# A Pseudo-measurement Modelling Strategy for Active Distribution Networks Considering Uncertainty of DGs

Dongliang Xu, Junjun Xu, Cheng Qian, Zaijun Wu, and Qinran Hu

***Abstract*—Active distribution networks utilize advanced sensors, communication, and control technologies to achieve flexible and intelligent power distribution management. Reliable state estimation (SE) is crucial for distribution management systems to monitor these networks. Historically, the scarcity of measurement resources has hindered the application of SE technology in distribution networks. Establishing a dependable pseudo-measurement model for active distribution networks can significantly enhance the feasibility of SE. This paper proposes a pseudo-measurement model that aligns with the actual operating status of the distribution network, considering the uncertainty in output from distributed generations (DGs) such as wind turbines and photovoltaics. Firstly, it analyzes and models the uncertainty of high-penetration DG output, establishing a reliable output model that incorporates the physical characteristics of wind and photovoltaic output. Secondly, it proposes a pseudo-measurement modeling method based on support vector machine (SVM), where the kernel function of the SVM is weighted according to the information entropy of fluctuations in historical operating data. This weighting ensures that the established pseudo-measurement model better reflects the actual operating status of the active distribution network. Finally, a mathematical model for optimizing pseudo-measurement selection is developed, with the minimum state estimation error as the objective function and the observability of the active distribution network system as the constraint. Case studies demonstrate the accuracy and effectiveness of this approach.**

***Index Terms*—Distribution network, pseudo-measurement, uncertainty of DGs, state estimation, entropy weighting method-support vector machine (EWM-SVM).**

Received: December 3, 2023
Accepted: May 10, 2024
Published Online: September 1, 2024
Dongliang Xu, Cheng Qian, Zaijun Wu (corresponding author) and Qinran Hu are with the School of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: 230198651@seu.edu.cn; 230179581@seu.edu.cn; zjwu@seu.edu.cn; qhu@seu.edu.cn).
Junjun Xu is with the College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: andy-xu199@hotmail.com).
DOI: 10.23919/PCMP.2023.000189

## I. Introduction

For distribution networks with incomplete information, enhancing the quantity and quality of measurement equipment, improving communication conditions, and bolstering data storage and processing capabilities can significantly amplify the monitoring effectiveness of the distribution management system over the distribution network [1]. Concurrently, techniques like data interpolation and model prediction can address data gaps, maximizing data availability and accuracy. Pseudo-measurement modeling technology leverages system models and existing partial measurement data to extrapolate missing measurements, effectively enhancing the completeness of distribution network data [2], [3]. This, in turn, fortifies system monitoring and control capabilities while refining the verification and optimization processes of system models.

Distribution network state estimation necessitates a certain redundancy in measurement data [4], [5]. Given the aforementioned analysis, establishing a pseudo-measurement model that accurately reflects the true operating status of an active distribution network with incomplete information holds paramount importance. This endeavor not only enhances the accuracy of active distribution network state estimation but also ensures the reliable monitoring and control of the active distribution network [6].

When power measurement equipment is not configured at the load buses, obtaining historical load data becomes challenging. Pseudo-measurement models are typically constructed based on typical load curves and real-time measurement data [7], [8]. However, these typical load curves, established through methods such as fault simulation, overall measurement, and statistical synthesis, often involve a significant number of empirical values in the data input process [9], [10]. Consequently, the pseudo-measurement model frequently deviates significantly from the actual operating state of the system.

When historical load data is available, pseudo-measurement modeling schemes based on probabil-

ity analysis and prediction methods are widely utilized. The use of Gaussian process regression (GPR) to establish a pseudo-measurement model is proposed in [11], enhancing the prediction accuracy of pseudo-measurement confidence intervals. In [12], the Gaussian mixture model (GMM) is employed to analyze the probability density of loads and establish pseudo-measurement information with the maximum expected solution. Reference [13] presents a statistically based pseudo-measurement generation method that effectively enhances estimation performance by updating historical measurement data through a Kalman filter. However, methods such as GPR and GMM have high computational costs when dealing with complex nonlinear relationships and are sensitive to the selection of model parameters, which may affect the generalization ability of the pseudo-measurement model.

In recent years, with the continuous advancement of artificial intelligence technology, the robust predictive capabilities of neural networks have provided valuable insights for pseudo-measurement modeling in distribution networks. In [14], researchers devised a pseudo-measurement generation scheme based on artificial neural networks (ANNs). This approach combines typical load distributions to decompose errors associated with the generated pseudo-measurements into several components, making it suitable for weighted least squares (WLS) state estimation. Recurrent neural networks (RNNs) have also been employed in pseudo-measurement modeling, as demonstrated in [5], where RNNs are enhanced through an attention mechanism that analyzes power sequences in both the time and frequency domains at the source and load ends. This analysis leads to the establishment of a period-dependent extrapolation model for characterizing power sequences in these domains. The complex mapping functions in the model are automatically represented by A-RNNs to obtain a period-dependent pseudo-measurement generation model based on A-RNNs. Although such neural network-based pseudo-measurement models perform well in prediction ability, their training process relies on a large amount of annotated data and is prone to overfitting problems. In addition, most of these methods have a certain demand for the computing power of distribution network management systems, and there may be difficulties in popularizing their application in actual distribution management systems.

Additionally, in [15], user-level measurement data is utilized to train gradient boosting tree models for generating pseudo-measurements. Subsequently, the trapezoidal iterative state estimator is applied to these pseudo-measurements to solve the system state, showing potential for enhancing the robustness of distribution network system state estimation. Furthermore, a pseudo-measurement generation method based on distributed generators' characteristics and static voltage control mode is proposed in [16]. Its iterative approach improves the estimation of distributed generators' state and output, thereby enhancing the accuracy of fault current calculation. Although scholars have made many explorations in this field, most existing pseudo-measurement modeling methods lack quantitative analysis of many uncertainties in distribution networks, resulting in unstable performance of the models in practical applications. Moreover, although some studies combine multiple algorithms, the method of integrating multiple algorithms has high complexity and poor real-time performance, which is not conducive to the promotion and application in actual distribution networks.

In light of the above analysis, this paper proposes an advanced pseudo-measurement modeling strategy for active distribution networks. The following are the main contributions of this paper:

1) The pseudo-measurement modeling strategy in this paper considers the output characteristics of wind power and PV systems separately. It establishes output prediction models that account for their uncertainties, reliably providing estimated pseudo-measurement information for distributed generation (DG) buses in active distribution networks.

2) The strategy proposed in this paper innovatively introduces an improved pseudo-measurement generation scheme for conventional load buses based on an entropy-weighted support vector machine (SVM) method.

3) An optimization model for selecting pseudo-measurement sets is also presented, effectively enhancing estimation accuracy.

In the case studies conducted on the testing system, the proposed pseudo-measurement modeling strategy can reliably establish pseudo-measurements that align with the actual operational state of the distribution network. This capability effectively enhances the performance of state estimation in active distribution networks.

The remainder of this paper is organized as follows. Section II presents a pseudo-measurement modeling method considering the uncertainty of DG access buses, while Section III discusses the pseudo-measurement modeling strategy based on EWM-SVM for conventional load buses. Section IV demonstrates the analysis of pseudo measurement errors and the selection of optimization models to enhance the effectiveness of state estimation. In Section V, the effectiveness and reliability of the proposed method were verified through simulation experiments based on modified IEEE-33 bus and IEEE-123 bus system, and Section VI summarizes and prospects the research of the whole paper.

## II. Pseudo-measurement Modelling of DGs Buses Output Uncertainty

In the realm of active distribution networks, the intermittent generation of high-penetration DGs and po-

tential measurement errors from numerous real-time measurement devices can introduce additional uncertainties that must be factored into the state estimation process of these networks [17], [18]. Current theories concerning state estimation in active distribution networks, including those rooted in probabilistic statistical distributions and fuzzy theory, necessitate a substantial number of probability density functions or membership functions for the injected power pseudo-measurements originating from DGs and load buses. Recognizing the uncertainty inherent in DGs' output is vital to ensuring that the established pseudo-measurement models accurately reflect the actual operational state of the distribution network, a factor of significant practical importance [19].

### *A. Wind Power Output Uncertainty Modelling Analysis*

The variability in wind power generation output primarily hinges on wind speed and direction. Leveraging the intrinsic characteristics of wind power generation, along with an extensive dataset encompassing numerical weather forecasts and historical records such as measured wind power output, provides dependable data for pseudo-measurement models aimed at forecasting wind power in the ultra-short/short term.

The kernel-based extreme learning machine (KELM) offers the advantages of fast learning speed, strong generalization ability, and robustness under limited data. Specifically, KELM achieves rapid training speeds by solving a simple linear system rather than relying on iterative optimization. Meanwhile, KELM excels in generalization, effectively managing nonlinearities by mapping input data into a higher-dimensional feature space. Additionally, KELM performs well even with small datasets, making it suitable for scenarios where extensive historical data is unavailable. This study introduces a method for predicting wind turbine output uncertainty based on a double-layer KELM approach [20]. The first layer of the KELM model is adjusted to accommodate variations in wind speed, while the bidirectional output of the second layer delineates the uncertain range of wind turbine output. The schematic of the double-layer KELM structure is depicted in Fig. 1.

The $\boldsymbol{v}_n$ in Fig. 1 represents forecasting wind speed data provided by the meteorological bureau; $\boldsymbol{v}_{t-1}$ represents the wind speed data measured at the moment before the prediction point; $\boldsymbol{v}_{\sin}$ and $\boldsymbol{v}_{\cos}$ are the sine and cosine values of the wind direction, respectively; $L$ and $U$ are the lower and upper bounds of the output interval; $\boldsymbol{v}_{co}$ represents the corrected wind speed by the first layer KELM; and $\rho$ is the local air density.

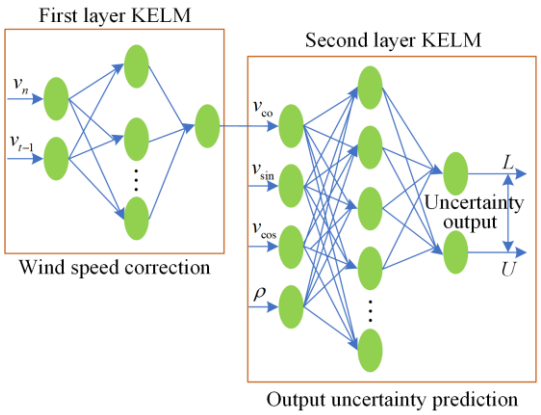


Fig. 1. The structure of the proposed double-layer KELM.

The specific implementation process is as follows.

**Step 1**: Data processing. Collect and process the measured wind turbine output power and tower wind speed obtained from the data collection and monitoring system, along with meteorological data such as wind speed, direction, temperature, humidity, and pressure provided by the local meteorological department. Normalize these data to the range of [0,1]. Thus, the dual input vector $\boldsymbol{I}$ of the double-layer KELM is:

$$\boldsymbol{I}=[\boldsymbol{v}_n,\boldsymbol{v}_{t-1},\boldsymbol{v}_{\sin},\boldsymbol{v}_{\cos},\rho]^{\mathrm{T}} \tag{1}$$

$$\rho=\frac{1276(P-3.78\times10^{-3}hP_{\mathrm{b}})}{1+3.66\times10^{-3}T_{\mathrm{c}}} \tag{2}$$

where $T_{\mathrm{c}}$ and $P_{\mathrm{b}}$ are celsius temperature at the location where the wind turbine is arranged and the saturated vapor pressure value at that temperature, respectively; $P$ and $h$ are the atmospheric pressure and relative air humidity, respectively.

**Step 2**: First layer KELM wind speed correction. A small change in wind speed during the transition from below rated wind speed to rated wind speed can cause significant changes in the output power of wind turbines, necessitating a certain degree of correction to the wind speed forecast data. This paper utilizes the first layer of the KELM network to simulate and correct the non-linear relationship between predicted wind speed and measured wind speed.

**Step 3**: Prediction of uncertainty intervals for wind turbine output. The wind speed value corrected by the first layer KELM, along with the sine and cosine of the wind direction and air density, is used as the input for the second layer KELM to predict the lower and upper bounds of the current wind turbine output as the output for the second city KELM. KELM essentially belongs to a single hidden layer feedforward neural network, and the given raw sample $\{\boldsymbol{x}_i,\ \boldsymbol{T}_i\}$ $(i=1,2,\cdots,N)$, the ELM with a number of nodes with $N$ input layers, $L$ hidden layers can be represented as:

$$\boldsymbol{f}(\boldsymbol{x})=\sum_{i=1}^{L}\boldsymbol{\beta}_i\boldsymbol{H}_i=\boldsymbol{H}\boldsymbol{\beta}=\boldsymbol{T} \tag{3}$$

$$\boldsymbol{H}\begin{bmatrix} g(w_1x_1+b_1) & \cdots & g(w_Kx_1+b_K) \\ \vdots & & \vdots \\ g(w_1x_N+b_1) & \cdots & g(w_Kx_N+b_K) \end{bmatrix} \tag{4}$$

$$\boldsymbol{\beta}=\boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{5}$$

$$\boldsymbol{H}^{\dagger}=\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})^{-1} \tag{6}$$

where $\boldsymbol{H}$ is the hidden layer matrix; $K$ represents the number of the hidden layer neurons; $g(\cdot)$ is the activation function; $b$ is the bias of hidden layer neurons; $\boldsymbol{\beta}$ is the connection weights between hidden layer neurons and output layer neurons and $\boldsymbol{w}$ is the connection weights between the input layer neurons and hidden layer neurons; $\boldsymbol{H}^{\dagger}$ is the generalized inverse.

Introducing kernel functions into ELM and replacing random mappings in ELM with kernel mappings, the kernel matrix is:

$$\boldsymbol{\Omega}_{\mathrm{KELM}}=\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{K}(x_i,x_j) \tag{7}$$

The output of the KELM can be represented as:

$$\boldsymbol{f}_{\mathrm{KELM}}(\boldsymbol{x})=\left[\boldsymbol{K}(x,x_1),\cdots,\boldsymbol{K}(x,x_1)\right]^{\mathrm{T}}\left(\frac{\boldsymbol{I}}{C}+\boldsymbol{\Omega}_{\mathrm{KELM}}\right)^{-1}\boldsymbol{T} \tag{8}$$

In this paper, the kernel functions of the double-layer KELM are all selected as RBF kernels.

### *B. PV Output Uncertainty Modelling Analysis*

The double-output feedforward neural network (DFNN) offers advantages such as handling multiple outputs, complex pattern recognition, flexibility, and scalability. Specifically, DFNNs are particularly suitable for scenarios requiring simultaneous predictions of multiple variables, such as PV outputs, providing a comprehensive modeling approach. Additionally, DFNNs can model complex, nonlinear relationships between inputs and outputs due to their deep structure, capturing intricate dependencies in the data. DFNNs are also flexible and can scale with the complexity of the dataset, enabling the model to adapt to various types of input data and patterns found in renewable energy systems.

In this study, we address the uncertainty associated with PV output and devise a DFNN to delineate the upper and lower bounds of the PV generation system's output [21]. Leveraging the distribution of PV output interval values, we propose interval evaluation metrics, which are further enhanced through integration with the particle swarm optimization (PSO) algorithm [22]. This comprehensive optimization approach fine-tunes the indicators, leading to the determination of optimal output weights for the neural network model. As a result, we achieve precise interval prediction for the PV generation system's output.

As shown in Fig. 2, $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)$ is the input variable, $n$ represents the number of input layer neurons; $\boldsymbol{U}=(u_1,u_2,\cdots,u_m)$ and $\boldsymbol{H}=(h_1,h_2,\cdots,h_m)$ are the input and output of the hidden layer, respectively, $m$ represents the number of hidden layer neurons; $\boldsymbol{L}=(l_1,l_2)$ and $\boldsymbol{Y}=(y_1,y_2)$ are the input and output of the output layer, respectively, $y_1,y_2$ are the upper and lower bounds of PV output. The weight matrix connecting the input layer $\boldsymbol{W}$ and the hidden layer and the weight matrix connecting the hidden layer and the hidden layer $\boldsymbol{V}$ can be represented as follows:

$$\boldsymbol{W}=\begin{bmatrix} \omega_{11} & \cdots & \omega_{1m} \\ \vdots & & \vdots \\ \omega_{n1} & \cdots & \omega_{nm} \end{bmatrix} \tag{9}$$

$$\boldsymbol{V}=\begin{bmatrix} v_{11} & v_{12} \\ \vdots & \vdots \\ v_{m1} & v_{m2} \end{bmatrix} \tag{10}$$

The outputs of the hidden layer and the output layer can be calculated by the following formulas, respectively.

$$\boldsymbol{h}_j=\boldsymbol{f}(u_j)=\boldsymbol{f}\left(\sum_{i=1}^{n}x_i\omega_{ij}+a_j\right),\ j\in[1,m] \tag{11}$$

$$\boldsymbol{y}_j=\boldsymbol{g}(u_j)=\boldsymbol{g}\left(\sum_{j=1}^{m}h_iv_{jk}+b_k\right),\ k=1,2 \tag{12}$$

where $\boldsymbol{f}(\cdot)$ and $\boldsymbol{g}(\cdot)$ are the activation function of the proposed DFNN, respectively; $a_j$ and $b_k$ are the bias of the hidden layer neuron $j$ and output layer neuron $k$, respectively.

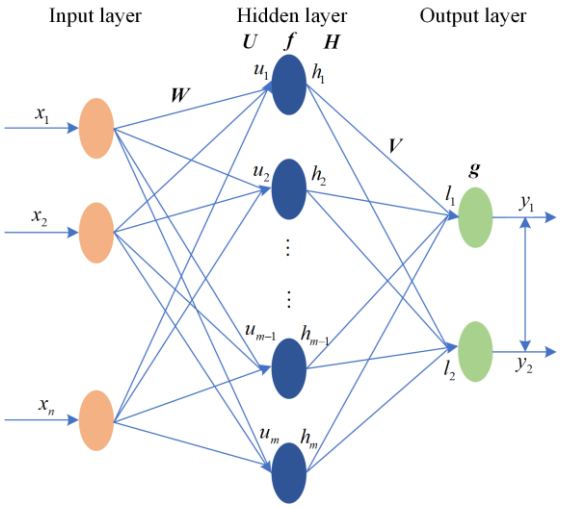


Fig. 2. The structure of the DFNN-based interval prediction for PV generation system output.

To refine the predicted output range from the aforementioned DFNN model, it is crucial to establish corresponding interval evaluation metrics to derive the optimal interval value. Predictive interval coverage probability (PICP) and predictive interval normalized average width (PINAW) serve as crucial indicators for assessing interval prediction performance. In this paper,

$\xi_{\text{PICP}}$ is used to represent the numerical value of PICP, which measures the percentage of the target output encompassed within the prediction interval, reflecting the prediction accuracy [23]. Here is the formal definition of $\xi_{\text{PICP}}$:

$$\xi_{\text{PICP}} = \frac{1}{N}\sum_{i=1}^{N}\tau_i \tag{13}$$

where $N$ is the scale of data; if the output data $y_i$ belongs to the prediction interval $[l_i, u_i]$, $\tau_i = 1$, otherwise $\tau_i = 0$.

$\xi_{\text{PINAW}}$ is used to represent the numerical value of PINAW, which measures measures the average width of the prediction interval, characterizing the uncertainty of the prediction, defined as:

$$\xi_{\text{PINAW}} = \frac{1}{NR}\sum_{i=1}^{N}(u_i - l_i) \tag{14}$$

where $R$ is the target width of the prediction interval.

From a practical application perspective, the ideal prediction interval aims to maximize $\xi_{\text{PICP}}$ while minimizing the $\xi_{\text{PINAW}}$. Achieving these goals concurrently is challenging, primarily because of the inherent trade-off between them. Increasing $\xi_{\text{PICP}}$ typically leads to a broader $\xi_{\text{PINAW}}$, and conversely, reducing $\xi_{\text{PINAW}}$ often results in a decrease in $\xi_{\text{PICP}}$. This paper introduces an interval prediction model, coverage width criterion (CWC), which considers a comprehensive evaluation of both PICP and PINAW indicators, shown as $\xi_{\text{CWC}}$:

$$\xi_{\text{CWC}} = \xi_{\text{PINAW}}(1 + \vartheta e^{-\eta(\xi_{\text{PICP}} - \mu)}) \tag{15}$$

where $\xi_{\text{CWC}}$ can assign a larger weight to $\xi_{\text{PICP}}$, and continue to optimize $\xi_{\text{PINAW}}$ when it meets the confidence level; as $\xi_{\text{PINAW}}$ approaches the expected value, the comprehensive metric also converges to the target width of the prediction interval; $\eta$ and $\mu$ are a pair of hyperparameters that determine the punishment given to the lower prediction interval of PICP; $\mu$ is a confidence level; $\eta$ is an amplification factor that amplifies the differences between the actual interval coverage PICP from the expected interval coverage $\mu$; when $\text{PICP} \geqslant \mu$, $\vartheta = 0$, and when $\text{PICP} < \mu$, $\vartheta = 1$.

This method simplifies the optimization of interval prediction model indicators by consolidating multiple objectives into one, using the PSO algorithm to iteratively solve the objective function. This iterative process identifies the optimal output weight for the dynamic feedforward neural network (DFNN), enhancing the accuracy of interval predictions for PV output. The key steps for the PV output range prediction process are as follows.

**Step 1**: The historical PV dataset is partitioned into training and testing sets at a predetermined ratio. Factors such as solar irradiance, PV array size, and ambient temperature affect PV output. Historical data on PV output and related meteorological conditions are gathered at consistent intervals and then normalized for use as DFNN training data.

**Step 2**: The training data are segmented into training and validation subsets. Initial weights for input vectors and biases for hidden layer neurons are set randomly. The PV output values in the training set are slightly adjusted to create initial prediction bounds, which are used to determine the initial output weight vector.

**Step 3**: Set the scale and inertia parameters for the PSO algorithm. Use the initial PV output interval from Step 2 as the starting population for the PSO, with the evaluation function from (15) serving as the fitness criterion. Execute several algorithm iterations to find the global optimum-the optimal output weight vector.

## III. Pseudo-measurement Modelling of Load Buses Based on EWM-SVM

SVM regression leverages the power of kernel functions and insensitivity to outliers, demonstrating high tolerance for noisy data and requiring relatively low computational resources. As a result, it excels in nonlinear regression tasks [24].

The core concept of SVM is to map data into a high-dimensional feature space via kernel functions, which increases the likelihood of linear separability. This capability is especially beneficial for creating pseudo-measurements of load buses when actual measurement equipment is not available [25]. This paper presents an information entropy weighting method (EWM) designed to optimize the SVM kernel function. The objective is to construct a pseudo-measurement model that accurately represents the true state of load buses.

According to the analysis presented in [24], the SVM algorithm employs the Lagrangian method to transform the aforementioned optimization problem into a dual form:

$$\begin{cases} \text{obj: } \min \dfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^{n}\alpha_i \\ \text{s.t. } \sum_{i=1}^{n} y_i \alpha_i = 0, \quad 0 \leqslant \alpha_i < c \end{cases} \tag{16}$$

where $K(x_i, x_j)$ is the kernel function used in the SVM algorithm; $\alpha$ is the Lagrangian penalty term and $c$ is the penalty factor.

In this paper, based on the radial basis function (RBF) kernel function and utilizing its advantages of multi-scale kernels, a high-dimensional multi-scale kernel function is designed.

For the RBF kernel function $\exp\left(-\dfrac{\| x_i - x_j \|^2}{2\sigma^2}\right)$, choosing different $\sigma$ can be applied to multi-scale scenarios, and the multi-scale RBF kernel function set $\boldsymbol{S}_{\text{exp}}$ can be represented as follows:

$$\boldsymbol{S}_{\text{exp}} = \left[\exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma_1^2}\right), \cdots, \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma_n^2}\right)\right] \tag{17}$$

where $\sigma_1 < \sigma_2 < \cdots < \sigma_n$. When the distribution experiences significant fluctuations, employing a smaller value of $\sigma$ for multi-scale RBF can yield superior

output outcomes. Conversely, when the distribution fluctuation is minimal, utilizing a larger $\sigma$ for multi-scale RBF can lead to better output results.

Weighted synthesis is applied to kernel functions of varying scales, resulting in an enhanced kernel function with adaptive weights. The optimization problem, which has been transformed into a dual form using the Lagrangian algorithm, can be expressed as:

$$\begin{cases} \min \dfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} y_i y_j \alpha_i \alpha_j W_k \exp\left(-\dfrac{\| x_i - x_j \|^2}{2\sigma_k^{\,2}}\right) - \\ \quad \sum_{i=1}^{n}\alpha_i \\ \text{s.t.} \quad \sum_{i=1}^{n} y_i \alpha_i = 0, \quad 0 \leqslant \alpha_i < c \end{cases} \tag{18}$$

Information entropy, a concept from information theory, quantifies the uncertainty of a random variable or the magnitude of information [26]. In information theory, information entropy is used to measure the level of uncertainty regarding the occurrence of an event. Higher information entropy indicates greater uncertainty of a random variable, while lower information entropy suggests reduced uncertainty.

Given a specific support vector, the information entropy linked to the scale of the kernel function can be expressed as follows:

$$I_k = -\ln(m)^{-1}\sum_{j=1}^{m} p_{kj} \ln p_{kj} \tag{19}$$

$$p_{kj} = \frac{\varphi(x_{kj})}{\sum_{j=1}^{m}\varphi(x_{kj})} \tag{20}$$

where $p_{kj}$ is mapped from support vectors by kernel functions, when $p_{kj}$=0, it is defined that $\sum_{j=1}^{m} p_{kj} \ln p_{kj} = 0$; $m$ is the number of samples; $\varphi(x_{kj})$ represents the measurement value of the $k$th indicator in the $j$th sample.

According to the calculated information entropy, the adaptive weight $W_k$ is:

$$W_k = \frac{1 - I_k}{n - \sum_{k=1}^{n} I_k}, k = 1, 2, \cdots, n \tag{21}$$

Generally, in distribution network state estimation (SE), short-term or ultra-short-term load forecasting data from the distribution network serves as pseudo-measurement information. This data is typically derived by analyzing historical operational data of bus loads and external factors such as climate data through regression techniques. To ensure that the pseudo-measurement model aligns closely with the actual operational state of the distribution network, the model proposed in this section not only considers historical bus load data and climate information but also incorporates real-time measurement data of branch power from the SCADA system.

Based on this analysis, the input variables of the pseudo-measurement model using EWM-SVM in this section include real-time power data from SCADA, historical load operation data, date information, and climate data. The output variable aims to provide power prediction information for pseudo-measurement buses and branches.

During the training phase of EWM-SVM, it is crucial to choose the penalty factor corresponding to the kernel function under different adaptive weights. This penalty factor $c$ determines the model's generalization ability and complexity. In this context, a multi-scale RBF kernel function set $\boldsymbol{S}_{\exp}$ with various $\sigma$ values was selected, along with the corresponding adaptive weight set $\boldsymbol{W}$ based on EWM. The penalty factor should align with the multi-scale penalty factor set $\boldsymbol{c}$. The selection of penalty factor set $\boldsymbol{c}$ for the optimization problem of EWM-SVM should correspond to the Cartesian product of the adaptive weight set $\boldsymbol{W}$ and the multi-scale RBF kernel function set $\boldsymbol{S}_{\exp}$. To simplify notation, $\boldsymbol{W} \times \boldsymbol{S}_{\exp}$ is represented by the set $\boldsymbol{\kappa}$. It is evident that the predictive efficacy of EWM-SVM hinges on the selection of two parameters: $\boldsymbol{c}$ and $\boldsymbol{\kappa}$.

In this study, the grid search method is employed to select two parameters, supplemented by cross-validation techniques to prevent overfitting or underfitting of the model. The minimum root mean square error serves as the criterion for identifying the optimal selection of $\boldsymbol{c}$ and $\boldsymbol{\kappa}$ parameters.

Based on the aforementioned analysis, the steps of the pseudo-measurement modelling method based on EWM-SVM are outlined as follows.

**Step 1**: Input data, integrate real-time power data, historical load operation data, date information, and climate information measured by SCADA into specified training input $\boldsymbol{T}_{\mathrm{i}}$, prediction input $\boldsymbol{F}_{\mathrm{i}}$. Simultaneously setting the output vectors for model training and prediction, training output $\boldsymbol{T}_{\mathrm{o}}$, and prediction output $\boldsymbol{F}_{\mathrm{o}}$:

$$\boldsymbol{T}_{\mathrm{i}} = \left[P_{\mathrm{M,TI}}\ Q_{\mathrm{M,TI}}\ P_{\mathrm{B,TI}}\ Q_{\mathrm{B,TI}}\ P_{\mathrm{L,TI}}\ Q_{\mathrm{L,TI}}\ D_{\mathrm{TI}}\ C_{\mathrm{TI}}\ H_{\mathrm{TI}}\right]^{\mathrm{T}} \tag{22}$$

$$\boldsymbol{T}_{\mathrm{o}} = \left[P_{\mathrm{B,TO}}\ Q_{\mathrm{B,TO}}\ P_{\mathrm{L,TO}}\ Q_{\mathrm{L,TO}}\right]^{\mathrm{T}} \tag{23}$$

$$\boldsymbol{F}_{\mathrm{i}} = \left[P_{\mathrm{M,FI}}\ Q_{\mathrm{M,FI}}\ P_{\mathrm{B,FI}}\ Q_{\mathrm{B,FI}}\ P_{\mathrm{L,FI}}\ Q_{\mathrm{L,FI}}\ D_{\mathrm{FI}}\ C_{\mathrm{FI}}\ H_{\mathrm{FI}}\right]^{\mathrm{T}} \tag{24}$$

$$\boldsymbol{F}_{\mathrm{o}} = \left[P_{\mathrm{B,FO}}\ Q_{\mathrm{B,FO}}\ P_{\mathrm{L,FO}}\ Q_{\mathrm{L,FO}}\right]^{\mathrm{T}} \tag{25}$$

where $P_{\mathrm{M,TI}}, Q_{\mathrm{M,TI}}$ are real time measurement of active and reactive power for training inputs, respectively; $P_{\mathrm{B,TI}}, Q_{\mathrm{B,TI}}$ are active and reactive power of the training input branch historical data, respectively; $P_{\mathrm{L,TI}}, Q_{\mathrm{L,TI}}$ are training input bus load historical data active and reactive power, respectively; $D_{\mathrm{TI}}, C_{\mathrm{TI}}, H_{\mathrm{TI}}$ are training input date information, temperature information, and humidity information, respectively; $P_{\mathrm{B,TO}}, Q_{\mathrm{B,TO}}$ are training output of branch active and reactive power, respectively; $P_{\mathrm{L,TO}}, Q_{\mathrm{L,TO}}$ are training output bus load active and reactive power, respectively; $P_{\mathrm{M,FI}}, Q_{\mathrm{M,FI}}$ are forecasting input real time measurement of active and reactive power, respectively; $P_{\mathrm{B,FI}}, Q_{\mathrm{B,FI}}$ are forecasting

input active and reactive power of the branch historical data, respectively; $P_{\mathrm{L,FI}}, Q_{\mathrm{L,FI}}$ are forecasting input bus load historical data active and reactive power, respectively; $D_{\mathrm{FI}}, C_{\mathrm{FI}}, H_{\mathrm{FI}}$ are forecasting input date information, temperature information, and humidity information, respectively; $P_{\mathrm{B,FO}}, Q_{\mathrm{B,FO}}$ are forecasting output of branch active and reactive power, respectively; $P_{\mathrm{L,FO}}, Q_{\mathrm{L,FO}}$ are forecasting output bus load active and reactive power, respectively.

$\boldsymbol{T}_{\mathrm{i}}$ and $\boldsymbol{F}_{\mathrm{i}}$ are derived from the processing of historical data, effectively partitioned into specific datasets for their respective purposes.

$\boldsymbol{T}_{\mathrm{i}}$ is utilized to train the proposed EWM-SVM pseudo-measurement model, leveraging historical data encompassing measurements and observations over a specific period. This dataset aids the model in learning underlying patterns and relationships within the data. $\boldsymbol{F}_{\mathrm{i}}$, on the other hand, is employed for predicting future states and conditions. It includes anticipated scenarios based on historical data, facilitating the model's ability to forecast future measurements.

**Step 2**: Normalize the data to ensure consistency in feature scales, avoid the complexity introduced during the optimization process due to inconsistent feature scales, ensure that the features of different data maintain effective weights in the algorithm, and improve the solving speed and generalization ability of the EWM-SVM algorithm.

**Step 3**: Calculate the information entropy based on the fluctuation of historical operational data, and use it to compute the corresponding adaptive weights for different parameters of the multi-scale kernel function.

**Step 4**: The grid search method is used to initiate the selection of $\boldsymbol{c}$ and $\boldsymbol{\kappa}$ parameters, followed by model training based on the optimal parameters. To prevent overfitting or underfitting of the model, a five-fold cross-validation method is employed for parameter optimization. The dataset is randomly divided into five parts, with one part designated as the validation set and the remaining four parts as the training set. After training and validating the model, the objective is to minimize the root mean square error on the training set to optimize and determine the optimal $\boldsymbol{c}$ and $\boldsymbol{\kappa}$ values. This process establishes the optimal penalty factor set $\boldsymbol{c}$ adaptive weight set $\boldsymbol{W}$ and corresponding multi-scale RBF kernel function set $\boldsymbol{S}_{\mathrm{exp}}$.

**Step 5**: Input prediction data and output prediction results. Provide power prediction information for pseudo-measurement buses and branches that need to be provided.

## IV. Pseudo-measurement Error Analysis and Selection Optimization Model

The pseudo-measurement model based on the EWM-SVM algorithm mentioned above generally includes prediction errors, termed pseudo-measurement errors, when providing pseudo-measurements of node load power and branch power compared to their actual values in active distribution network operation. In traditional distribution network state estimation processes, the weights assigned to these pseudo-measurement data are typically determined based on the reciprocal of the variance of the pseudo-measurement errors. A smaller variance in pseudo-measurement errors generally indicates higher reliability of the pseudo-measurement values, resulting in larger weights. Conversely, a larger variance in pseudo-measurement errors indicates lower reliability of the pseudo-measurement values, leading to smaller weights. This variance-based weight allocation method allows for a more reasoned consideration of pseudo-measurement values in the state estimation process, thereby enhancing the computational accuracy of state estimation. By assigning higher weights to pseudo-measurement values with smaller variances, the system can rely more on these values for state estimation, while pseudo-measurement values with larger variances have reduced impact on state estimation to minimize error propagation.

Given the complex operational dynamics of active distribution networks and the increased uncertainty due to the significant integration of distributed power sources in recent years, the probability density distribution of pseudo-measurement errors exhibits a diverse range that cannot be adequately represented by a standard Gaussian distribution. This article employs a Gaussian mixture model (GMM) to create multiple Gaussian components that conform to normal distributions and assigns varying weights to the pseudo-measurement model derived from EWM-SVM. Essentially, GMM represents the weighted combination of several Gaussian components, expressed mathematically as follows:

$$f(e \mid \zeta) = \sum_{i=1}^{N} \omega_i f(e \mid \mu_i, \delta_i^2) \tag{26}$$

where $\omega_i$ is the weight of the $i$th Gaussian component; $\mu_i, \delta_i^2$ are the mean and variance of the $i$th Gaussian component, respectively; $\zeta$ is the parameters selected for GMM, which is obtained by the maximum expected value method.

The above $N$ Gaussian components can be used to fit the probability density distribution of the time stamped pseudo-measurement error $e_n$ output by EWM-SVM. The correlation coefficient between the pseudo-measurement error $e_n$ and the $i$th Gaussian component can be represented by the edge density function:

$$c(e_n \mid \zeta_i) = \frac{\omega f(e_n \mid \mu_i, \delta_i^2)}{\sum_{i=1}^{N} \omega_i f(e_n \mid \mu_i, \delta_i^2)} \tag{27}$$

The Gaussian component corresponding to the maximum value of the correlation coefficient $c$ obtained by the edge density function is the pseudo-measurement variance obtained, and the reciprocal of the pseudo-measurement variance can be used as the weight of the pseudo-measurement data at different times.

The selection of pseudo-measurement data significantly impacts the accuracy of subsequent state estimation calculations, thereby influencing the observability of the system. Choosing an appropriate set of pseudo-measurements

is crucial for the practical implementation of state estimation.

In Section II, this paper explores distributed power loads, which can induce significant variations in voltage amplitude and phase angle. Unlike conventional loads that cause minor changes, these distributed loads require special attention. Therefore, our focus is on optimizing the accuracy of power estimation data output by thestate estimator. This optimization aims to minimize power estimation errors while ensuring system observability. Ensuring observability involves utilizing the full rank of the Jacobian matrix as a constraint condition in our optimization model. This transformation effectively turns the task of pseudo-measurement set selection into an optimization problem, outlined as follows:

$$\begin{cases} \text{obj:min} \sqrt{\dfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N_B+N_L}\left[(\hat{P}_{j,t}-P_{j,t})^2+(\hat{Q}_{j,t}-Q_{j,t})^2\right]} \\ \text{s.t.} \quad \text{rank}(\boldsymbol{H})=2N_n-1 \end{cases} \tag{28}$$

where $\hat{P}_{j,t}$ and $\hat{Q}_{j,t}$ are the estimated values of the $j$th active power and reactive power in the $t$th state estimation, respectively; $P_{j,t}$ and $Q_{j,t}$ are the true values of the power flow for the $j$th active power and reactive power in the $t$th state estimation, respectively; $T$, $N_B$ and $N_L$ are the total number of state estimates, the number of system branches, and the number of load buses, respectively; $N_n$ is the number of buses in the system; $\boldsymbol{H}$ is the Jacobian matrix, and rank ($\boldsymbol{H}$) represents the rank of the Jacobian matrix.

The optimization model described above exhibits strong nonlinear characteristics, posing computational challenges for traditional mathematical methods in solving planning problems. To address these challenges, enhance solution efficiency, and improve the accuracy of optimization results, the football team training algorithm (FTTA) proposed in [27] is employed. FTTA is a metaheuristic algorithm that incorporates global optimization, constraint handling, and local optimization techniques.

Based on the analysis presented, the overall flowchart of the pseudo-measurement modeling strategy, which considers DG uncertainty, is depicted in Fig. 3.

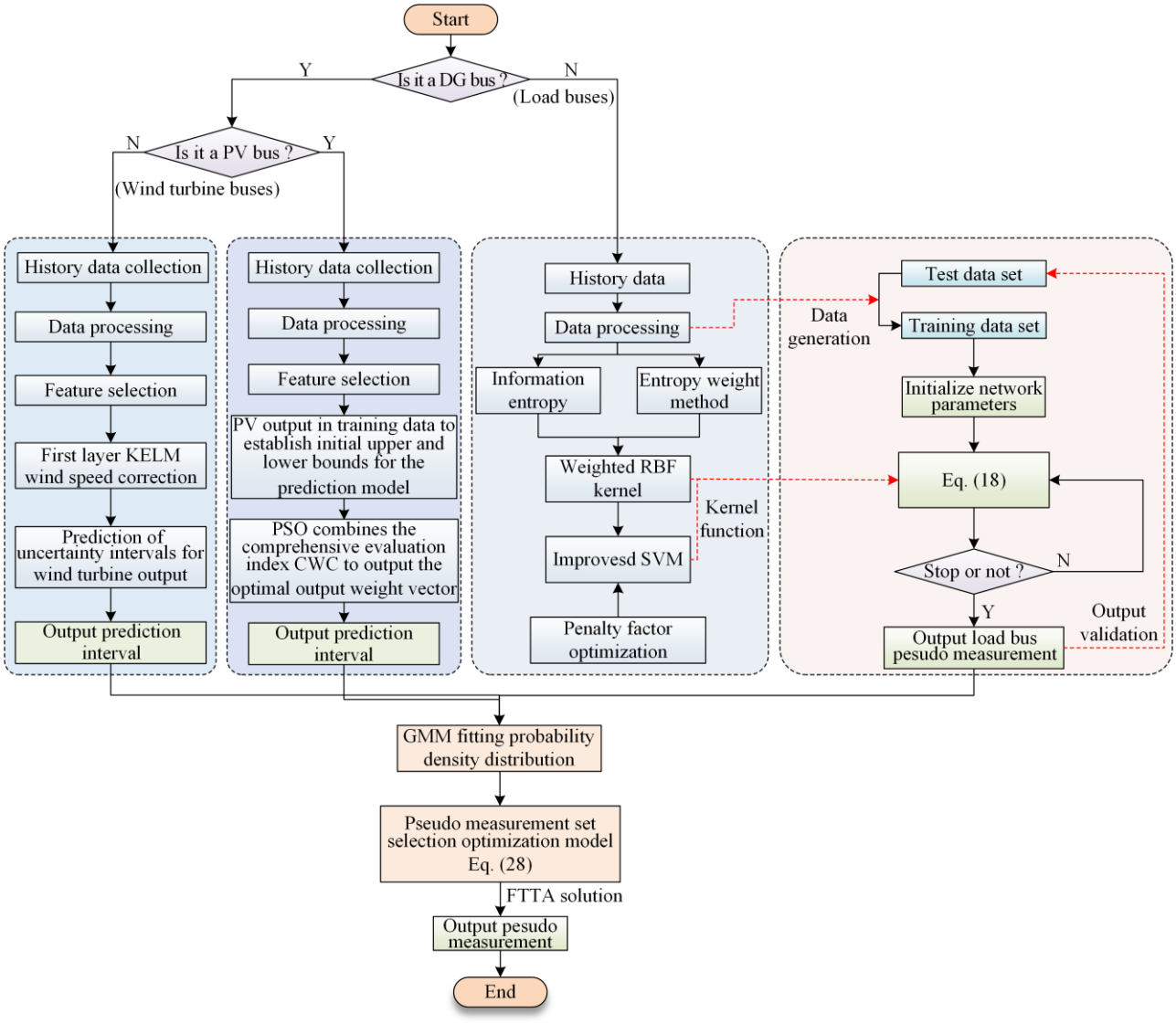


Fig. 3. Overall flowchart of the pseudo-measurement modeling strategy.

## V. Case Study

In order to verify the effectiveness of the proposed pseudo-measurement modelling strategy, a modified IEEE 33-bus distribution network system was employed in this paper. The system topology, measurement layout, and DGs access are shown in Fig. 4 and Table I.

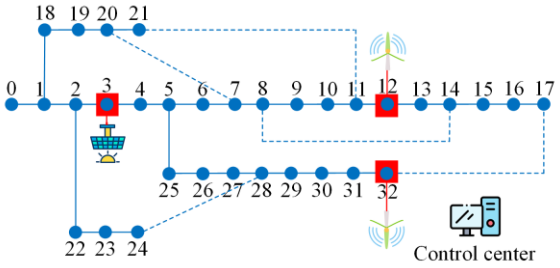


Fig. 4. IEEE 33-bus test system.

TABLE I
Measurement Arrangement of Testing System

| Equipment type | Arrangement buses |
|---|---|
| μPMU | 0 |
| SCADA | 1–18, 2–22, 23–24, 20–21, 5–25, 5–6, 27–28, 29–30, 8–9, 14–15, 16–17 |

All algorithms are implemented on a Ryzen R7-5800 H 3.2 GHz CPU and 32 G RAM laptop based on Matlab R2019b.

### *A. Analysis of DGs Buses Pseudo Measurement Model*

In this case, the temperature, direct radiation, scattered radiation, and total radiation of the surrounding environment of the photovoltaic power generation system, along with measured data of wind speed, temperature, pressure, etc., around the wind turbine power generation system, serve as the parameter sources for predicting the uncertainty in the output of both the photovoltaic power generation system and the wind turbine power generation system in a certain area of East China during the summer of May (with a sampling interval of 15 min) [31]. Additionally, it incorporates measured data such as wind speed, temperature, pressure, and other relevant factors around the wind turbine power generation system. These parameters form the basis for developing an interval prediction algorithm for both PV and wind turbine power generation outputs, with a sampling interval of 15 min.

#### *1) Wind Power Output Uncertainty*

The wind turbine power generation system, with an installed capacity of 1000 kW, utilizes historical measured data spanning the past 29 days. This includes power output values collected from the wind turbine system and meteorological data such as wind speed and other variables around the system during the same period. Additionally, forecasted data for the next day, normalized across different meteorological variables, serves as initial training samples for the first layer of the KELM network. The first layer of the KELM network is designed to simulate the corrected wind speed at the forecast point.

Moreover, the second layer of the KELM network incorporates wind direction cosine and sine values from the past 29 days, along with air density values, as input parameters. For the KELM network, initial input weights are randomly assigned, and the hidden layer consists of 100 neuron nodes using the Sigmoid activation function and employing the RBF kernel.

The interval prediction method for the wind turbine power generation system generates output curves at various time intervals within the next day, as depicted in Fig. 5. The figure illustrates that the upper and lower bounds of the wind power output prediction effectively cover the dispersed range of true values, demonstrating high prediction accuracy.

In this study, the offline training of the KELM model typically requires 20–30 min. Subsequently, we conducted 50 efficiency tests, finding that the average time consumption for online prediction on the experimental platform is 4.71 s.

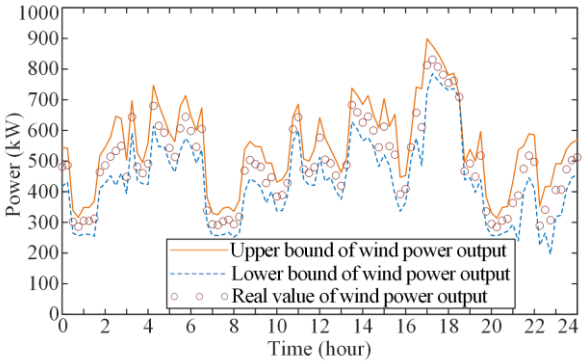


Fig. 5. Predictive curve of typical daily wind turbine power generation system output range.

#### *2) PV Output Uncertainty*

The installed capacity of the PV power generation system is 500 kW. Historical measured data from the past 29 days includes power output values from the PV system, environmental temperature values, and solar radiation values around the system. These data, along with the forecasted data for the next day, are used as training data for the DFNN model.

To ensure accuracy and avoid interference caused by differences in the scale of temperature and radiation values, all input variables are standardized to range between –1 and 1. The input weights and hidden layer biases of the DFNN model are randomly assigned, with the hyperbolic tangent Sigmoid function chosen as the activation function.

In the PSO process, a population size of 100 is selected with a maximum iteration limit of 1000. The learning factors are set to 2, and the weight is set to 0.5.

The interval prediction method for the PV power generation system roughly generates curves of the sys-

tem's output at different time intervals within the next day, as shown in Fig. 6. It can be observed that the upper and lower bounds for predicting PV output also encompass the scattered distribution area of true values, indicating high prediction accuracy.

The DFNN model training proposed in this paper, like wind power prediction, is also conducted offline, usually taking about 30 min. After completing offline training, we also conducted 50 efficiency tests. The average time consumption for online prediction on the experimental platform in this article is 6.14 s.

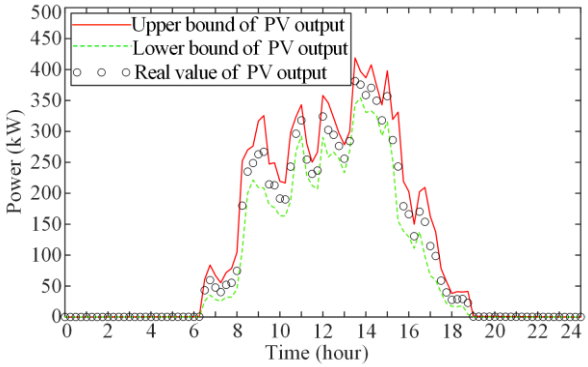


Fig. 6. Predictive curve of typical daily PV generation system output range.

### *B. Efficiency Analysis of the Load Buses Pseudo Measurement Model*

To verify the effectiveness of the proposed EWM-SVM method in pseudo-measurement modeling of load nodes, this paper utilized a power load dataset representing 1000 residential units over the course of one year to simulate distribution system operations [28], [29]. Specifically, within the IEEE-33 bus distribution network, there are 32 sets of load information, each containing approximately 31 (1000/32) residential electricity loads.

The pseudo-measurement information for the load buses is generated at the start of state estimation and updated 96 times per day, in accordance with the standard state estimation frequency (every 15 min).

To assess the effectiveness of the proposed EWM-SVM method against existing methods, three models were trained using the residential load information described earlier, obtained through power flow calculations: the EWM-SVM method proposed in this paper, the SVM method, and the ANN method from the literature [14]. Using load bus 26 in the system without measurement devices as a case study, Fig. 7 presents a comparison between the active power provided by the pseudo-measurement model and the actual values.

As depicted in Fig. 7, it is evident that the outputs generated by the pseudo-measurement methods based on ANN and SVM exhibit considerable deviations from the actual power flow values. Upon closer examination in the enlarged view, the EWM-SVM method proposed in this paper, utilizing a dynamic kernel function accounting for information entropy in historical data, demonstrates outputs closer to the actual values compared to conventional SVM methods. In order to demonstrate the overall performance of the three trained models, in this paper, 500 times tests have been conducted and root mean square error (RMSE) is selected as a comparison parameter, which is:

$$r_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_i)^2} \tag{29}$$

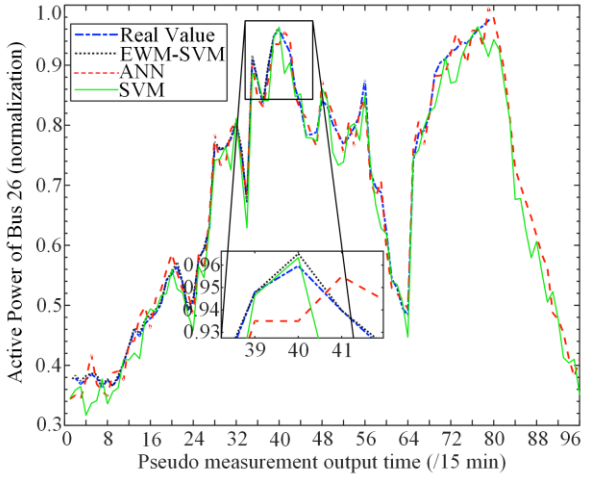


Fig. 7. The effectiveness comparison of different pseudo-measurement models at the load buses.

The RMSE and efficiency comparison of the three models' pseudo-measurements are shown in Table II. Efficiency comparison only considers the time consumption of the online prediction part, as model training is conducted offline. Therefore, this paper temporarily excludes consideration of offline training efficiency. Despite potential performance degradation of the proposed method during sudden load changes, it still offers significant advantages over the other two models. In terms of prediction efficiency, the ANN method exhibits the lowest efficiency among the three comparison methods. The average time consumption of the basic SVM method is lower than that of the method proposed in this paper because the proposed method requires an additional process of calculating information entropy to adjust the weights of the multi-scale RBF kernel function. Although this slightly reduces prediction efficiency, it effectively enhances prediction accuracy.

TABLE II
ACCURACY AND EFFICIENCY COMPARISON OF PSEUDO-MEASUREMENT FROM DIFFERENT MODELS

| Method | Max RMSE ($\times 10^{-4}$) | Min RMSE ($\times 10^{-4}$) | Average RMSE ($\times 10^{-4}$) | Average time consumption (s) |
|---|---|---|---|---|
| ANN | 77.19 | 24.87 | 55.18 | 11.56 |
| Basic SVM | 89.43 | 2.57 | 61.38 | 4.59 |
| Proposed method | 26.15 | 0.83 | 5.29 | 5.72 |

The histogram depicting the error probability density distribution of the pseudo-measurement load model using the EWM-SVM method proposed in this paper is shown in Fig. 8. Based on the histogram distribution, a 5th-order GMM model was selected for fitting. Table III provides the parameters of the fitted GMM model. The pseudo-measurement variance of each load bus can be

determined through Gaussian component analysis of the established 5th-order GMM model, facilitating the determination of appropriate pseudo-measurement weights.

TABLE III
THE PARAMETERS OF EACH GAUSSIAN COMPONENT IN GMM

| Parameter | GMM 1 | GMM 2 | GMM 3 | GMM 4 | GMM 5 |
|---|---|---|---|---|---|
| Mean | –0.0894 | –0.0641 | –0.0283 | 0.0084 | 0.0573 |
| Covariance ($\times 10^{-5}$) | 1.3431 | 1.9422 | 12.161 | 21.441 | 6.7362 |
| Weight | 0.0299 | 0.0543 | 0.3843 | 0.4234 | 0.1081 |

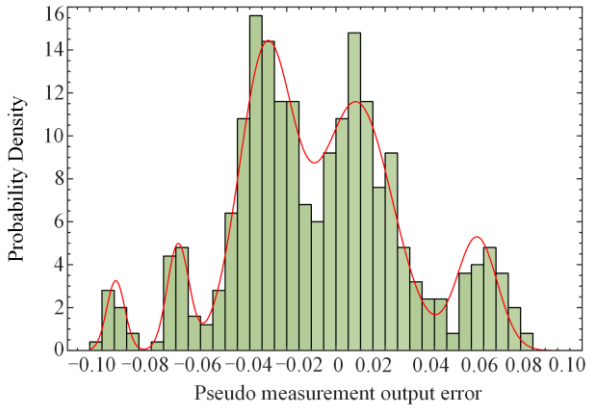


Fig. 8. Schematic diagram of GMM fitting pseudo-measurement error.

### *C. Analysis of the Selection Optimization and Estimation Results*

To verify the effectiveness of the proposed optimization model, all generated pseudo-measurement data will be input into the optimization model. This paper focuses on pseudo-measurement modeling, employing basic WLS for static state estimation of distribution networks in its experiments. Detailed discussions on interval state estimation, distributed estimation, and forecasting-aided state estimation are deferred for future consideration. Therefore, considering the uncertainty in the output intervals of DGs, the mean of the upper and lower bounds of wind or PV output at the same pseudo-measurement time serves as the pseudo-measurement information for DG buses.

For comparative experiments, we trained three models: the FTTA optimization proposed before, a greedy algorithm without local optimization from [30], and the PSO method without selection optimization.

Figure 9 illustrates the optimization process and iteration status. All three optimization models converge after iterations, significantly enhancing the effectiveness of state estimation compared to the unoptimized approach. The PSO and greedy algorithms, lacking constraint correction and local optima considerations, fail to converge to ideal states. Though requiring more iterations for convergence, the method proposed here reliably improves state estimation results.

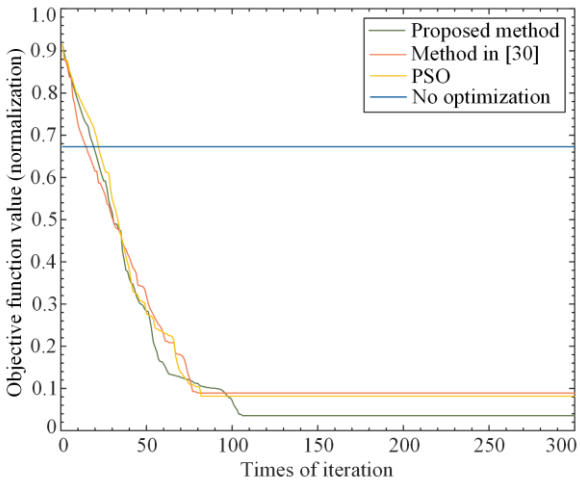


Fig. 9. Optimization process of pseudo-measurement selection for different models.

To comprehensively assess the effectiveness of the pseudo-measurement modeling strategy proposed in this article, four scenarios were designed for comparative experiments:

**Scenario 1:** EWM-SVM modeling with pseudo-measurement selection optimization.

**Scenario 2:** Basic SVM modeling with pseudo-measurement selection optimization.

**Scenario 3:** EWM-SVM modeling without pseudo-measurement selection optimization.

**Scenario 4:** Basic SVM modeling without pseudo-measurement selection optimization.

Figure 10 and Fig. 11 provide a comparison of estimation results, focusing on the RMSE of voltage amplitude and phase angle across 500 tests for four distinct scenarios. These figures illustrate that Scenario 1, utilizing the pseudo-measurement generation scheme and the selected optimization model proposed in this article, achieves significantly superior estimation performance compared to the other scenarios.

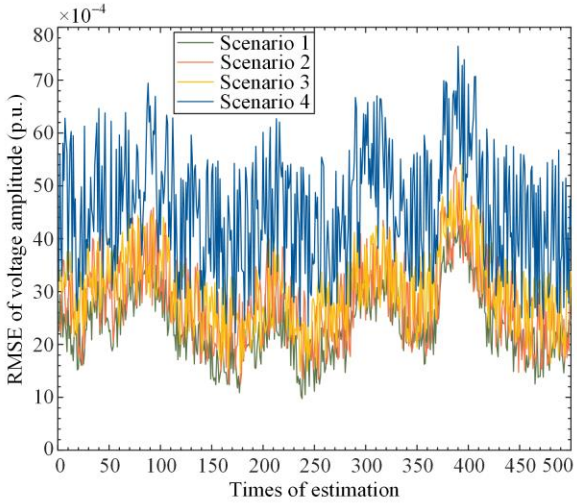


Fig. 10. Voltage amplitude RMSE comparison of four scenarios.

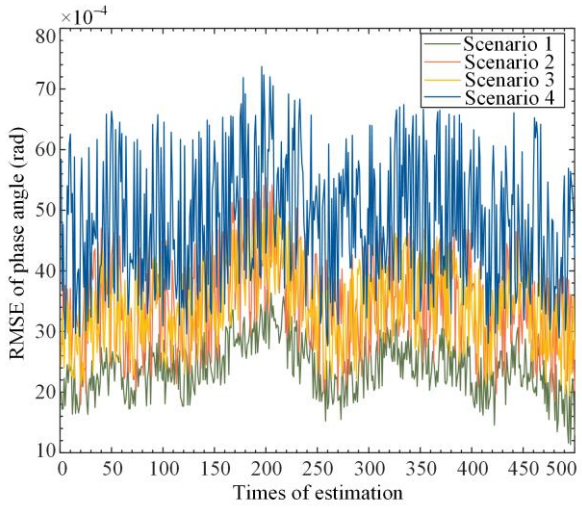


Fig. 11. Phase angle RMSE comparison of four scenarios.

Specifically, in Fig. 10, the RMSE of voltage amplitude across scenarios highlights that Scenario 1 exhibits markedly lower errors than Scenarios 2, 3, and 4. This superiority underscores the effectiveness of incorporating information entropy from historical operational data into the pseudo-measurement generation process. By aligning pseudo-measurement data more closely with actual operational states, Scenario 1 enhances state estimation accuracy for voltage amplitude.

Figure 11 compares the RMSE of phase angle across the same scenarios. Once again, Scenario 1 outperforms the others, demonstrating lower RMSE values. The comparison between Scenario 1 and Scenario 2 indicates that the consideration of information entropy contributes significantly to improving the alignment between pseudo-measurement outputs and actual system states, thereby reducing phase angle estimation errors. Furthermore, the comparison with Scenario 3 underscores the role of the selected optimization model in enhancing state estimation accuracy. This model effectively minimizes the objective function encompassing errors in active and reactive power while optimizing the selection of pseudo-measurement datasets.

Overall, Fig. 10 and Fig. 11 affirm that Scenario 1, combining the EWM-SVM modeling approach with optimized pseudo-measurement selection, offers the most effective strategy for improving voltage amplitude and phase angle estimation in distribution network state estimation tasks.

### *D. Scalability in Large Distribution Network*

In practical distribution systems, the number of buses typically ranges around 100 or more. To thoroughly assess the scalability of the proposed method on large-scale systems, we employed a modified IEEE-123 bus test system [32], depicted in Fig. 12, which includes detailed topology and distribution of DGs.

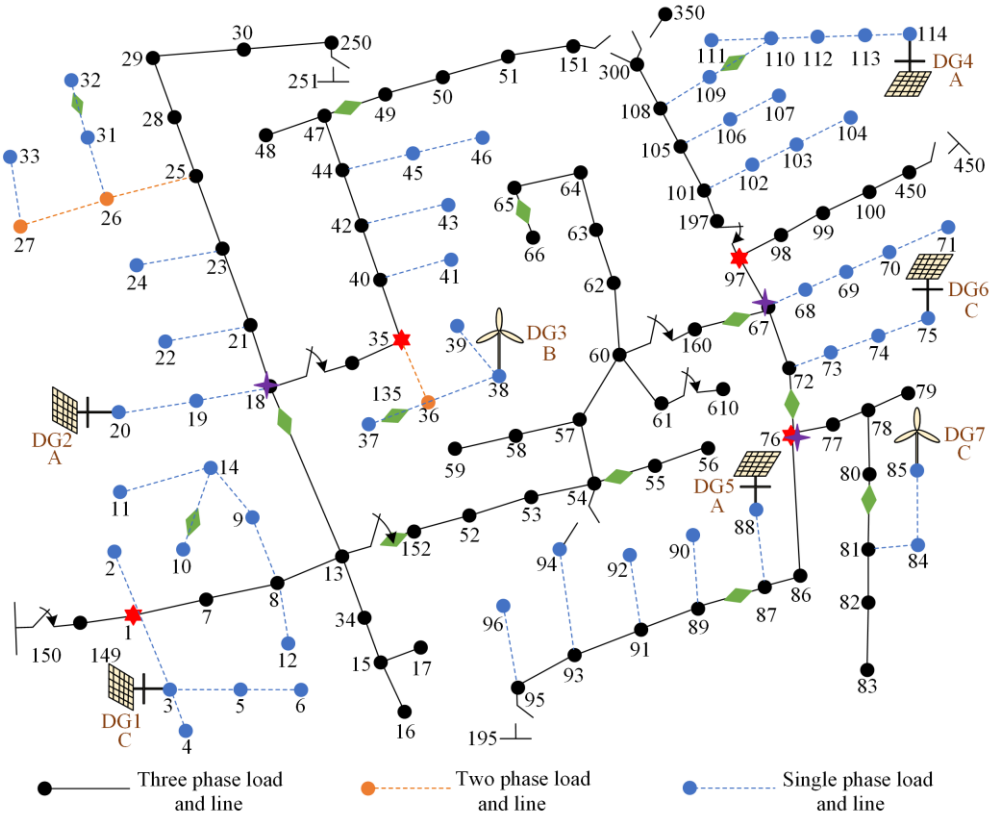


Fig. 12. The topology structure of the modified IEEE123-bus testing system.

For scalability testing, we designed four scenarios—labeled as a, b, c, and d—to evaluate the applicability of the pseudo-measurement modeling strategy proposed in this paper on larger systems.

**Scenario a:** Adopts the proposed pseudo-measurement modeling method and pseudo-measurement selection optimization.

**Scenario b:** Uses a pseudo-measurement generation scheme with conventional mean plus Gaussian noise (20% Gaussian noise) and the pseudo-measurement selection optimization proposed in this paper.

**Scenario c:** Adopts the proposed pseudo-measurement modeling method without pseudo-measurement selection optimization.

**Scenario d:** Utilizes a pseudo-measurement generation scheme with conventional mean and Gaussian noise (20% Gaussian noise) without pseudo-measurement selection optimization.

Table IV demonstrates that even in larger testing systems, the proposed pseudo-measurement modeling method and pseudo-measurement selection optimization can achieve accurate estimation results. Scenarios a and c notably exhibit higher accuracy compared to scenarios b and d. This finding suggests that the approach of separately modeling load buses and DG-connected buses with the proposed method enhances the effectiveness of state estimation, even without selection optimization.

TABLE IV
ACCURACY AND EFFICIENCY COMPARISON OF PSEUDO-MEASUREMENT FROM DIFFERENT MODELS

| Estimation object | Error of Scenario a ($\times10^{-4}$) | | Error of Scenario b ($\times10^{-4}$) | | Error of Scenario c ($\times10^{-4}$) | | Error of Scenario d ($\times10^{-4}$) | |
|---|---|---|---|---|---|---|---|---|
| Voltage amplitude error (p.u.) | Max | 62 | Max | 129 | Max | 103 | Max | 157 |
| | Min | 21 | Min | 84 | Min | 64 | Min | 101 |
| | Average | 32 | Average | 101 | Average | 86 | Average | 138 |
| Voltage phase angle error (rad) | Max | 59 | Max | 117 | Max | 99 | Max | 174 |
| | Min | 14 | Min | 78 | Min | 57 | Min | 98 |
| | Average | 21 | Average | 89 | Average | 71 | Average | 143 |

## VI. CONCLUSION

This study introduces an advanced pseudo-measurement modeling strategy for active distribution network state estimation, specifically targeting the uncertainty associated with wind power and PV systems. It employs double-layer KELM and DFNN models to predict and analyze the output uncertainty of wind turbines and PV buses, respectively. Additionally, the study integrates a historical data entropy-weighted kernel function to enhance pseudo-measurement generation for conventional load buses using SVM models. Moreover, optimizing the selection of pseudo-measurement datasets, guided by state estimation performance as the objective function, significantly enhances the effectiveness of the proposed strategy in providing pseudo-measurements for the estimator. The simulation example validates the efficacy of this strategy using modified IEEE-33 and IEEE-123 bus distribution network systems. The results demonstrate a notable improvement in the accuracy of active distribution network state estimation, offering innovative solutions to longstanding challenges in distribution network monitoring due to inaccurate pseudo-measurements.

Looking ahead, our objective is to expand the application of reliable pseudo-measurement modeling technology to forecasting-aided state estimation and multi-area distributed computing. This will further enhance the practical utility of our research in large-scale distribution network monitoring. Our forthcoming research will delve deeper into these areas, showcasing our ongoing exploration and contributions to the field.

## ACKNOWLEDGMENT

Not applicable.

## AUTHORS' CONTRIBUTIONS

Dongliang Xu: conceptualization, investigation, writing-original draft, methodology, software and formal analysis, writing-review and editing. Junjun Xu: writing-review & editing. Cheng Qian: methodology. Zaijun Wu: conceptualization, investigation, writing-original draft preparation, funding acquisition. Qinran Hu: validation.

## FUNDING

This work is supported by the National Natural Science Foundation of China (No. 52377086).

## AVAILABILITY OF DATA AND MATERIALS

Not applicable.

## DECLARATIONS

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.

## AUTHORS' INFORMATION

**Dongliang Xu** received the B.S degree in electrical engineering from Nanjing Normal University in 2019. He is currently working toward the Ph.D. degree in electrical engineering with the School of Electrical Engineering, Southeast University, Nanjing, China. His

research interests include state estimation and distribution network topology identification.

**Junjun Xu** received the B.S. degree in power system and its automation from the Nanjing Institute of Technology, Nanjing, China, in 2012, and the M.S. degree in agricultural electrification and automation from the School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China, in 2015, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2019. He is currently an associate professor of Nanjing University of Posts and Telecommunications. His research interests include distribution network state estimation, self-healing control, and uncertainty modeling approaches.

**Cheng Qian** received the B.S degree in electrical engineering from Zhengzhou University 2017. He is currently working toward the Ph.D. degree in electrical engineering with the School of Electrical Engineering, Southeast University, Nanjing, China. His research interests include load forecasting and non-invasive measurement.

**Zaijun Wu** received the B.S. degree in power system and its automation from the Hefei University of Technology, Hefei, China, in 1996, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2004. From 2012 to 2013, he was a visiting scholar with Ohio State University, Columbus, OH, USA. He is currently a professor of electrical engineering with the School of Electrical Engineering, Southeast University. He has authored or co-authored more than 130 referred journal papers, and is a reviewer of several journals. His research interests include microgrids, active distribution networks, and power quality.

**Qinran Hu** received the B.S. degree in electrical engineering from the Chien-Shiung Wu Honors College, Southeast University, Nanjing, China, in 2010, and the M.S. and Ph.D. degrees in electrical engineering from the University of Tennessee, Knoxville, TN, USA, in 2013 and 2015, respectively. He was a postdoctoral fellow with Harvard University, Cambridge, MA, USA, from 2015 to 2018. He joined the School of Electrical Engineering, Southeast University, in Oct. 2018. His research interests include distributed energy resources aggregation and power system operation optimization.

## REFERENCES

[1] Y. Ju and Y. Huang. "State estimation for an AC/DC hybrid power system adapted to non-smooth characteristics," *Power System Protection and Control*, vol. 51, no. 2, pp. 141-150, Jan. 2023. (in Chinese)

[2] J. Xu, Z. Wu, and X. Yu *et al*., "A dynamic robust restoration framework for unbalanced power distribution networks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6301-6312, Jan. 2020.

[3] J. Xu, J. Liu, and Z. Wu *et al*., "Multi-area state estimation for active distribution networks under multiple uncertainties: an affine approach," *International Journal of Electrical Power & Energy Systems*, vol. 155, no. 109632, pp. 1-11, Jan. 2024.

[4] C. Li, Y. Xi, and Y. Lu *et al*., "Resilient outage recovery of a distribution system: co-optimizing mobile power sources with network structure," *Protection and Control of Modern Power Systems*, vol. 7, no. 3, pp. 1-13, Jul. 2022.

[5] Y. Wang, J. Gu, and L. Yuan, "Distribution network state estimation based on attention-enhanced recurrent neural network pseudo-measurement modeling," *Protection and Control of Modern Power Systems*, vol. 8, no. 2, pp. 1-16, Apr. 2023.

[6] K. Dehghanpour, Y. Yuan, and Z. Wang *et al*., "A game-theoretic data-driven approach for pseudo-measurement generation in distribution system state estimation," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 5942-5951, Jan. 2019.

[7] A. Grandjean, J. A. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 9, pp. 6539-6565, Dec. 2012.

[8] A. Bernieri, G. F. Betta, and C. Liguori *et al*., "Neural networks and pseudo-measurements for real-time monitoring of distribution systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 645-650, Apr. 1996.

[9] L. Wang, H. Gao, and G. Zou, "Modeling methodology and fault simulation of distribution networks integrated with inverter-based DG," *Protection and Control of Modern Power Systems*, vol. 2, no. 4, pp. 1-9, Oct. 2017.

[10] R. Yan, Y. Yuan, and Z. Wang *et al*., "Active distribution system synthesis via unbalanced graph generative adversarial network," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4293-4307, Oct. 2022.

[11] X. Zhang, W. Yan, and M. Huo *et al*., "Robust interval state estimation for distribution systems considering pseudo-measurement interval prediction," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 1, pp. 179-188, Jun. 2023.

[12] R. Singh, B. C. Pal, and R. A. Jabr, "Distribution system state estimation through Gaussian mixture model of the load as pseudo-measurement," *IET Generation, Transmission & Distribution*, vol. 4, no. 1, pp. 50-59, Jan. 2010.

[13] B. Özsoy and M. Göl, "A hybrid state estimation strategy with optimal use of pseudo-measurements," in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe*, Sarajevo, Bosnia and Herzegovina, Oct. 2018, pp. 1-6.

[14] E. Manitsas, R. Singh avindra, and B. C. Pal *et al*., "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1888-1896, Apr. 2012.

[15] Z. Cao, Y. Wang, and C. Chu *et al.*, "Robust pseudo-measurement modeling for three-phase distribution systems state estimation," *Electric Power Systems Research*, vol. 180, no. 106138, pp. 1-10, Mar. 2020.

[16] H. Z. Margossian, "Iterative state estimation with weight tuning and pseudo-measurement generation," *IEEE Systems Journal*, vol. 15, no. 3, pp. 3165-3172, Mar. 2021.

[17] S. K. Injeti and V. K. Thunuguntla, "Optimal integration of DGs into radial distribution network in the presence of plug-in electric vehicles to minimize daily active power losses and to improve the voltage profile of the system using bio-inspired optimization algorithms," *Protection and Control of Modern Power Systems*, vol. 5, no. 1, pp. 1-15, Jan. 2020.

[18] S. K. Bilgundi, R. Sachin, and H. Pradeepa *et al.*, "Grid power quality enhancement using an ANFIS optimized PI controller for DG," *Protection and Control of Modern Power Systems*, vol. 7, no. 1, pp. 1-14, Jan. 2022.

[19] L. Mehigan, J. P. Deane, and B. P. Gallachóir *et al.*, "A review of the role of distributed generation (DG) in future electricity systems," *Energy*, vol. 163, no. 1, pp. 822-836, Nov. 2018.

[20] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote-sensing image classification," *Remote Sensing Letters*, vol. 4, no. 9, pp. 853-862, Jun. 2013.

[21] T. T. Truong, D. Dinh-Cong, and J. Lee *et al.*, "An effective deep feedforward neural networks (DFNN) method for damage identification of truss structures using noisy incomplete modal data," *Journal of Building Engineering*, vol. 30, no. 101244, pp. 1-20, Jul. 2020.

[22] F. Marini and B. Walczak, "Particle swarm optimization (PSO). A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, no. 1, pp. 153-165, Dec. 2015.

[23] A. Khosravi, S. Nahavandi, and D. Creighton *et al.*, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp.337-346, Mar. 2011.

[24] M. E. Mavroforakis and S. Theodoridis, "A geometric approach to support vector machine (SVM) classification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 671-682, May 2006.

[25] M. A. Zulfiqar, M. Kamran, and M. B. Rasheed *et al.*, "A hybrid framework for short term load forecasting with a navel feature engineering and adaptive grasshopper optimization in smart grid," *Applied Energy*, vol. 338, no. 120829, pp. 1-20, May 2023.

[26] R. M. Gra, "Distortion and entropy," in *Entropy and information theory*, 1st ed., California, USA: Springer Science & Business Media, 2011, pp. 147-169.

[27] Z. Tian and M. Gai, "Football team training algorithm: a novel sport-inspired meta-heuristic optimization algorithm for global optimization," *Expert Systems with Applications*, vol. 245, no. 123088, pp. 1-18, Jul. 2024.

[28] Commission for Energy Regulation (CER), "CER smartmetering project electricity customer behaviour trial, 2009-2010," *Irish Social Sci. Data Arch.*, vol. 1, no. 1, 2012.

[29] J. Zhang, Y. Wang, and Y. Weng *et al.*, "Topology identification and line parameter estimation for non-PMU distribution network: a numerical method," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4440-4453, Sept. 2020.

[30] M. Mao, Z. Wu, and D. Xu *et al.*, "Community-detection-based approach to distribution network partition," *CSEE Journal of Power and Energy Systems*, [Online], pp. 1-11. Available: https://ieeexplore. ieee. org/abstract/document/9666824

[31] J. Xu, Z. Wu, and T. Zhang *et al.*, "A distributed interval state estimation framework of distribution networks based on multi-source measurements," *Proceedings of the CSEE*, vol. 42, no. 24, pp. 8888-8900, Sept. 2022.

[32] D. Xu, Z. Wu, and J. Xu *et al.*, "A multi-area forecasting aided state estimation strategy for unbalance distribution networks," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 806-814, Jan. 2024.