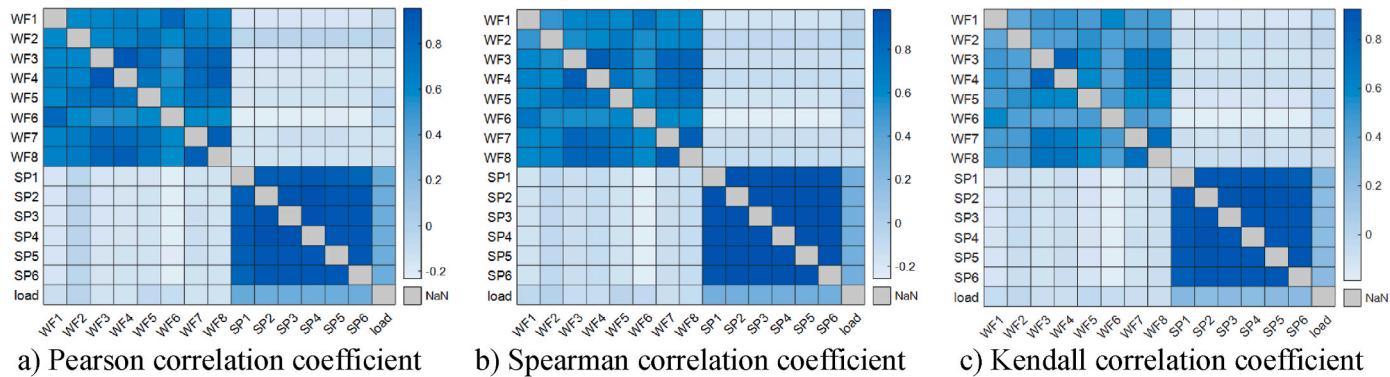


Fig. 5. Correlation analysis of wind farm power, solar plant power, and system load.

Table 5

Average correlation coefficients between wind power, solar power, and system load.

	Pearson			Spearman			Kendall		
	WF	SP	Load	WF	SP	Load	WF	SP	Load
WF	0.693	-0.149	-0.110	0.683	-0.147	-0.088	0.531	-0.108	-0.060
SP	-0.149	0.908	0.323	-0.147	0.975	0.296	-0.108	0.892	0.218
Load	-0.110	0.323	-	-0.088	0.296	-	-0.060	0.218	-

The Kendall rank correlation coefficient is a statistic used to measure the correlation between two pairs of random variables. It is also a rank correlation coefficient, but the object it calculates is a categorical variable.

If we compare two measurement units from our sample (indexed i and j), then any pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if the ranks for both elements agree: i.e. if both $(x_i > x_j \text{ and } y_i > y_j)$ or if both $(x_i < x_j \text{ and } y_i < y_j)$. They are said to be discordant, if $(x_i > x_j \text{ and } y_i < y_j)$ or if $(x_i < x_j \text{ and } y_i > y_j)$. If $(x_i = x_j \text{ and/or } y_i = y_j)$, the pair is neither concordant nor discordant.

When there is no $x_i = x_j$ or $y_i = y_j$ in the set:

$$\tau = 2 \frac{C - D}{N(N - 1)} \quad (28)$$

where C is the number of these pairs that are concordant and D is the number of discordant pairs, N is the sample size.

When there is $x_i = x_j$ or $y_i = y_j$ in the set:

$$\tau = \frac{C - D}{\sqrt{(T_0 - T_1)(T_0 - T_2)}} \quad (29)$$

where, $T_0 = \frac{1}{2}N(N - 1)$, $T_1 = \sum \frac{t_i(t_i - 1)}{2}$, $T_2 = \sum \frac{u_i(u_i - 1)}{2}$; t_i (or u_i) is the number of tied values in the i_{th} group of quantity X (or Y).

Pearson linear correlation coefficient is calculated on the basis of variance and covariance of the original data, so it is sensitive to outliers. Spearman and Kendall rank correlation coefficients are both obtained on the basis of rank and relative size of observed values. They are a more general non-parametric method, less sensitive to outliers, and more tolerant.

Three correlation coefficients are used to quantify the correlation between the actual power of each WF, each SP, and the system load, the results are as shown in Fig. 5 and Table 5. As can be seen.

- (1) The power of each WF and the power of each SP is strongly correlated. The correlation between the power of each SP is stronger than that between the power of each WF.
- (2) Wind power and solar power show a negative correlation.
- (3) Wind power and system load show a weak negative correlation.

The relationship between wind power and load is more in line with linear correlation.

Table 6
Methods of wind and solar power deterministic forecasting.

Basic model	WPF		SPF		Abbr.
	inputs	outputs	inputs	outputs	
SVM	NWP wind speed per WF	wind power per WF	NWP irradiance and temperature per SP	solar power per SP	SVM-s2s
LightGBM					LGBM-s2s
FCNN					FCNN-s2s
SVM	NWP wind speed of WFs	wind power per WF	NWP irradiance and temperature of SPs	solar power per SP	SVM-m2s
LightGBM					LGBM-m2s
FCNN					FCNN-m2s
SVM	NWP wind speed of WFs	wind power of WFs	NWP irradiance and temperature of SPs	solar power of SPs	SVM-m2m
LightGBM					LGBM-m2m
FCNN					FCNN-m2m
FCNN	inputs: NWP wind speed of WFs, NWP irradiance and temperature of SPs		outputs: wind power of WFs, solar power of SPs		VAM-MTL

- (4) Solar power and system load show a positive correlation. The relationship between solar power and load is more in line with linear correlation.

4.3. Results of deterministic forecasting

The results of the proposed integrated short-term forecasting method based on VAM-MTL under 50%-quantile are presented as deterministic forecasting results in this section, and compared with the forecasting results of traditional methods to verify the effectiveness and superiority of the proposed method. The regional wind/solar power forecasting results are obtained by summing the power forecasting results of each WF/each SP.

The benchmark methods of wind and solar power deterministic forecasting are shown in Table 6, consisting of three basic models and

Table 7
Methods of load deterministic forecasting.

Basic model	inputs	outputs	Abbr.
SVM	historical load data	load data on the date to be predicted	SVM
LightGBM			LightGBM
FCNN			FCNN
LSTM	historical load data, regional WFs power and regional SPs power on the date to be predicted	load data on the date to be predicted	VAM-ML

three types of inputs and outputs. The basic models including SVM [53], LightGBM (LGBM) [54], and FCNN. The inputs and outputs of the forecasting models including 1) NWP data of single WF/SP as inputs and power of single WF/SP as outputs, i.e., single to single (s2s). b) NWP data of multiple WFs/SPs as inputs and power of single WF/SP as outputs, i.e., multiple to single (m2s). c) NWP data of multiple WFs/SPs as inputs and power of multiple WFs/SPs as outputs, i.e., multiple to multiple (m2m). The configuration of inputs and outputs is to illustrate the necessity of multi-point weather prediction information in wind and solar power forecasting on the one hand; on the other hand, it is to compare the proposed integrated forecasting method with traditional

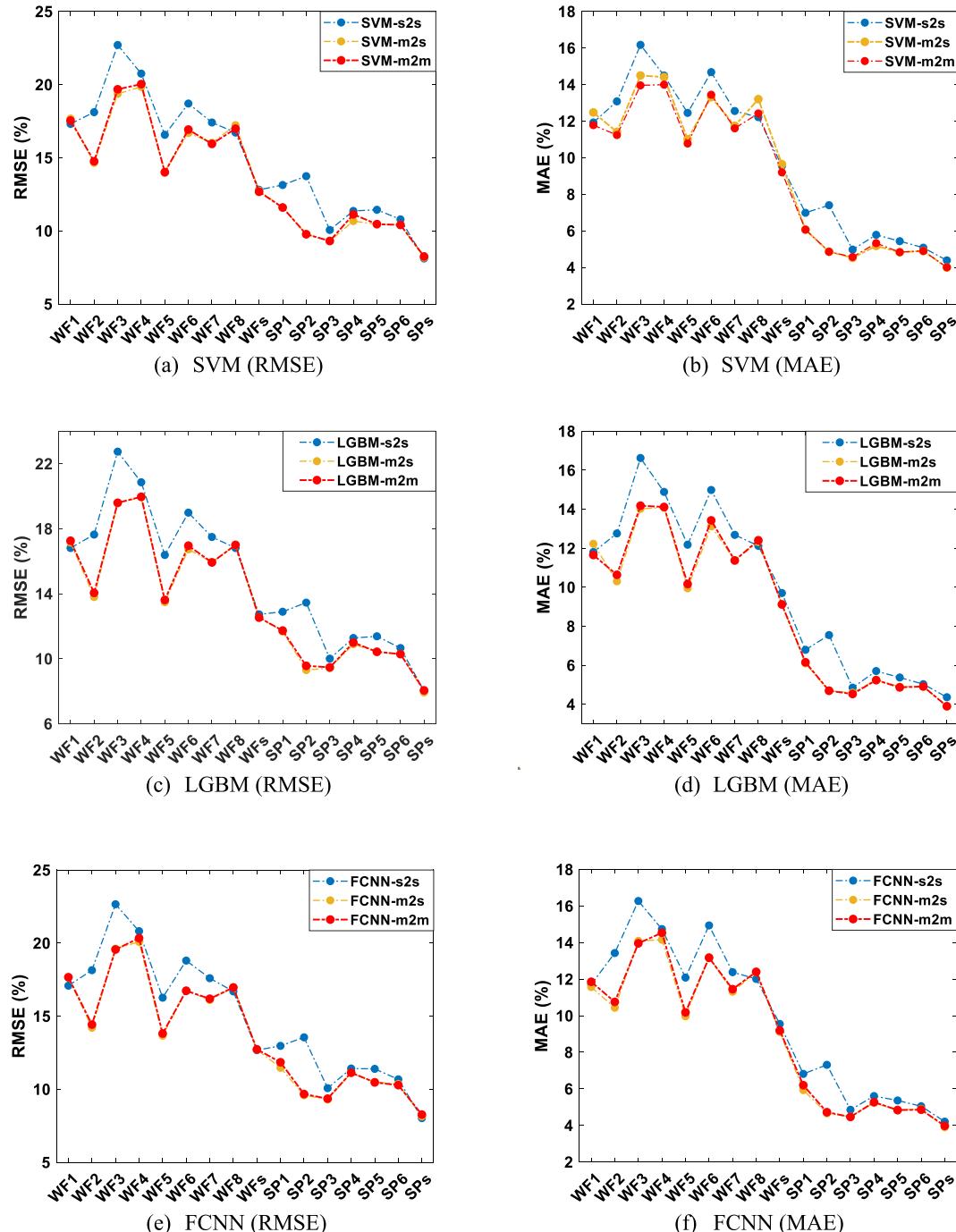


Fig. 6. Wind and solar power deterministic forecasting error with different inputs and outputs.

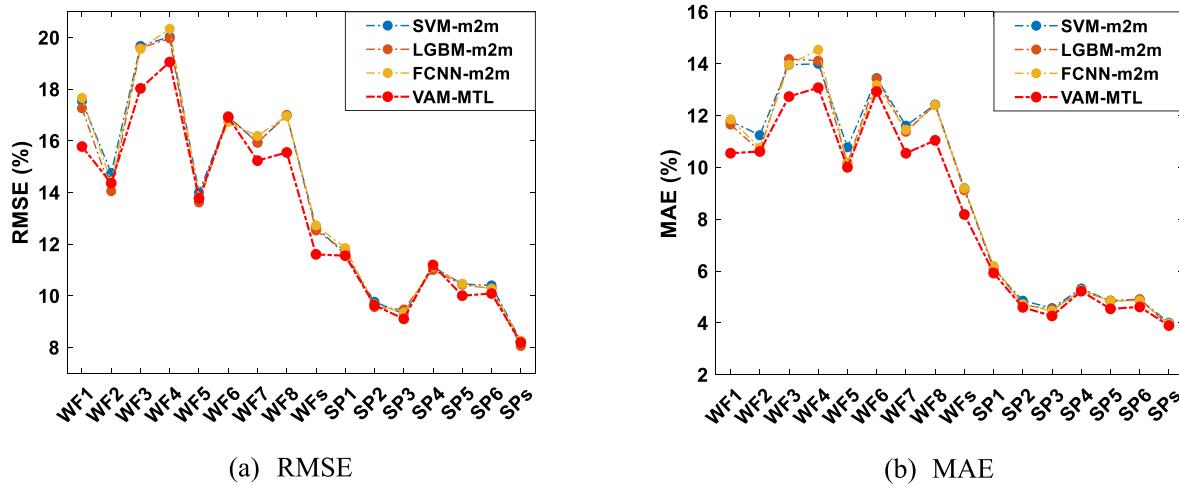


Fig. 7. Comparison of wind and solar power deterministic forecasting error using different methods with multi-input and multi-output.

multi-input and multi-output forecasting method.

The benchmark methods of load deterministic forecasting include SVM, LGBM, and LSTM, which only use the historical load data as the model inputs, as shown in Table 7.

The influence of the spatial scales of inputs and outputs on the forecasting accuracy is analyzed at first. The variation of the deterministic forecasting error of wind and solar power under different inputs and outputs of the forecasting model is depicted in Fig. 6. It can be seen that, under the same basic forecasting model.

- (1) the power forecasting accuracy of each WF/SP is significantly improved when the inputs are NWP data at multiple spatial stations, compared with NWP data at a single station. Besides, the improvement of solar power forecasting accuracy is more obvious. Taking FCNN as an example, compared with FCNN-s2s, RMSE of FCNN-m2s in each wind farm and each solar plant can be maximum reduced by 21.67 % (WF2) and 29.23 % (SP2), with an average decrease of 8.63 % and 10.33 %, respectively; RMSE of FCNN-m2m in each wind farm and each solar plant can be maximum reduced by 20.49 % (WF2) and 28.63 % (SP2), with an average decrease of 8.19 % and 9.78 %, respectively. The forecasting accuracy of multi-input and single-output, multi-input and multi-output are basically identical under the same basic forecasting model. The input information of the forecasting model significantly impacts forecasting accuracy.
- (2) The power forecasting accuracy of regional wind farms/solar plants has little difference under different inputs and outputs. Taking FCNN as an example, RMSE of regional wind power is 12.68 %, 12.72 %, and 12.73 %, MAE of regional wind power is 9.55 %, 9.11 %, 9.19 %; RMSE of regional solar power is 8.02 %, 8.14 %, and 8.27 %, MAE of regional solar power is 4.21 %, 3.90

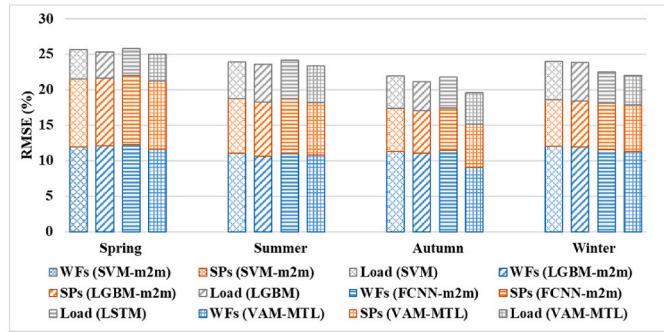


Fig. 8. Seasonal deterministic forecasting error by using different methods with multi-input and multi-output.

%, 3.98 %, under the methods of FCNN-s2s, FCNN-m2s, and FCNN-m2m, respectively.

Secondly, the forecasting error in each wind farm and each solar plant with multi-input and multi-output using different methods are shown in Fig. 7. As can be seen, the proposed integrated forecasting method based on VAM-MTL achieves higher forecasting accuracy in each wind farm, each solar plant, regional wind farms, and regional solar plants compared with the traditional method FCNN-m2m, the RMSE of VAM-MTL in each wind farm can be maximum reduced by 10.67 % (WF1), with an average decrease of 4.84 %; in each solar plant can be maximum reduced by 4.39 % (SP5), with an average decrease of 1.86 %.

The deterministic forecasting error of regional wind power, regional solar power, and system load with multi-input and multi-output using different methods are further summarized, as listed in Table 8. As can be

Table 8
Deterministic forecasting error using different methods with multi-input and multi-output.

		Forecasting error				Error reduction compared with VAM-MTL		
		SVM-m2m	LGBM-m2m	FCNN-m2m	VAM-MTL	SVM-m2m	LGBM-m2m	FCNN-m2m
RMSE	WFs	12.67 %	12.53 %	12.73 %	11.61 %	8.42 %	7.38 %	8.82 %
	SPs	8.25 %	8.06 %	8.27 %	8.19 %	0.67 %	-1.72 %	0.86 %
MAE	WFs	9.20 %	9.12 %	9.19 %	8.18 %	11.07 %	10.26 %	11.02 %
	SPs	4.01 %	3.90 %	3.98 %	3.90 %	2.69 %	0.01 %	1.95 %
Forecasting error								
RMSE	Load	4.91 %	4.78 %	4.63 %	4.49 %	8.55 %	6.07 %	3.02 %
	Load	3.83 %	3.63 %	3.56 %	3.45 %	9.92 %	4.96 %	3.09 %

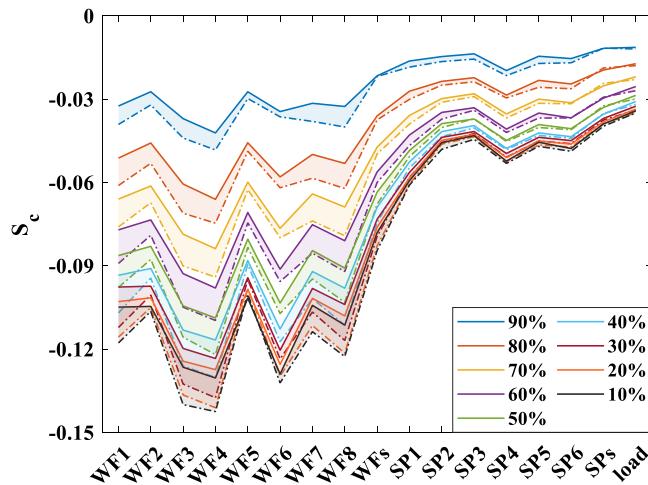


Fig. 9. Uncertainty forecasting results of wind power, solar power, and system load by using the proposed method and FCNN-m2m/LSTM.

Table 9

Uncertainty forecasting results of regional wind power, regional solar power, and system load by using the proposed method and FCNN-m2m/LSTM.

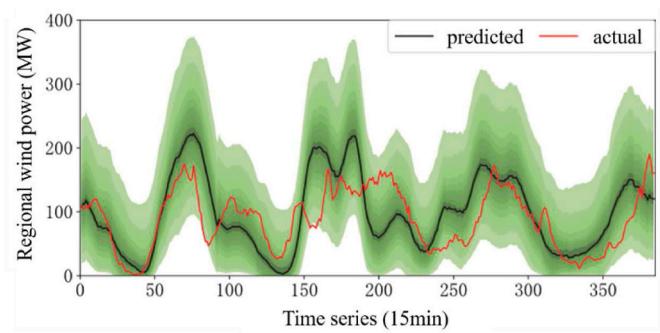
$S_c (10^{-2})$	90 %	80 %	70 %	
WFs	FCNN-m2m	-2.19	-3.75	-4.98
	VAM-MTL	-2.17	-3.61	-4.74
	Error reduction	1.01 %	3.79 %	4.90 %
SPs	FCNN-m2m	-1.17	-1.96	-2.53
	VAM-MTL	-1.18	-1.93	-2.5
	Error reduction	-0.68 %	1.28 %	0.96 %
Load	LSTM	-1.25	-1.81	-2.32
	VAM-MTL	-1.14	-1.73	-2.2
	Error reduction	8.27 %	4.31 %	5.14 %

seen, the proposed integrated forecasting method achieves higher accuracy in wind-solar-load forecasting compared with the traditional methods (except for LGBM-m2m). 1) In wind power forecasting, RMSE and MAE of regional wind power is reduced by 8.82 % and 11.02 % by applying VAM-MTL, compared with the forecasting results of FCNN-m2m. 2) In solar power forecasting, RMSE and MAE of regional solar power is reduced by 0.86 % and 1.95 % by applying VAM-MTL, compared with the forecasting results of FCNN-m2m. 3) In load forecasting, RMSE and MAE of system load is reduced by 3.02 % and 3.09 % by applying VAM-MTL, compared with the forecasting results of LSTM. Compared with the corresponding traditional methods, the forecasting accuracy improvement of the proposed method for different objects is the wind power forecasting accuracy improves most, followed by load, and the solar power forecasting accuracy improves the least.

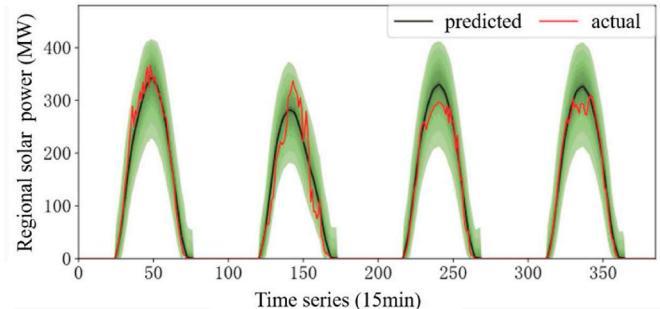
Then, to calculate the seasonal deterministic forecasting error using different methods with multi-input and multi-output, as depicted in Fig. 8. By using the proposed integrated forecasting method, 1) the forecasting accuracy of regional wind power improves the most in autumn, which can improve by 24.25 % on average. 2) The forecasting accuracy of regional solar power improves the most in summer, which can improve by 3.74 % on average. 3) The forecasting accuracy of system load improves the most in winter, which can improve by 20.66 % on average.

4.4. Results of uncertainty forecasting

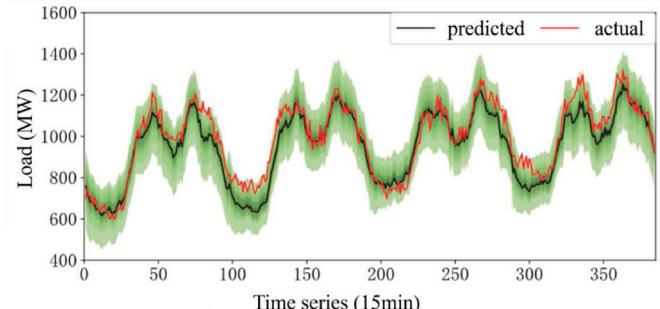
The uncertainty forecasting results by using the proposed integrated short-term forecasting method based on VAM-MTL and the benchmark methods are compared in detail in this section. The benchmark method for wind power uncertainty forecasting is FCNN-m2m, with NWP wind



(a) Regional wind power forecasting time series



(b) Regional solar power forecasting time series



(c) Load forecasting time series

Fig. 10. Wind power, solar power and load forecasting time series by using the proposed method.

speed of multiple WFs as inputs and power of multiple WFs as outputs. The benchmark method for solar power uncertainty forecasting is FCNN-m2m, with NWP irradiation and temperature of multiple SPs as inputs and power of multiple SPs as outputs. The benchmark method for load uncertainty forecasting is LSTM which only takes historical load data as inputs. The uncertainty forecasting is realized through the quantile regression algorithm.

The uncertainty power forecasting results of each wind farm, each solar plant, regional wind farms, regional solar plants, and system load are shown in Fig. 9, where the solid lines represent the proposed method, and the dotted lines represent the corresponding traditional forecasting methods. The skill scores of regional wind power, regional solar power, and system load under three confidence levels, including 90 %, 80 %, and 70 %, are summarized in Table 9.

As can be seen, the proposed integrated forecasting method achieves a better performance of wind power, solar power, and system load uncertainty forecasting at different confidence levels than the corresponding traditional methods. At the confidence level of 80 %, the uncertainty forecasting accuracy of regional wind power, regional solar power, and system load can be improved by 3.79 %, 1.28 %, and 4.31 %, respectively.

Table 10

Deterministic forecasting error by using different methods.

		Forecasting error				Error reduction compared with VAM-MTL		
		SVM-m2m	LGBM-m2m	FCNN-m2m	VAM-MTL	SVM-m2m	LGBM-m2m	FCNN-m2m
RMSE	WFs	15.10 %	14.72 %	15.02 %	14.49 %	4.00 %	1.56 %	3.50 %
	SPs	9.40 %	9.39 %	9.38 %	9.17 %	2.40 %	2.33 %	2.19 %
MAE	WFs	11.43 %	11.01 %	11.10 %	10.58 %	7.40 %	3.88 %	4.72 %
	SPs	4.75 %	4.70 %	4.68 %	4.47 %	5.98 %	4.81 %	4.40 %
		Forecasting error				Error reduction compared with VAM-MTL		
RMSE	Load	SVM	LGBM	LSTM	VAM-MTL	SVM	LGBM	LSTM
	MAE	4.65 %	4.51 %	4.54 %	4.44 %	4.49 %	1.36 %	2.22 %
MAE	Load	3.68 %	3.54 %	3.62 %	3.52 %	4.38 %	0.60 %	2.65 %

Table 11

Uncertainty forecasting results of regional wind power, regional solar power, and system load by using the proposed method and FCNN-m2m/LSTM.

$S_c (10^{-2})$		90 %	80 %	70 %
WFs	FCNN-m2m	-2.63	-4.19	-5.30
	VAM-MTL	-2.50	-4.07	-5.02
	Error reduction	4.85 %	2.79 %	5.26 %
SPs	FCNN-m2m	-1.40	-2.11	-2.60
	VAM-MTL	-1.38	-2.03	-2.61
	Error reduction	1.49 %	3.84 %	-0.28 %
Load	LSTM	-1.15	-1.69	-2.13
	VAM-MTL	-1.07	-1.59	-2.04
	Error reduction	6.87 %	6.05 %	4.35 %

Four days are randomly selected from the integrated forecasting results, as shown in Fig. 10. As can be seen, 1) the variation trend of predicted values is basically consistent with the actual values under each quantile, which directly demonstrates the effectiveness of the proposed method. 2) The forecasting error of wind power, solar power, and system load is gradually decreased, and the forecasting interval is gradually narrowed, but the actual values can be basically included.

4.5. Robustness verification of the proposed method

To verify the robustness of the proposed integrated forecasting method, the operation data of 2 wind farms and 3 solar plants, and the load data of a city near the above stations are used for case study in this part. The length of data is 10 months, and the time resolution is 15 min. The operation data of wind farms and solar plants include the NWP wind speed, NWP irradiance, and the actual power data of stations. The data in the first 18 days per month is used as the training sample, the data in the last 7 days per month is used as the testing sample, and the remaining data per month is used as the verification sample. The forecasting horizon is 24 h in future.

The comparison methods of wind power, solar power, and system load forecasting are the same as Section 4.3 and 4.4. The deterministic forecasting error by using different methods are listed in Table 10. As can be seen, the proposed method achieves higher accuracy in wind-solar-load forecasting compared with the traditional methods. 1) In wind power forecasting, RMSE and MAE of regional wind power is reduced by 3.50 % and 4.72 % by applying VAM-MTL, compared with the forecasting results of FCNN-m2m. 2) In solar power forecasting, RMSE and MAE of regional solar power is reduced by 2.19 % and 4.40 % by applying VAM-MTL, compared with the forecasting results of FCNN-m2m. 3) In load forecasting, RMSE and MAE of system load is reduced by 2.22 % and 2.65 % by applying VAM-MTL, compared with the forecasting results of LSTM.

The skill scores of regional wind power, regional solar power, and system load under three confidence levels, including 90 %, 80 %, and 70 %, are summarized in Table 11. As can be seen, the proposed integrated forecasting method achieves a better performance of wind power, solar power, and system load uncertainty forecasting at different confidence

levels than the corresponding traditional methods.

4.6. Discussion

This paper uses FCNN and LSTM as the basic models in the proposed integrated forecasting method based on VAM-MTL. However, the proposed method is not limited to the basic forecasting models. The theoretical basis for the proposed method to improve the forecasting accuracy is the dynamic coupling relationship among wind power, solar power, and system load in the same area.

Specifically, the improvement of forecasting accuracy is due to the following points. a) A feature extraction module of linkage characteristics of wind-solar-load is built based on variable attention mechanism at first, which can extract the key information effectively and reduce the complexity of the forecasting model. b) The multi-task learning model that can automatically calculate the optimal loss weights for different forecasting tasks is constructed to simultaneously accomplish the wind and solar power forecasting tasks. c) The historical load and power forecasting information are fused in the load forecasting model. The proposed integrated forecasting method has strong robustness to different forecasting models. The application scenarios are concentrated in power system, and the applicable objects are the wind farms and solar plants centrally developed around cities and the system load of cities.

5. Conclusion

According to the interactive coupling relationship of wind power, solar power, and system load, an integrated short-term forecasting method of wind-solar-load based on variable attention mechanism and multi-task learning is proposed in this paper. The proposed method can reduce the number of forecasting models, lessen forecasting workload, and achieve the simultaneous improvement of wind-solar-load forecasting accuracy. The operation data of 8 wind farms and 6 solar plants, and the load data of a city near the above new energy stations are used for instance analysis. The main conclusions are as follows.

- (1) With the same basic forecasting model, inputs, and outputs, i.e., compared to FCNN-m2m, the deterministic power forecasting error (RMSE) of each wind farm and each solar plant is reduced by 4.84 % and 1.86 % on average, RMSE of regional wind power and regional solar power is reduced by 8.82 % and 0.86 %, respectively when using the proposed integrated forecasting method.
- (2) With the same basic forecasting model, i.e., compared to LSTM, the deterministic power forecasting error (RMSE) of load is reduced by 3.02 % when using the proposed integrated forecasting method, which considers the forecasting information of wind and solar power on the predicted day.
- (3) The proposed integrated forecasting method achieves better performance at different confidence levels compared with the corresponding traditional methods (FCNN-m2m for wind and solar power forecasting, LSTM for load forecasting). The uncertainty power forecasting error (S_c) of regional wind power,

regional solar power, and system load is improved by 5.91 %, 1.17 %, and 4.17 %, respectively.

CRediT authorship contribution statement

Han Wang: Writing – original draft, Methodology, Data curation, Conceptualization. **Jie Yan:** Writing – review & editing, Supervision, Conceptualization. **Jiawei Zhang:** Methodology, Formal analysis. **Shihua Liu:** Visualization, Methodology. **Yongqian Liu:** Writing – review & editing, Validation, Resources. **Shuang Han:** Validation, Methodology. **Tonghui Qu:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgement

This work is supported by National Key Research and Development Program of China (2022YFE0117600), State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (Grant No. LAPS24016).

References

- [1] Zhuo Z, Du E, Zhang N, et al. Cost increase in the electricity supply to achieve carbon neutrality in China. *Nat Commun* 2022;13(1):3172.
- [2] Castillo V, De Boer H, Muñoz R, et al. Future global electricity demand load curves. *Energy* 2022;258:124741.
- [3] Huang S, Lu H, Chen M, et al. Integrated energy system scheduling considering the correlation of uncertainties. *Energy* 2023;283.
- [4] Xu S, Wang Y, Xu X, et al. A multi-step wind power group forecasting seq2seq architecture with spatial-temporal feature fusion and numerical weather prediction correction. *Energy* 2024;291:130352.
- [5] Chen Y, Bai M, Zhang Y, et al. Proactively selection of input variables based on information gain factors for deep learning models in short-term solar irradiance forecasting. *Energy* 2023;284:129261.
- [6] Sharma A, Kumar J. A novel seasonal segmentation approach for day-ahead load forecasting. *Energy* 2022;257:124752.
- [7] Tziolis G, Spanias C, Theodoride M, et al. Short-term electric net load forecasting for solar-integrated distribution systems based on Bayesian neural networks and statistical post-processing. *Energy* 2023;271:127018.
- [8] Ye L, Li Y, Pei M, et al. A novel integrated method for short-term wind power forecasting based on fluctuation clustering and history matching. *Appl Energy* 2022;327:120131.
- [9] Castorini A, Gentile S, Gerald E, et al. Increasing spatial resolution of wind resource prediction using NWP and RANS simulation. *J Wind Eng Ind Aerod* 2021; 210:104499.
- [10] Lorenz E, Scheidsteiger T, Hurka J, et al. Regional PV power prediction for improved grid integration. *Prog Photovoltaics Res Appl* 2011;19(7):757–71.
- [11] Yang M, Guo Y, Huang Y. Wind power ultra-short-term prediction method based on NWP wind speed correction and double clustering division of transitional weather process. *Energy* 2023;282:128947.
- [12] Brester C, Kallio-Myers V, Lindfors A, et al. Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations. *Renew Energy* 2023;207:266–74.
- [13] Markovics D, Mayer M. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renew Sustain Energy Rev* 2022;161:112364.
- [14] Wang H, Han S, Liu Y, et al. Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system. *Appl Energy* 2019;237:1–10.
- [15] Li Y, Wang R, Li Y, et al. Wind power forecasting considering data privacy protection: a federated deep reinforcement learning approach. *Appl Energy* 2023; 329:120291.
- [16] Ren Y, Suganthan P, Srikanth N. Ensemble methods for wind and solar power forecasting-A state-of-the-art review. *Renew Sustain Energy Rev* 2015;50:82–91.
- [17] Yang M, Wang D, Zhang W. A short-term wind power prediction method based on dynamic and static feature fusion mining. *Energy* 2023;280:128226.
- [18] Cui Y, Chen Z, He Y, et al. An algorithm for forecasting day-ahead wind power via novel long short-term memory and wind power ramp events. *Energy* 2023;263: 125888.
- [19] Dewangan C, Singh S, Chakrabarti S. Combining forecasts of day-ahead solar power. *Energy* 2020;202:117743.
- [20] Zhao Z, Zhang Y, Yang Y, et al. Load forecasting via Grey Model-Least Squares Support Vector Machine model and spatial-temporal distribution of electric consumption intensity. *Energy* 2022;255:124468.
- [21] Lai C, Wang Yu, Fan K, et al. An improved forecasting model of short-term electric load of papermaking enterprises for production line optimization. *Energy* 2022; 245:123225.
- [22] Song W, Cattani C, Chi C. Multifractional Brownian motion and quantum-behaved particle swarm optimization for short term power load forecasting: an integrated approach. *Energy* 2020;194:116847.
- [23] Yin C, Mao S. Fractional multivariate grey Bernoulli model combined with improved grey wolf algorithm: application in short-term power load forecasting. *Energy* 2023;269:126844.
- [24] He Y, Zheng Y. Short-term power load probability density forecasting based on Yeo-Johnson transformation quantile regression and Gaussian kernel function. *Energy* 2018;154:143–56.
- [25] Yang D, Guo J, Li Y, et al. Short-term load forecasting with an improved dynamic decomposition-reconstruction-ensemble approach. *Energy* 2023;263(Part A): 125609.
- [26] Zhang J, Wang S, Tan Z, et al. An improved hybrid model for short term power load prediction. *Energy* 2023;268:126561.
- [27] Wan A, Chang Q, AL-Bukhaiti K, et al. Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism. *Energy* 2023;282:128274.
- [28] Liang J, Tang W. Ultra-short-term spatiotemporal forecasting of renewable resources: an attention temporal convolutional network-based approach. *IEEE Trans Smart Grid* 2022;13(5):3798–812.
- [29] Zhang M, Yu Z, Xu Z. Short-term load forecasting using recurrent neural networks with input attention mechanism and hidden connection mechanism. *IEEE Access* 2020;8:186514–29.
- [30] Jiang L, Wang Y. A wind power forecasting model based on data decomposition and cross-attention mechanism with cosine similarity. *Elec Power Syst Res* 2024; 229:110156.
- [31] Hou X, Ju C, Wang B. Prediction of solar irradiance using convolutional neural network and attention mechanism-based long short-term memory network based on similar day analysis and an attention mechanism. *Helion* 2023;9(11):110156.
- [32] Tong C, Zhang L, Li H, et al. Attention-based temporal-spatial convolutional network for ultra-short-term load forecasting. *Elec Power Syst Res* 2023;220: 109329.
- [33] Hu Y, Liu H, Wu S, et al. Temporal collaborative attention for wind power forecasting. *Appl Energy* 2024;357:122502.
- [34] Kang H, Kang P. Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism. *Knowl Base Syst* 2024;111507.
- [35] Wen Z, Li Y, Tan Y, et al. A combined forecasting method for renewable generations and loads in power systems[C]. In: 2015 IEEE PES Asia-Pacific power and energy engineering conference (APPEEC); 2015. p. 1–5. Brisbane, QLD, Australia.
- [36] Browell J, Fasiolo M. Probabilistic forecasting of regional net-load with conditional extremes and gridded NWP. *IEEE Trans Smart Grid* 2021;12(6):5011–9.
- [37] Alipour M, Aghaei J, Norouzi M, et al. A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration. *Energy* 2020; 205:118106.
- [38] van der Meer D, Shepero M, Svensson A, et al. Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes. *Appl Energy* 2018;213:195–207.
- [39] Sáez D, Ávila F, Olivares D, et al. Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids. *IEEE Trans Smart Grid* 2015;6(2): 548–56.
- [40] Laouafi A, Mordjaoui M, Dib D. One-hour ahead electric load and wind-solar power generation forecasting using artificial neural network[C]. IREC2015 The Sixth International Renewable Energy Congress, Sousse, Tunisia 2015:1–6.
- [41] Quan H, Srinivasan D, Khosravi A. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Transact Neural Networks Learn Syst* 2014;25(2):303–15.
- [42] Caruana R. Multi-task learning[M]. *Mach Learn* 1997;28:41–75.
- [43] Tan M, Liao C, Chen J, et al. A multi-task learning method for multi-energy load forecasting based on synthesis correlation analysis and load participation factor. *Appl Energy* 2023;343:121177.
- [44] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[CI]. 2018 IEEE/CVF conference. Salt Lake City, UT, USA: On Computer Vision and Pattern Recognition; 2018. p. 1–6.
- [45] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. The MIT Press; 2016.
- [46] Fazlpori Z, Mashhour E, Joorabian M. A deep model for short-term load forecasting applying a stacked autoencoder based on LSTM supported by a multi-stage attention mechanism. *Appl Energy* 2022;327:120063.
- [47] Liu S, Tian J, Ji Z, et al. Research on multi-digital twin and its application in wind power forecasting. *Energy* 2024;292:130269.
- [48] Yu H, Zhong F, Du Y, et al. Short-term cooling and heating loads forecasting of building district energy system based on data-driven models. *Energy Build* 2023; 298:113513.
- [49] Pinson P, Nielsen H, Møller J, et al. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy* 2010;10(6):497–516.

- [50] Rahadian H, Bandong S, Widyotriatmo A, et al. Image encoding selection based on Pearson correlation coefficient for time series anomaly detection. *Alex Eng J* 2023; 82:304–22.
- [51] Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904;15(1):72–101.
- [52] Puth MT, Neuhäuser M, Ruxton GD. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Anim Behav* 2015;102:77–84.
- [53] Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. *IEEE Trans Neural Network* 1999;10(5):1055–64.
- [54] Ke G, Meng Q, Finley T, et al. LightBGM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 2017:30.