

ORIGINAL RESEARCH

Open Access



Distribution network state estimation based on attention-enhanced recurrent neural network pseudo-measurement modeling

Yaojian Wang¹ , Jie Gu^{1*} and Lyuzerui Yuan¹

Abstract

Because there is insufficient measurement data when implementing state estimation in distribution networks, this paper proposes an attention-enhanced recurrent neural network (A-RNN)-based pseudo-measurement modeling method. First, based on analyzing the power series at the source and load end in the time and frequency domains, a period-dependent extrapolation model is established to characterize the power series in those domains. The complex mapping functions in the model are automatically represented by A-RNNs to obtain an A-RNNs-based period-dependent pseudo-measurement generation model. The distributed dynamic state estimation model of the distribution network is established, and the pseudo-measurement data generated by the model in real time is used as the input of the state estimation model together with the measurement data. The experimental results show that the method proposed can explore in depth the complex sequence characteristics of the measurement data such that the accuracy of the pseudo-measurement data is further improved. The results also show that the state estimation accuracy of a distribution network is very poor when there is a lack of measurement data, but is greatly improved by adding the pseudo-measurement data generated by the model proposed.

Keywords State estimation, Pseudo measurement, Recurrent neural network, Attention mechanism, Time-frequency domain analysis, Distribution network

1 Introduction

Power system state estimation is the process of determining the internal state of an energy system (e.g., node voltage vectors), by “fusing” a mathematical model and input/output data measurements. State estimation is fundamental to many analysis, monitoring, and energy management tasks, and its accuracy has an important impact on power system security and stability [1]. The amount of measurement resources is an important factor affecting the estimation accuracy. Adequate measurement data

can help to obtain high accuracy estimation results, while on the contrary, the estimation results may not reliably reflect the real state of the system or even make the state estimation problem insoluble if measurement data are inadequate [2].

In practice, the measurement system of a distribution network (DN) cannot provide sufficient resources because of the constraints of cost, communication delay, and uneven distribution of measurement points [3]. To solve this problem, researchers have proposed the solution of pseudo-measurement generation, i.e., artificially generating currently available “measurement data” using existing data (e.g., load power, distributed power output, etc.) without adding extra measurement equipment [3, 4]. Pseudo-measurement generation for DN state estimation has been widely studied. The literature in this field can be roughly categorized into two groups: statistical

*Correspondence:

Jie Gu

gujie@sjtu.edu.cn

¹ Department of Electrical Engineering, School of Electronics, Information and Electrical Engineering, Shanghai Jiao Tong University, 209 of Smart Grid Building, 800 Dong Chuan Rd, Shanghai 200240, China

and probabilistic methods, and machine-learning methods. Statistical and probabilistic methods can explicitly characterize the distribution of the data. Reference [5] considers the non-normal distribution properties of loads and the correlation of loads at different nodes, and fits a load density function satisfying log-normal and beta distributions based on hourly historical load data. However, the static characteristics of the load cannot be fully described by the mean and variance alone, and the generated load pseudo-measurement types are limited.

Compared with statistical and probabilistic methods, the machine-learning-based pseudo-measure generation method does not need to build an explicit mathematical model, and can directly mine and learn hidden characteristics from the data. This offers a stronger feature characterization capability. In [7], historical load data, adjacent time interval load difference, real-time temperature, humidity, and date are jointly used as inputs to train a Gaussian process regression model to obtain load pseudo-measurements. Reference [8] proposes a distributed game-theoretic-based correlation vector machine model to capture the uncertainty of node injection power and generate pseudo-measurements. However, Gaussian process regression models and correlated vector machines have very limited nonlinear fitting capability and cannot dig deeper into sequence features, and thus it is difficult for ordinary neural networks to mine long-term correlations in time series data (such as load and DG power). The proposed recurrent neural networks (RNNs) can solve this problem, because RNNs retain a memory of what had already been processed and thus can learn from previous iterations during the training. For example, reference [9] takes advantage of this by training recurrent neural network models through historical measurement data to generate more accurate pseudo-measurement data.

Although RNNs have a superior temporal correlation feature learning capability, their initial attention to each element in the input dataset is same, which is not conducive to fast screening of high-value features. The attention mechanism can help to capture the model hidden layer dimensional relationships, and was initially used mainly for tasks such as natural language processing, where its focus was on the features needed for the target scenario. Reference [10] proposes a short-term load forecasting method using a dual attention mechanism to improve the traditional gated recurrent unit (GRU). This weakens the influence of each input feature on the grid load situation and enhances the RNNs to capture the long-time dependence of the load data. The results from application of the algorithm show that the prediction accuracy is improved, to different degrees, compared with previous models.

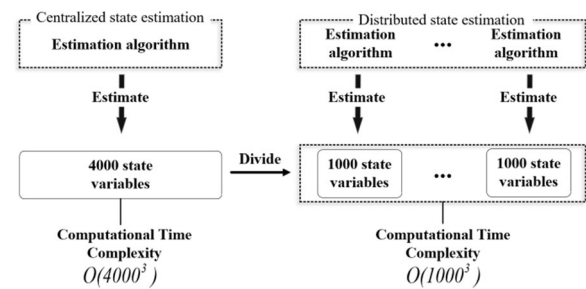


Fig. 1 Schematic diagrams of centralized state estimation and distributed state estimation

Therefore, this paper considers combining RNNs with an attention mechanism to generate pseudo-measurement data with higher accuracy.

For DN state estimation, a well-developed and widely used estimation algorithm is Weighted Least Squares (WLS). However, when WLS is directly applied to the state estimation problem, there are significant drawbacks in terms of computational speed and estimation accuracy. Specifically, in terms of computational speed, the computational time complexity of WLS grows in a power series with the number of state variables, and thus the centralized WLS state estimation of all state variables in the DN takes a lot of time, which makes it difficult to provide timely data support for other real-time tasks in dispatch management. In terms of estimation accuracy, the dynamic uncertainty of the state of the DN is greatly enhanced because of the randomness and fluctuation of the power of active components such as renewable energy, while as WLS only uses the current measurement information to obtain the optimal estimation, it is difficult to capture the dynamic change characteristics of the system. This leads to the final estimation results deviating from the real situation and not being able to provide reliable state data for dispatch management. To address the above problems, this paper adopts the technical idea of combining distributed estimation with the dynamic state estimation algorithm, also known as distributed dynamic state estimation. As shown in Fig. 1, the distributed estimation approach can divide a high-dimensional state variable into multiple low-dimensional state variables for parallel estimation, which significantly reduces the computational complexity of the estimation. As shown in Fig. 2, the dynamic state estimation algorithm is different from the static estimation. It uses not only the current measurement information but also the historical state information to portray the trend of system state changes, to improve the estimation accuracy.

The main contributions of this paper are as follows:

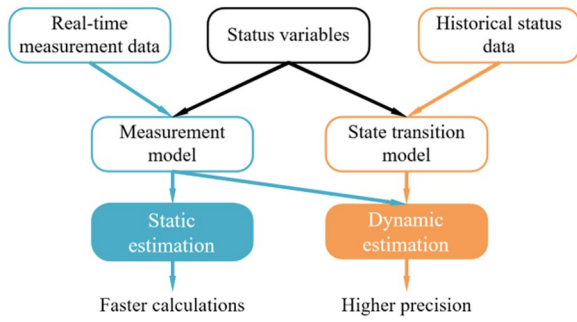


Fig. 2 Static state estimation and dynamic state estimation

- (i) Introducing a distributed dynamic state estimation model to achieve fast state estimation of a large-scale distribution network.
- (ii) Establishing a period-correlation extrapolation model to characterize the time–frequency domain of power series.

It introduces a multi-headed attention mechanism to focus on data with high relevance to the target and combines it with RNNs to construct attention-enhanced RNNs (A-RNNs). This can generate more accurate pseudo-measurement data to improve the accuracy of state estimation.

The remainder of this paper is organized as follows. In Sect. 2, the basic principles and algorithms of distributed dynamic state estimation are introduced, while in Sect. 3 the proposed periodic-correlation pseudo-measurement generation model based on A-RNNs is described. Section 4 summarizes data sources and experiment implementation, while results of the proposed model are presented in Sect. 5, where comparisons with existing models are discussed. Section 6 concludes the paper.

2 Distributed dynamic state estimation model

To address the problem of long estimation time caused by DN high-dimensional state variables, and to further improve the accuracy of distribution network state estimation, distributed dynamic state estimation is adopted. The distributed estimation can greatly reduce the dimension of state variables and shorten the time of state estimation, and the use of dynamic estimation increases the effective use of state transition information and improves the accuracy of state estimation.

2.1 Distributed dynamic state estimation model and problem description

From a limited amount of measurement data which contain noise, it is difficult to obtain the true system state. However, multiple available valid information (e.g., measurement

data and historical state estimates) can create conditions for computing optimal or suboptimal estimates of the true state. Specifically, the state transfer characteristics and their noise distribution obtained from the historical state estimates provide a priori information that reflects the trend of the system, while the measurement data and their error statistics can help measure the likelihood that the estimated state is consistent with the true state. Based on state transfer properties, measurement data and noise statistics, the dynamic estimation algorithm can find the optimal or sub-optimal estimation of the system state [16].

2.1.1 State transition equation for distributed dynamic state estimation

Let a DN have s sub-areas, where the k^{th} area has n_k nodes. Denote the set of nodes constituting the k^{th} area as A_k and its state variable at moment t is $x_t^{(k)} = [\theta_{t,1}^{(k)}, \dots, \theta_{t,n_k}^{(k)}, V_{t,1}^{(k)}, \dots, V_{t,n_k}^{(k)}]^T$, where $\theta_{t,i}^{(k)}$ and $V_{t,i}^{(k)}$ represent the phase angle and amplitude of the voltage of the i^{th} node at moment t , respectively. The state transition trend of $x_t^{(k)}$ can be described with a discrete first-order Markov model, as:

$$x_{t+1}^{(k)} = f^{(k)}(x_t) + w_t^{(k)} \quad (1)$$

where $w_t^{(k)}$ is the state transition noise and its distribution properties need to be described by a known probability distribution function (pdf), which is called prior pdf, i.e., $w_t^{(k)} \sim p(x_{t+1}|x_t)$. The most commonly used Gaussian pdf is chosen to characterize $w_t^{(k)}$. Let $w_t^{(k)}$ obey the mean of 0 and the covariance matrix of W_t with a Gaussian distribution, denoted as $w_t^{(k)} \sim N(0, W_t)$. $f^{(k)}(\cdot)$ is the state transfer function of $x_t^{(k)}$, and in general it is nonlinear. In order to reduce the computational complexity, $f^{(k)}(\cdot)$ is usually linearized to be explicitly expressed [20], i.e.:

$$x_{t+1}^{(k)} = F_t^{(k)} x_t^{(k)} + G_t^{(k)} + w_t^{(k)} \quad (2)$$

Set $\tilde{x}_{t-1}^{(k)}$ as the status value calculated by the state transition equation, $\hat{x}_{t-1}^{(k)}$ as the final state estimation results at time $t-1$, $F_t^{(k)}$ and $G_t^{(k)}$ in (2) can then be obtained by using Holt-Winters exponential smoothing method, as:

$$F_t^{(k)} = \varepsilon(1 + \gamma) \quad (3)$$

$$G_t^{(k)} = (1 + \gamma)(1 - \varepsilon)\tilde{x}_{t-1}^{(k)} - \gamma a_{t-1}^{(k)} + (1 - \gamma)b_{t-1}^{(k)} \quad (4)$$

where γ and ε are artificial smoothing parameters, $a_{t-1}^{(k)}$ is the state change of horizontal component, and $b_{t-1}^{(k)}$ is the state change of trend component at time $t-1$. The value of both at time t can be updated to:

$$a_t^{(k)} = \varepsilon \hat{x}_{t-1}^{(k)} + (1 - \varepsilon) \tilde{x}_{t-1}^{(k)} \quad (5)$$

$$b_t^{(k)} = \gamma(a_t^{(k)} - a_{t-1}^{(k)}) + (1 - \gamma)b_{t-1}^{(k)} \quad (6)$$

2.1.2 Measurement equations for distributed dynamic state estimation

The mathematical relationship between the measurement data and the system state $x_t^{(k)}$ can be expressed using the measurement equations. Let all the measurement data that can be received at time t be the measurement vector $z_t^{(k)}$, and assume that all the measurement noise $r_t^{(k)}$ are independent of each other. Then the mathematical relationship between $x_t^{(k)}$ and $z_t^{(k)}$ can be described as:

$$z_t^{(k)} = h^{(k)}(x_t) + r_t^{(k)} \quad (7)$$

Similar to w_i in (1), $r_t^{(k)}$ is also described by a given pdf, called the likelihood pdf $p(z_t|x_t)$, i.e. $r_i^{(k)} \sim p(z_t|x_t)$. Similarly, the most commonly used Gaussian pdf is chosen, i.e., $r_t^{(k)}$ obeys a Gaussian distribution with mean 0 and covariance matrix R_t of the Gaussian distribution, denoted as $r_i^{(k)} \sim N(0, R_t)$. $h^{(k)}(\cdot)$ represents the nonlinear measurement function, which is determined by the type of measurement data and the state variable.

2.1.3 Distributed dynamic state estimation problem and solution

The objective of DN state estimation is to minimize the estimation errors of all nodes in the DN, i.e., the difference between the estimated state and the true state. According to [11], the optimization objective of the distributed dynamic estimation problem at time t can be expressed as:

$$\min \sum_{k=1}^x E[(x_t^{(k)} - \hat{x}_t^{(k)})^T S_t^{(k)} (x_t^{(k)} - \hat{x}_t^{(k)})] \quad (8)$$

where $E(\cdot)$ is the expectation function, while $x_t^{(k)}$ and $\hat{x}_t^{(k)}$ represent the true and estimated states, respectively. $S_t^{(k)}$ is a custom weight function, which is usually set based on the measurement error. The constraints of the DN state estimation problem are the state transition equation of (1) and the measurement equation of (7).

Distributed dynamic state estimation for DN is accomplished quickly through three steps: local state estimation, boundary state consistency and local state update for global estimation. Among the three steps, both the local state estimation step and the local state update step can be done based on existing dynamic estimation algorithms, while the boundary state consistency step relies on boundary consistency estimation methods that can compute global estimates of boundary nodes. In the following paragraphs, the three steps of the distributed

dynamic estimation method will be introduced as an example of solving the global estimate of the sub-area k at time $t + 1$.

(1) Local state estimation: the computational process of local state estimation can be expressed as a whole as:

$$\hat{x}_{t+1}^{l,(k)} = b_{t+1}^{(k)}(\hat{x}_t^{(k)} + \hat{\Lambda}_{t+1}^{(k)}) \quad (9)$$

where $b_{t+1}^{(k)}(\cdot)$ is the mapping function between the global state estimate $\hat{x}_t^{(k)}$ at time t of sub-area k and $\hat{x}_{t+1}^{l,(k)}$ at time $t + 1$. $\hat{\Lambda}_{t+1}^{(k)}$ is the set of parameters of function $b_{t+1}^{(k)}(\cdot)$, which includes process noise, measurement noise, local measurement data, etc.

(2) Boundary state consistency: $\hat{x}_{t+1}^{l,(k)}$ is the estimation result obtained for the sub-area k based on locally valid information, without taking into account the external valid information. This makes it different from the corresponding global estimation results. Thus, the sub-area k needs to perform consistent estimation of the boundary state, i.e., exchanging information with all neighboring sub-areas and obtaining all external information related to its boundary nodes, so that the local estimate of the boundary nodes in $\hat{x}_{t+1}^{l,(k)}$ of the local estimate is corrected to the global estimate.

(3) Local state update: after obtaining the global estimation result $\hat{x}_{t+1}^{l,(k,ib)}$ of the boundary nodes in sub-area k , the local estimation of the internal nodes (all nodes except the boundary nodes) in sub-area k result $\hat{x}_{t+1}^{l,(k)}$ is corrected by using it as the base, and finally the global state estimation result of sub-area k is obtained as:

$$\hat{x}_{t+1}^{(k)} = e_{t+1}^{(k)}(\hat{x}_{t+1}^{l,(k)}, \hat{x}_{t+1}^{l,(k,ib)}; \Upsilon_{t+1}^{(k)}) \quad (10)$$

where $e_{t+1}^{(k)}(\cdot)$ is the mapping function of sub-area k with its local state estimate $\hat{x}_{t+1}^{l,(k)}$, boundary global estimate $\hat{x}_{t+1}^{l,(k,ib)}$ and sub-area k global state estimate $\hat{x}_{t+1}^{(k)}$, after achieving the boundary state consistency. $\Upsilon_{t+1}^{(k)}$ is the set of parameters of function $e_{t+1}^{(k)}(\cdot)$, which includes locally valid information for the sub-area as well as externally valid information obtained from other sub-areas.

2.2 Distributed state estimation algorithm

In existing studies, the theoretical basis for solving dynamic state estimation problems is Bayesian filtering [12]. According to (1) and (7), the power system is a dynamic system with first-order Markovianity, and the measurement at any moment depends only on the state at that moment and is independent of the measurement and state at other moments. Let the initial state of the power system be x_0 , the state at time t is x_t , the measurement data is z_t . Its priori pdf is denoted as $p(x_t|x_{t-1})$, and the likelihood pdf is $p(z_t|x_t)$. According to Bayesian theory, the a posterior pdf that contains the complete

statistical information of x_t can be obtained by calculating the a priori pdf and the likelihood pdf. Therefore, the statistical properties of the state variables characterized by the posterior pdf are considered to be approximately equal to the statistical properties of the real state, i.e., the optimal estimate of the state variables can be obtained from the posterior pdf. According to the state transition equation and the measurement equation, Bayesian filtering can obtain the posterior pdf of x_t through the prediction step and the update step.

(1) Prediction step: using the previous $t - 1$ moments of the measure $z_{1:t-1} = [z_1, z_2, \dots, z_{t-1}]$, and the posterior pdf $p(x_{t-1}|z_{1:t-1})$ at the time $t - 1$ and the prior pdf $p(x_t|x_{t-1})$ at the time t , the predicted probability function at time t can be calculated as:

$$p(x_t|z_{1:t-1}) = \int_{-\infty}^{+\infty} p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1})dx_{t-1} \quad (11)$$

It can be seen that the prediction step of Bayesian filtering is actually a fusion of statistical properties of the state variables of posterior information at $t - 1$ and prior information at t .

(2) Update step: On the basis of the prediction step, the measurement z_t at time t and the corresponding likelihood pdf $p(z_t|x_t)$ are added to obtain the posterior pdf of the state variables at time t :

$$\begin{aligned} p(x_t|z_{1:t}) &= \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})} \\ &= \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{\int_{-\infty}^{+\infty} p(z_t|x_t)p(x_t|z_{1:t-1})dx_t} \\ &= \eta p(z_t|x_t)p(x_t|z_{1:t-1}) \end{aligned} \quad (12)$$

where $\eta = \int_{-\infty}^{+\infty} p(z_t|x_t)p(x_t|z_{1:t-1})dx_t$ is a constant unrelated to x_t . Usually, the expectation of the posterior pdf is the result of the state estimation at time t , as:

$$\hat{x}_t = \int x_t p(x_t|z_{1:t})dx_t \quad (13)$$

The flow chart and the calculation procedure of the distributed dynamic state estimation of the DN are shown in Figs. 3 and 4, respectively. It can be seen that the solution of the state estimation problem is influenced by the quality of the measurement data in each sub-region of the DN. However, because of economic constraints, only a small number of PMUs are installed in the DN sub-region, and most of the available measurement data are still provided by traditional measurement systems, including SCADA systems and Advanced Metering Infrastructure (AMI). Although SCADA systems can provide voltage amplitude and node injection power measurements every second, the

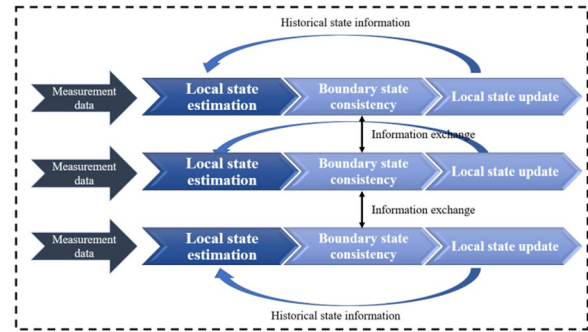


Fig. 3 Schematic diagrams of distributed dynamic state estimation

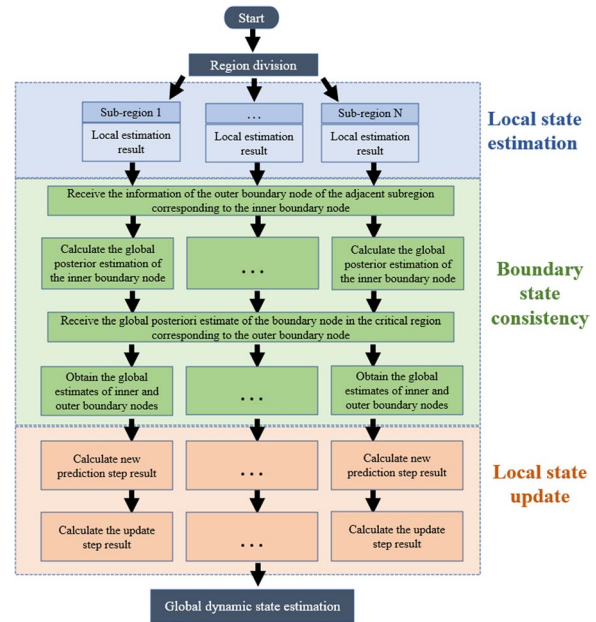


Fig. 4 Calculation procedure of the distributed dynamic state estimation

number of measurements obtained may be fewer than the number of state variables because of communication delays and limited measurement points, while the data upload period of AMI is 5 or 15 min, which cannot meet the real-time requirements for state estimation [17]. Therefore, the DN sub-region suffers from a scarcity of measurements in the second-level. Therefore, in this paper, a periodic-correlation pseudo-measurement generation model based on A-RNNs is considered to generate pseudo-measurement data in real time to alleviate the measurement scarcity problem in the DN sub-region.

3 A periodic-correlation pseudo-measures generation model based on attention-enhanced recurrent neural networks

Analysis and modeling of time–frequency domain characteristics of load and distributed power output.

3.1 Periodic-correlation extrapolation model

Generally, the source-load power series has a strong correlation with the data that lags itself by several time steps. The autocorrelation of the different time series can be described by the maximum lag time step and the minimum autocorrelation function (ACF) value, where the maximum lag time step is selected by humans and the minimum ACF value is the ACF value corresponding to the maximum lag time step. When the minimum ACF value is larger, the historical power data within the maximum lag time step has a stronger correlation with the power data at the current moment.

In addition to the autocorrelation of power series demonstrated by ACF, power data are also related to external factors such as weather and date. In this subsection, the Jensen-Shannon divergence (JS) is used to quantify the similarity between load, PV and wind power data and their corresponding external factor series. The JS value is 0 when the two time series data are distributed independently and is 1 when the opposite is true.

The power series not only has autocorrelation and external similarity in the time domain, but also exhibits periodic fluctuations in the frequency domain. These can be derived from Fourier spectrum analysis.

Combining the above analysis, a periodic-correlation extrapolation model is constructed to map the time-frequency domain characteristics of the power series. The independent variables of this model consist of two parts. One is the autocorrelated periodic series $X_{t,T}$, which reflects the autocorrelation and periodic volatility of the power series, and the other is the externally correlated periodic series $E_{t,T}$, which reflects the similarity and periodic volatility of the power series and the external factor series, and they are collectively called the periodic-correlation input. The periodic-correlation extrapolation model characterizing the target power x_t at the time of t with its periodic-correlation input is given as:

$$x_t = W_{S2}\phi_S(W_{S1}^T X_{t,T}) + W_{E2}\phi_E(W_{E1}^T E_{t,T}) + b_r \quad (14)$$

where b_r is the residual coefficient, W_{S1} and W_{E1} are the weight matrices of the mapping functions $\phi_S(\cdot)$ and $\phi_E(\cdot)$, respectively. W_{S2} and W_{E2} represent the weight matrices of $X_{t,T}$ and $E_{t,T}$, respectively. The autocorrelated periodic series $X_{t,T}$ is based on the power series data $\{x\}$ and can be expressed as:

$$X_{t,T} = [X_{t,T,1}, X_{t,T,1}, \dots, X_{t,T,k}] \quad (15)$$

where $X_{t,S} = [x_{t-1}, \dots, x_{t-m}]$ is an autocorrelated sequence of m lagged time steps of x_t . X_{t,T_i} ($i = 1, \dots, k$) is a periodic cycle sequence with x_t and $X_{t,T}$ in a strong cycle T_i , i.e.:

$$X_{t,T_i} = [X_{t,T_i,1}, X_{t,T_i,2}, \dots, X_{t,T_i,c}], \quad i = 1, \dots, k \quad (16)$$

where the j th ($j = 1, \dots, c$) element of X_{t,T_i} is $X_{t,T_i,j} = [x_{t-jT_i}, x_{t-1-jT_i}, \dots, x_{t-m-jT_i}]$, and c is the number of iterations of X_{t,T_i} . The external correlation periodic sequence $E_{t,T}$ is based on the external factor sequence $\{\varepsilon\}$, and its construction is similar to the above process, which can be expressed as:

$$E_{t,T} = [E_{t,S}, E_{t,T_1}, \dots, E_{t,T_k}] \quad (17)$$

$$E_{t,S} = [\varepsilon_{t-1}, \dots, \varepsilon_{t-m}] \quad (18)$$

$$E_{t,T_i} = [E_{t,T_i,1}, E_{t,T_i,2}, \dots, E_{t,T_i,c}], \quad i = 1, \dots, k \quad (19)$$

$$E_{t,T_i,j} = [\varepsilon_{t-jT_i}, \varepsilon_{t-1-jT_i}, \dots, \varepsilon_{t-m-jT_i}], \quad j = 1, \dots, c \quad (20)$$

Thus, after determining the autocorrelation maximum lag term m , k strong period $T_1, T_2, \dots, T_k$, and the number of cycles c , the cycle correlation input of the power series can be constructed based on the historical power data and external factor data. If the optimal weight matrix $\{\hat{W}_{S1}, \hat{W}_{S2}, \hat{W}_{E1}, \hat{W}_{E2}\}$, residuals $\hat{b}_r$ and mapping functions $\hat{\phi}_S(\cdot)$ and $\hat{\phi}_E(\cdot)$ exist, let

$$\min \Delta x_t = |x_t - \hat{W}_{S2}\hat{\phi}_S(\hat{W}_{S1}^T X_{t,T}) + \hat{W}_{E2}\hat{\phi}_E(\hat{W}_{E1}^T E_{t,T}) + \hat{b}_r| \quad (21)$$

then the power data can be predicted using periodic-correlation input and periodic-correlation extrapolation models.

However, the mapping functions $\phi_S(\cdot)$ and $\phi_E(\cdot)$ are usually nonlinear and may be either explicit or implicit functions, making it difficult for traditional algorithms to find their optimal representations. Deep learning, as an emerging type of data-driven optimization algorithm, can mine the hidden shallow and deep features from a large amount of data through multiple neuron computation nodes and multiple layers of network operations, so as to approximate the correlation between input and output as much as possible and complete the automated representation of complex mappings. Therefore, a suitable deep learning framework is considered to achieve the optimization objective described in (21).

3.2 Pseudo-measurement data generation model

The powerful temporal correlation learning capability and nonlinear fitting ability of RNNs can approximate the periodic-correlation extrapolation models proposed in (21). However, when these RNN models take the autocorrelated periodic sequence $X_{t,T}$ as input, they will

focus equally on historical data with different lag steps, and incorporate invalid information from some data with large lag time steps into the feature representation, thus wasting computational resources or even affecting the accuracy of the fit. Therefore, RNNs should focus more on data with high relevance to the target in order to obtain more critical detailed information and suppress other useless information.

For this reason, this paper introduces a multi-headed attention mechanism to focus on data with high relevance to the target and combines it with RNNs to construct A-RNNs to better fit (21).

3.2.1 Multi-headed attention mechanism

The attention mechanism is an information focusing technique that imitates a human cognitive attention mechanism, which can help RNNs focus more on the less but important information in the input. The core idea is to train the generation of weight coefficients corresponding to the input under a given target, so as to determine the importance of information. As shown in the left side of Fig. 4, the attention mechanism represents the input as a "key-value pair", with "key" K used to calculate the attention distribution and "value" V used to calculate the aggregated attention value. Then the attention value of V is obtained by using the scaled dot product scoring function and the query vector Q .

The multi-head attention mechanism (MAM) uses multiple attention mechanisms to obtain multiple sub-feature spaces, thereby focusing on important features of the original information from different aspects. As shown on the right side of Fig. 5, it decomposes the input into multiple uncorrelated sub-feature spaces, each with its own "key" K_i , "value" V_i , and query vector Q_i . Then the attention results of each subspace are calculated, and all the results are spliced to get the final multi-head attention results. MAM is consistent with the proposed idea of mining power data features from multiple perspectives in the time–frequency domain. Therefore, this paper considers using a neural network combined with MAM and RNNs to approximate the periodic-correlation extrapolation model.

3.2.2 Attention-enhanced recurrent neural networks and pseudo-measurement generation model

A-RNNs, combining MAM with RNNs framework, can obtain more critical detail information from the input and suppress other, useless, information. As shown in Fig. 6, A-RNNs containing one layer of RNNs and one layer of MAMs are used as an example to introduce the model.

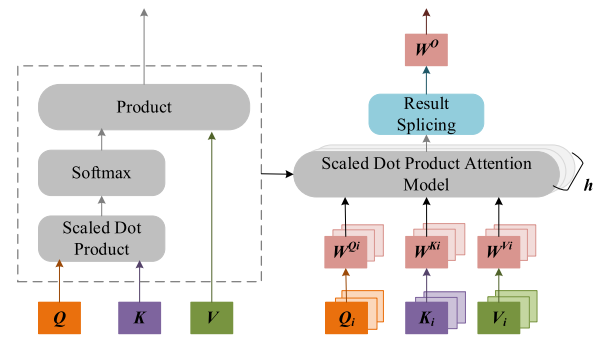


Fig. 5 Schematic diagram of a multi-head attention mechanism

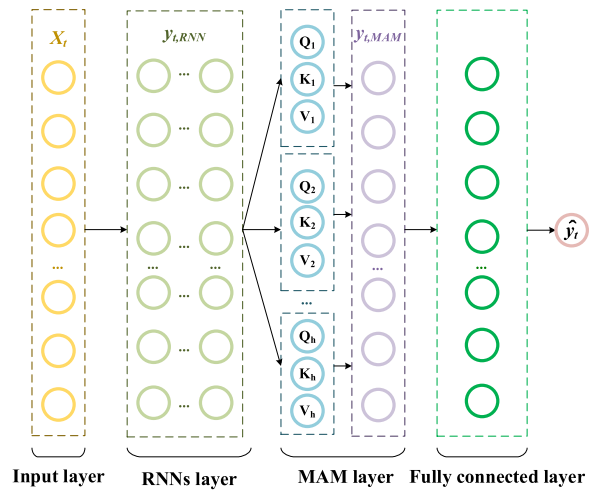


Fig. 6 Attention-enhanced recurrent neural network

(1) Input layer: At time t , the time series $X_t = [x_{t,1}, x_{t,2}, \dots, x_{t,n}]^T$ with n samples and d_k dimensions is received.

(2) RNNs layer: This layer receives the time series X_t from the input layer, and uses it as the input to the RNNs network structure. Using the ability of RNNs to process and memorize time series correlations over long periods of time, the results of sequence feature characterization are obtained without differentiating between their importance, i.e.:

$$y_{t,RNNs} = RNNs^{G_{RNNs}}(X_t) \quad (22)$$

where $RNNs(\cdot)$ represents the input-to-output mapping in RNNs, and can be derived from traditional RNN (TRNN), LSTM or GRU. The superscript G_{RNNs} represents the corresponding set of training parameters.

(3) MAM layer: this layer obtains all the feature information $y_{t,RNNs}$ about X_t from RNNs, and then extracts the key information highly associated with the target values from them while discarding useless information, so

that the subsequent networks can focus more on characterizing the target data using the key information. The input–output relationship of this layer is expressed as:

$$y_{t,MAM} = MAM^{G_{MAM}}(y_{t,RNNs}) \quad (23)$$

where $MAM(\cdot)$ represents the input to output mapping in MAM, and G_{MAM} represents the corresponding set of training parameters.

(4) Fully connected layer: this layer integrates the information computed in the MAM layer that is significantly associated with the target data, and constitutes the final output, as:

$$\hat{y}_t = \text{ReLu}(W_{fc}y_{t,MAM} + b_{fc}) \quad (24)$$

where W_{fc} and b_{fc} are the parameters to be trained in the fully connected layer, and $\text{ReLu}(\cdot)$ represents an activation function operating elementwise on vector v_k expressed as:

$$\text{ReLu}(v_k) = \max(v_k, 0) \quad (25)$$

All the training parameters in the input layer, RNN layer, MAM layer, and fully connected layer of the A-RNNs can be obtained using the temporal backpropagation algorithm.

Based on the above A-RNNs, a pseudo-measurement generation model can be constructed to fit the power series of the periodic-correlation extrapolation model, and its basic structural framework is shown in Fig. 7. Corresponding to (14), the power autocorrelation series and the external factor correlation series can be used as inputs to fit the functions $\phi_S(\cdot)$ and $\phi_E(\cdot)$ by A-RNNs respectively, and then the results of both can be integrated using a layer of fully connected layers to obtain the final pseudo-measurement generation results. This model is called the periodic correlation pseudo-measurement generation model based on A-RNNs, abbreviated as A-RNNs-PC. If the RNNs are LSTM models, they are abbreviated as A-LSTM-PC, and so on for other methods. The schematic diagrams of distributed dynamic state estimation with pseudo-measurement is shown in Fig. 8.

4 Case study

In this section, cases based on the SimBench database are analyzed to verify the effectiveness of the proposed A-RNNs-PC from the time domain and frequency domain respectively. The impact of the generated pseudo-measurement data on the state estimation accuracy of the DN is then shown. All experiments are run on a computer with Intel-i5-10400F CPU and 16 GB memory, and the programs are compiled by Python compiler.

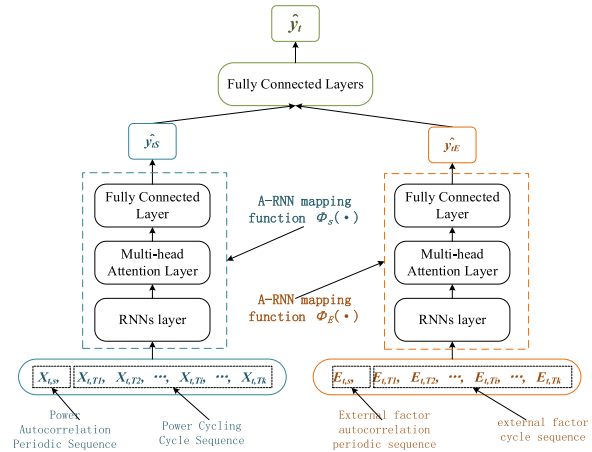


Fig. 7 Periodic correlation pseudo-measurement generation model based on attention-enhanced recurrent neural network

4.1 Data description and assessment metrics

The German SimBench project is a sub-project of the German Federal Government's "Research for an environmentally friendly, reliable and orderly energy supply" energy project, which aims to create a benchmark database for researchers to perform power system trend simulation, transient steady-state analysis, planning, etc. The SimBench database, which is built based on the actual German power system, contains network topology and parameter information of various transmission and distribution networks, as well as historical series data of generators, distributed power sources, and customer loads [13]. In this paper, some cases are selected from the SimBench database to verify the effectiveness of the proposed algorithm.

Corresponding to the time–frequency domain characterization, root mean square error (RMSE) is adopted as the time domain accuracy assessment metric. In the frequency domain, the energy spectra similarity (ESS) evaluation index is defined as:

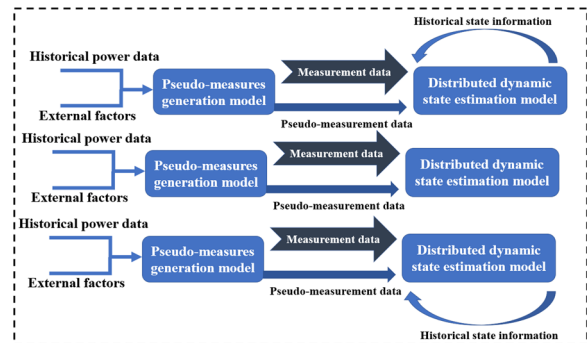


Fig. 8 Flow chart of distributed dynamic state estimation with pseudo-measurement

Table 1 Hyperparameters and algorithm characteristics of the pseudo measurement generation models

Models	Hyperparameters	Algorithm characteristic
GBDT	n_estimator=100, max_depth=3	GBDT is the state of the art of machine learning model with extremely high accuracy
RF	n_estimator=100	1. Good regression performance for high dimensional data. 2. No easy overfitting. 3. Fast Training speed, easy to make parallel method
BPNN	learning_rate=,hidden_layer=2, hidden_size=[128,128], drop_out=0.1	1. Strong robustness and fault tolerance. 2. Can fully approximate any complex nonlinear relation
TRNN, LSTM, GRU	learning_rate=0.0002–0.0005, hidden_layer=2, hidden_size=[128,128], drop_out=0.1	TRNN, LSTM, GRU are the most popular deep learning frameworks for processing time series data
A-TRNN-PC, A-LSTM-PC, A-GRU-PC	learning_rate=0.0001–0.0003, heads=2,hidden_layer=2, hidden_size=[128,128], drop_out=0.1	\

$$ESS = \frac{1}{l} \sum_{i=1}^l \sqrt{(A_{real}(f_i) - A_{pre}(f_i))^2} \quad (26)$$

where $A_{real}(i)$ and $A_{pre}(i)$ represent the energy amplitudes of the Fourier spectra of the real sequence and the pseudo-measurement sequence at the frequency f_i , respectively. The smaller the value of ESS is, the closer the spectrum of the pseudo-measurement sequence is to the spectrum of the real sequence.

4.2 Benchmark and parameter setting

Referring to [18, 19], six commonly used machine learning-based pseudo-measurement generation models are chosen, namely, Gradient boosting decision tree (GBDT), Random Forest (RF), Back propagation neural network (BPNN), TRNN, LSTM, and GRU as benchmark models. The inputs of these six benchmark models are the commonly adopted sequential inputs, i.e., continuous historical data with a lag of M time points of the target values are used as the inputs, while the inputs of A-RNNs-PC are determined by the autocorrelated maximum lag term of its periodic-correlation inputs m , the strong period $T_1, T_2, \dots, T_k$, and the number of cycles c . The A-RNNs-PC models tested in the experiments include A-TRNN-PC, A-LSTM-PC and A-GRU-PC, and the autocorrelation maximum lag term in its periodic-correlation input m is determined by the average maximum lag point with autocorrelation greater than 0.85 in the historical source-load data, while selecting the strong period from the two frequency points corresponding to the highest energy in the Fourier spectrum of the historical source-load data.

The hyperparameters of all the pseudo-measurement generation models are shown in Table 1, and their values are taken as the parameters with the smallest errors selected after multiple iterations. Among them,

n_estimator is the number of GBDT/RF estimators, max_depth is the maximum depth of GBDT, learning_rate represents the learning rate, hidden_layer and hidden_size are the number of hidden layers and the corresponding number of hidden neurons of the neural network, respectively. Heads is the number of heads of multi-headed attention, and drop_out is the neuron dropout rate, which is used to prevent the neural network from overfitting. For the neural network-based models, i.e., BPNN, TRNN, LSTM, GRU, A-TRNN-PC, A-LSTM-PC, and A-GRU-PC, Adam is adopted as the optimizer, and the number of samples for one training (batch size) is 1000. In particular, learning_rate is given as a range rather than a fixed value in Table 1 because the selection of the learning rate varies with the dataset and it usually requires several iterations to determine the optimal value. The range of learning rates listed in Table 1 is the optimal range selected after a large number of iterations, and the applicable learning rate can be determined within this range according to the dataset. The ratio of the amount of data in all training sets to the test set is 7:3.

5 Results and discussion

5.1 Analysis and modeling of load and PV output time-frequency domain characteristics

Three different sets of PV output curves and six different sets of load power curves with a sampling period of 15 min are selected from the SimBench database at random as examples to explore the characteristics of source-load power in the time domain and frequency domain. The three sets of photovoltaic (PV) output curves are recorded as PV1, PV2, PV3, and the six sets of load power curves are recorded as Load1, Load2, Load3, Load4, Load5, Load6.

The autocorrelation diagram of the series is shown in Fig. 9. Although the autocorrelation of PV output and load power varies widely, the source-load power series

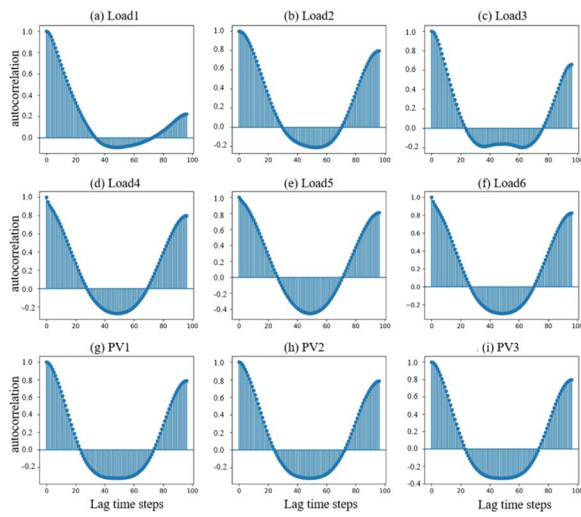


Fig. 9 Diagrams of autocorrelation of source-load power series

has a strong correlation with the data that lags itself by several time steps. This conclusion is also confirmed in [21, 22].

According to [14], external factors associated with power data generally include both temporal and climate data. Here, time and temperature, time and sunshine duration are selected as external factors affecting load power and PV output, respectively, and the JS divergence of power series and external factor series are calculated separately. The results are shown in Table 2. It is shown that power sequence has different similarity characteristics with different external factors. The load power has significant similarity with the time factor, but is not sensitive to the climate factor. The PV output is not sensitive to the time factor, while it is highly similar to the sunshine duration.

The power series not only has autocorrelation and external similarity in the time domain, but also exhibits periodic fluctuations in the frequency domain, fluctuations which can be derived from Fourier spectrum analysis. Figure 10 shows the Fourier spectra of the nine power series. It is evident that both the PV power output and

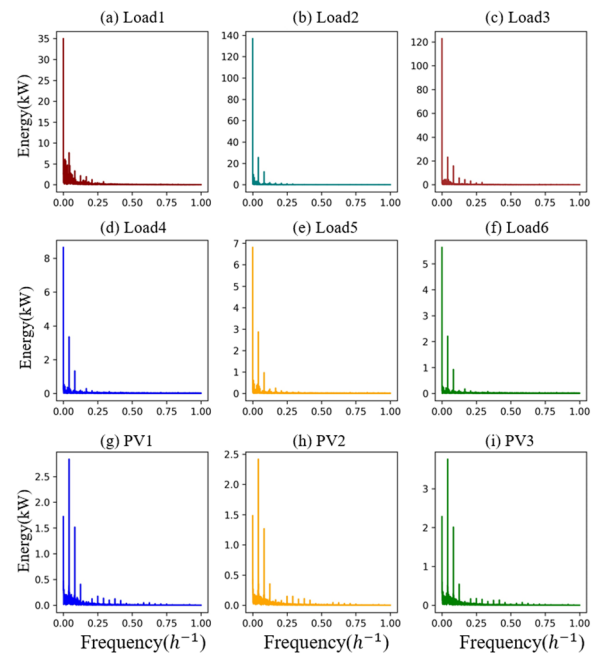


Fig. 10 Fourier spectra of source-load power sequences

load power have multiple significant frequencies, such as PV1 with higher energy at $\frac{1}{24}h^{-1}$, $\frac{1}{12}h^{-1}$ and $\frac{1}{8}h^{-1}$, and Load1 with higher energy at $\frac{1}{24}h^{-1}$ and $\frac{1}{12}h^{-1}$.

5.2 Accuracy of the A-RNNs-PC model

In this section, a 0.4 kV DN with the number "1-LV-rural3-0-no_sw" from the SimBench dataset is used as the test system. The network has 129 nodes and 127 lines, of which 118 nodes are loaded and 17 nodes are equipped with PV. Detailed network topology and data information can be found on the official SimBench project website [15]. The data contain the time series of load power and PV output in the test system from January 1, 2016 to December 31, 2016 with a collection period of 15 min. The network topology is shown in Fig. 12 in the "Appendix".

Table 2 JS values between source-load power sequences and their corresponding external factor sequences

External factors	Load1	Load2	Load3	Load4	Load5	Load6
Time factor	0.150	0.077	0.088	0.067	0.068	0.059
Climate factors	0.475	0.441	0.462	0.396	0.402	0.418
External factors	PV1	PV2	PV3			
Time factor	0.446	0.425	0.403			
Climate factors	0.107	0.144	0.108			

Table 3 Accuracy of pseudo measurements of PV output and active load under different models

Models	PV1	PV2	Load1	Load2	PV1	PV2	Load1	Load2
	RMSE ($\times 10^{-1}$ kW)				ESS ($\times 10^{-3}$ kW)			
GBDT	0.621	0.123	1.390	1.313	2.02	0.41	3.55	3.27
RF	0.606	0.119	1.378	1.319	1.73	0.34	3.41	3.27
BPNN	0.540	0.110	1.361	1.245	1.56	0.29	3.17	3.19
TRNN	0.490	0.104	1.332	1.231	1.47	0.27	3.05	3.11
LSTM	0.478	0.098	1.317	1.226	1.42	0.25	2.95	3.06
GRU	0.479	0.101	1.316	1.228	1.41	0.24	2.89	3.07
A-TRNN-PC	0.480	0.096	1.308	1.092	1.35	0.23	2.77	2.75
A-LSTM-PC	0.470	0.091	1.269	1.076	1.31	0.20	2.71	2.71
A-GRU-PC	0.471	0.090	1.272	1.075	1.33	0.20	2.73	2.71

In order to test the accuracy of the load power and PV output pseudo-measurement generated by A-RNNs-PC, the PV output data numbered "PV1" and "PV4" are chosen, and are then re-numbered as "PV1" and "PV2". For the load power data, the numbers "H0-A" and "H0-G" are chosen and are then re-numbered as "Load1" and "Load2". In addition, from the analysis results in Sect. 5.1, time and sunshine duration are selected as external factors of load and PV, respectively. The time series is the numbering sequence of sampling points in one day from 1 to 96 in chronological order. Based on the PV installation locations provided in [13], the sunshine duration data are obtained with a collection period of 10 min from the website of the German Meteorological Office for the corresponding locations, and these data are interpolated to make a sequence of sunshine duration with a collection period of 15 min. In order to avoid differences in magnitudes between different data and to eliminate the effect of different spans of values on the accuracy of pseudo-measurement generation, the values of all raw data are converted to the range [0, 1] by the maximum-minimum normalization method.

Based on the time domain evaluation index RMSE and the frequency domain evaluation index ESS, the accuracy of the pseudo-measurements generated by each model for PV output and active load under the DN is demonstrated to verify the effectiveness of the A-RNNs-PC model. All results are the average values after 10 replicate tests. The results of pseudo-measurement generation accuracy for PV output (PV1 and PV2) and active load (Load1 and Load2) under nine pseudo-measurement generation models are shown in Table 3.

In terms of the time-domain assessment metric RMSE, the accuracies of the pseudo-measurements generated by GBDT and RF are similar, with slight differences in performance with different data. However, their accuracies are generally inferior to that of the neural network-based

series models, because of the lack of learning ability of GBDT and RF for deeper features of the data. Among the various types of neural network models, BPNNs have the lowest accuracy in generating pseudo-measurements because they do not have the ability to mine temporal correlation properties in time-series data. RNNs models, i.e., TRNN, LSTM, and GRU, can compensate for this shortcoming, but their accuracies in the time domain are worse than those of LSTM and GRU because of the gradient loss/explosion problem that may occur when TRNNs deal with long time series. By comparing A-RNNs-PC model and RNNs model, it is shown that the average RMSEs of A-TRNN-PC, A-LSTM-PC, and A-GRU-PC in each source-load dataset are reduced by 6.28%, 6.24%, and 6.96% compared with TRNN, LSTM, and GRU, respectively. Clearly, the proposed A-RNNs-PC model can generate pseudo-measurement data closer to the real data. Analyzing the differences between RNNs and A-RNNs-PC structures, it can be concluded that the effective cooperation of the recurrent neural network structure with the attention mechanism, together with the periodically correlated inputs picking historical auto-correlated and externally correlated data with positive gain on the target value, enhance the ability of the neural network to characterize the data properties, enabling A-RNNs-PC to focus more on learning the data features from the valid information and then generating pseudo-measurement data with higher accuracy.

In addition to the time domain perspective, the performance of the A-RNNs-PC model can also be analyzed from the frequency domain perspective using the ESS evaluation index. From the ESS values shown in Table 3, it can be seen that the pseudo-measurement data obtained from the A-RNNs-PC model are the closest to the frequency domain characteristics of the real data, for both PV output and active load. A-LSTM-PC and A-GRU-PC have the highest frequency domain accuracy.

A-TRNN-PC is the next most accurate one, whereas the worst performers are GBDT and RF. A TRNN-PC, A-LSTM-PC, and A-GRU-PC have average ESS reductions of 10.96%, 11.81%, and 9.71% in each source-load dataset compared to those of TRNN, LSTM, and GRU, respectively. As an example, the average ESS of A-LSTM-PC is reduced by 26.14%, 21.86%, and 15.82% compared to those of GBDF, RF, and BPNN, respectively. Therefore, the proposed A-RNNs-PC model can capture more frequency domain feature information from the source-load historical data. This makes the generated pseudo-measurements close to the real data in the frequency spectrum, thus improving the accuracy of the pseudo-measurements in the frequency domain.

In summary, A-RNNs-PC has a better ability to capture the time–frequency domain characteristics of source-load power data than RNNs, and pseudo-measurement generation models based on machine learning such as GBDF, RF, and BPNN can provide more accurate pseudo-measurement data for the state estimation of DN.

5.3 Number of training parameters and training time of A-RNNs-PC model

The A-RNNs-PC model can be encapsulated into a fixed model after training, receive data from the measurement system, and generate the corresponding pseudo-measurement data by combining with the existing historical data in the database. Since the model parameters are already determined, the computation time required to generate pseudo-measurements is usually below microseconds, which has good real-time performance. However, the A-RNNs-PC model requires a large number of parameters to build, and this will occupy large storage space in the computing device. In addition, the model may need to consume some time at regular intervals to train and update the parameters to adapt to the changing characteristics of the power data at the source-load end.

Figure 11 shows the numbers of training parameters for A-RNNs-PCs, RNNs and BPNNs, as well as the running time for one epoch (i.e., all data are fed into the neural network and one forward pass and back propagation calculation is completed). It can be seen that the number of training parameters and training time of BPNN are the smallest, but its accuracy is also lower compared to A-RNNs-PC and RNNs. The training parameters of A-RNNs-PC are slightly fewer compared to RNNs, which saves storage space. This is because A-RNNs-PC trains two sub-RNNs models with power data and external factor data separately, and its periodic-correlation input method selects data with positive gain to the output as the input of the two sub-RNNs, which greatly reduces the number of neurons in the input layer. In contrast, RNNs splice power data and external factor data

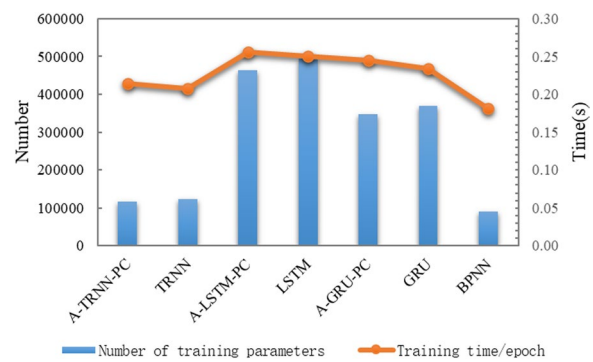


Fig. 11 Comparison of training parameters and training time of the pseudo measurement generation models based on neural network

into one-dimensional sequences for input to the model under the sequential input method, and thus, not only more input layer neurons are used, but also the number of weight parameters in the input and hidden layers is increased. In addition, A-RNNs-PCs and RNNs complete an epoch at a very similar rate, but the former generates pseudo-measurements with higher accuracy. It is shown that A-RNNs-PC can mine more effective information from the original source-load data without increasing the training parameters and training time, and quickly build a pseudo-measure generation model with superior performance.

5.4 Effect of pseudo-measurement on the accuracy of state estimation

In this subsection, the network numbered "1-EHVHV-mixed-all-0-no_sw" is selected from the SimBench database as the test system, and is then divided into one 380 V network, two 220 V networks and two 110 V networks according to the electrical connection relationship, with each sub-network renumbered. The 380 V network, whose network topology is shown in Fig. 13 in the "Appendix", is denoted as D380 and has 291 nodes, 354 branches, 222 generators distributed in 94 nodes, and 170 clean energy generators in 118 nodes, including 19 PVs and 123 wind generators. The 220 V network, as shown in Fig. 14 in the "Appendix", consists of 2 relatively independent sub-networks of D220-1 and D220-2, which are divided according to the electrical connection relationship. D220-1 has 223 nodes, 277 branches, 101 generators distributed on 52 nodes, and 6 generators distributed on 6 nodes, whereas D220-2 is smaller than D220-1, with only 57 nodes, 67 branches, 23 generators distributed in 13 nodes, and 13 clean energy generators distributed in 8 different nodes, including wind power generators only. The wind power generation equipment is included, as shown in Fig. 15 in the "Appendix". The two

110 V grid models are D110-1 and D110-2. D110-1 has 61 nodes, 59 branches, no generators, and 41 wind turbines distributed in 41 different nodes, and the network topology is shown in Fig. 16 in the “Appendix”. D110-2 has 81 nodes, 82 branches, no generators, and only 19 wind turbines distributed in 19 different nodes, and the network topology is shown in Fig. 17 in the “Appendix”.

Since this paper proposes a pseudo-measurement generation model for active power at the source-load side, it does not generate line power and node power. The measurement system shown in this table only adjusts the node power measurement of the original measurement system, i.e., reduces the ratio of nodal active/reactive power measurement, and adds nodal active pseudo-measurement. To analyze the impact of pseudo-measurement on distributed dynamic estimation accuracy, three measurement cases are considered:

- Case 1: Network measurements are sufficient and the measurement system is set up as shown in Table 4.
- Case 2: D110-1 and D110-2 are measurement-poor, PMU and SCADA measurement ratios are set as in Table 5, and no pseudo-measurement information is added.
- Case 3: D110-1 and D110-2 are measurement-poor, and the measurement ratio and pseudo-measurement ratio of the measurement system are as shown in Table 5.

Distributed dynamic state estimation is performed for this DN model in the three measurement cases. The comparison of voltage magnitude and voltage phase angle estimation accuracy and lifting degree are shown in Tables 6 and 7, respectively. Lifting degree represents the multiple of a model's calculation results superior to the standard model. Let the RMSE values of the model's calculation results be ν and the RMSE values of standard model's calculation results be ν_0 , then lifting degree $lift_{\nu, \nu_0}$ can be expressed as:

$$lift_{\nu, \nu_0} = -\frac{\nu - \nu_0}{\nu} \quad (27)$$

When the RMSE values of model results are taken as indicator, a lifting degree greater than 0 indicates that the model is better than the standard model, equaling 0 indicates that the model's performance is consistent with the standard model, whereas lower than 0 indicates that the model's calculation results are worse than the standard model.

Table 4 Measurement system settings

Area	PMU ratio (%)	SCADA Measurement ratio			Pseudo-measurement ratio
		Voltage amplitude (%)	Nodal power (%)	Line power (%)	Nodal active power (%)
D380	90	100	100	80	0
D220-1	70	100	100	80	0
D220-2	70	100	100	80	0
D110-1	40	90	90	80	0
D110-2	40	90	90	80	0

Table 5 Measurement system settings with pseudo-measurements

Area	PMU ratio (%)	SCADA Measurement ratio			Pseudo-measurement ratio
		Voltage amplitude (%)	Nodal power (%)	Line power (%)	Nodal active power (%)
D380	90	100	100	80	0
D220-1	70	100	100	80	0
D220-2	70	100	100	80	0
D110-1	40	90	40	80	50
D110-2	40	90	40	80	50

Table 6 RMSE values of voltage magnitudes obtained under three measurement conditions ($\times 10^{-4}$)

Area	Case 1	Case 2	Case 3	Lifting degree	
				Case 3 compared to Case 1 (%)	Case 3 compared to Case 2 (%)
D380	1.472	1.507	1.489	− 1.14	1.21
D220-1	1.911	1.970	1.941	− 1.55	1.49
D220-2	2.414	2.764	2.704	− 10.72	13.8
D110-1	3.712	5.540	4.868	− 23.75	13.80
D110-2	2.972	4.125	3.820	− 22.20	7.98

In Case 3, even though the pseudo-measurements are added to D110-1 and D110-2, the pseudo-measurements are limited to the active power of load, PV and wind power, and the reactive power data are lacking. Therefore, the accuracy of the pseudo-measurements is poorer than the real measurements, so the accuracy of global estimation in Case 3 is inferior to that of Case 1

Table 7 RMSE values of voltage phase angles obtained under three measurement conditions ($\times 10^{-4}$ rad)

Area	Case 1	Case 2	Case 3	Lifting degree	
				Case 3 compared to Case 1 (%)	Case 3 compared to Case 2 (%)
D380	7.005	7.152	7.089	1.20	0.89
D220-1	5.120	5.328	5.209	- 1.71	2.28
D220-2	6.266	7.169	7.013	- 10.65	2.22
D110-1	11.084	15.582	15.136	- 26.77	2.95
D110-2	10.070	14.776	13.943	- 27.78	5.57

(as shown in Tables 6 and 7). However, the inclusion of pseudo-measurements in Case 3 significantly improves the estimation accuracy of D110-1 and D110-2 compared to Case 2, which results in a positive effect on the boundary-consistent estimation results, leading to a small increase in the estimation accuracy of other sub-areas, as can be seen in Tables 6 and 7.

In summary, although the pseudo-measurement data obtained by the A-RNN pseudo-measurement generation model can hardly help the state estimation to reach the accuracy under the real measurement, it can improve the estimation without increasing the economic cost, and alleviate the negative impact of local measurement shortage on the global estimation of DN.

6 Conclusion

In order to alleviate the negative impact of insufficient measurement of DN on the accuracy of state estimation, and based on the solution idea of generating pseudo-measurement, this paper analyzes the time–frequency domain characteristics of the source-load historical power sequence, and the autocorrelation of the sequence and the probability similarity with external factors using autocorrelation function and JS dispersion. The periodic fluctuation of the sequence from the frequency domain is verified using Fourier spectrum. A-RNNs are then chosen to automate the representation of the mapping function of the model, so that an A-RNNs-PC pseudo-measurement generation model can be established for mining the power time–frequency domain characteristics of the source-load end. Finally, the generated pseudo-measurement data are used as the input to the distributed dynamic state estimation model.

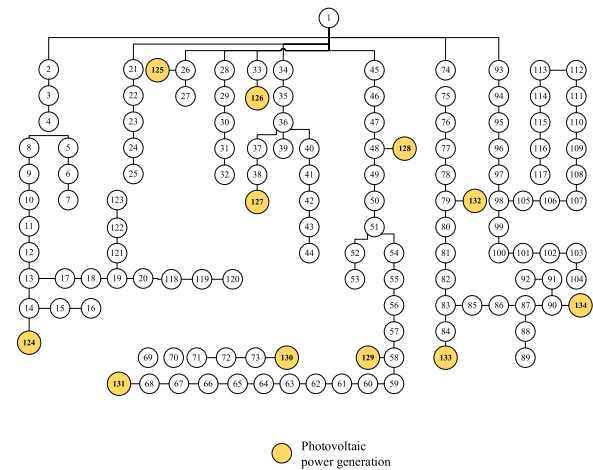
The experiments validate the performance of the proposed A-RNNs-PC pseudo-measurement generation model based on the load and PV output temporal data of the distribution network in the SimBench

database. The results show that A-RNNs-PC exhibits better accuracy in both time and frequency domains than six commonly used machine learning-based pseudo-measurement generation models, including RF, GBDT, BPNN, TRNN, LSTM, and GRU. In addition, it is also verified that A-RNNs-PC can obtain pseudo-measurements with higher accuracy using similar training parameters and training time as RNNs. The generated pseudo-measurement data can improve the accuracy of distributed dynamic state estimation without adding additional economic burden.

Considering the non-synchronism of measurement data caused by the different sampling frequencies and communication delays of measurement equipment in the state estimation of distribution network, as well as three-phase imbalance in some regions caused by imbalance of line parameters, single-phase load access etc., the solution of asynchronous measurement data in state estimation and distributed dynamic state estimation in the condition of three-phase imbalance of a local network will be explored in the future.

Appendix

See Figs. 12, 13, 14, 15, 16, and 17.

**Fig. 12** Network topology of "1-LV-rural3-0-no_sw"

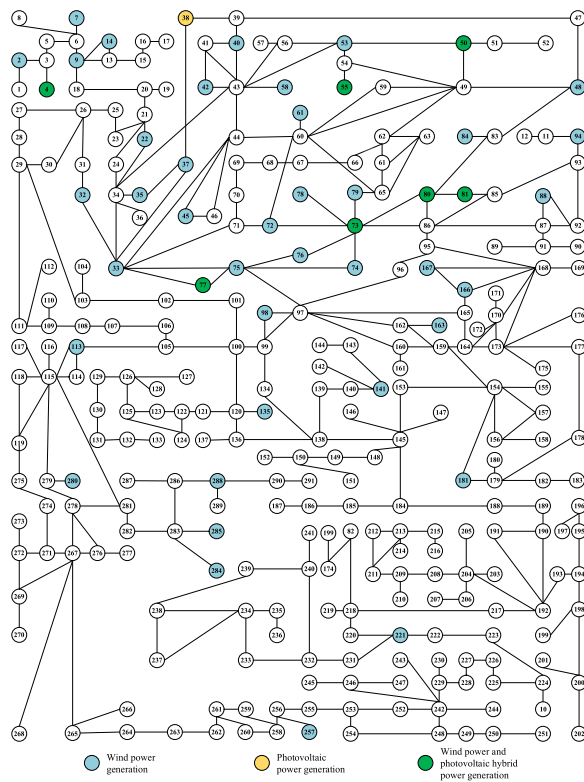


Fig. 13 Network topology of D380

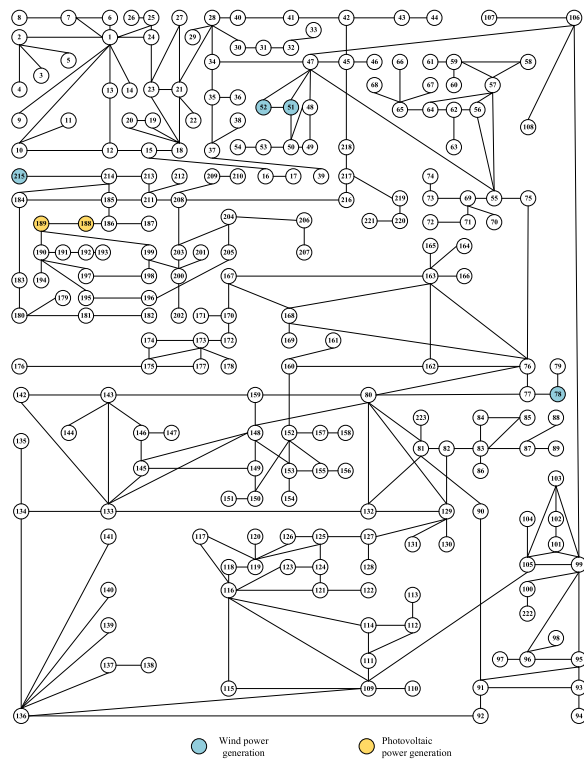


Fig. 14 Network topology of D220-1

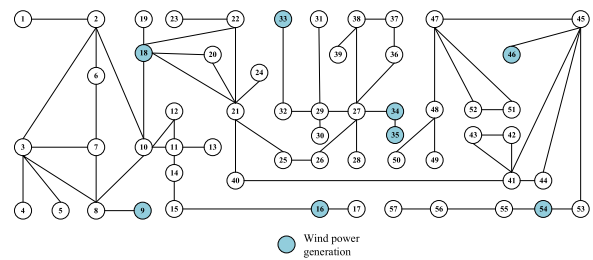


Fig. 15 Network topology of D220-2

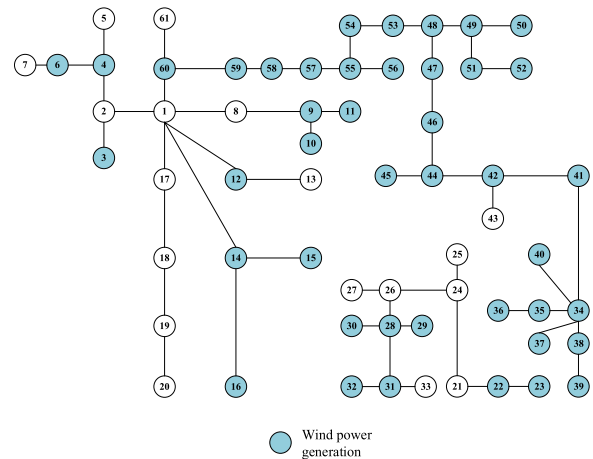


Fig. 16 Network topology of D110-1

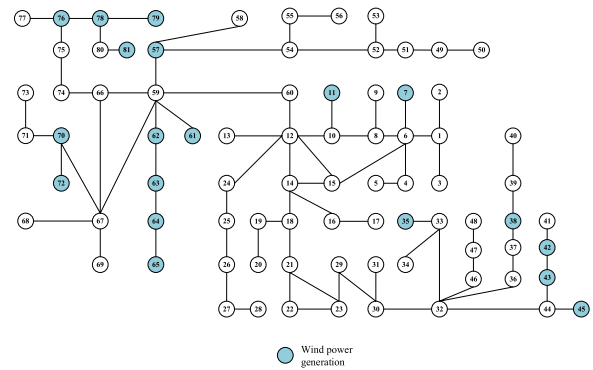


Fig. 17 Network topology of D110-2

Acknowledgements

Not applicable.

Author contributions

Demand research, literature survey, experiment design, data collection, mathematical modeling, case study, and paper writing. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Key Research Program of China (2016YFB0900100) and Key Project of Shanghai Science and Technology Committee (18DZ1100303).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations**Competing interests**

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.

Received: 13 October 2022 Accepted: 25 June 2023

Published online: 07 July 2023

References

- Meng, Y., Fan, S., Zheng, X., Xiao, J., Zhou, H., & He, G. (2022). Optimal operation of virtual power plant considering power-to-hydrogen systems. In: *2022 IEEE power & energy society general meeting (PESGM), Denver, CO, USA*, pp. 1–5.
- Do Coutto Filho, M. B., & Stacchini de Souza, J. C. (2009). Forecasting-aided state estimation—part I: Panorama. *IEEE Transactions on Power Systems*, 24(4), 1667–1677.
- Angioni, A., Schlösser, T., Ponci, F., & Monti, A. (2016). Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids. *IEEE Transactions on Instrumentation and Measurement*, 65(1), 70–77.
- Wu Zaijun, Xu, Xinghuo, J. Y., Qinran, Hu., Xiaobo, D., & Wei, Gu. (2017). Review of active distribution network state estimation technology. *Automation of Electric Power Systems*, 13, 182–191.
- Zhao, J., Zhang, G., Dong, Z. Y., & La Scala, M. (2018). Robust forecasting aided power system state estimation considering state correlations. *IEEE Transactions on Smart Grid*, 9(4), 2658–2666.
- Jikang, L., Danfeng, S., Xianghong, T., & Yangsheng, L. (2019). Stochastic state estimation based on a new method of adaptive sparse pseudo spectral approximation model and algorithm. *Chinese Journal of Electrical Engineering*, 39(01), 192–203.
- Dobbe, R., van Westering, W., Liu, S., Arnold, D., Callaway, D., & Tomlin, C. (2020). Linear single- and three-phase voltage forecasting and Bayesian state estimation with limited sensing. *IEEE Transactions on Power Systems*, 35(3), 1674–1683.
- Dehghanpour, K., Yuan, Y., Wang, Z., & Bu, F. (2019). A Game-theoretic data-driven approach for pseudo-measurement generation in distribution system state estimation. *IEEE Transactions on Smart Grid*, 10(6), 5942–5951.
- Zhang, L., Wang, G., & Giannakis, G. B. (2019). Distribution system state estimation via data-driven and physics-aware deep neural networks. *IEEE Data Science Workshop (DSW)*, 2019, 258–262.
- Zhenghao, L. I., & Mengfan, L. I. (2020). Smart load forecasting method based on deep learning. *Smart Power*, 48(10), 78–85.
- Yuan, L., Ma, J., Gu, J., Wen, H., & Jin, Z. (2020). Featuring periodic correlations via dual granularity inputs structured RNNs ensemble load forecaster. *International Transactions on Electrical Energy Systems*, 30, e12571.
- Dan, S. (2006). *Optimal state estimation* (pp. 123–140). New York: Wiley.
- Fox, V., Hightower, J., Liao, L., et al. (2003). Bayesian filtering for location estimation. *IEEE Pervasive Computing*, 2(3), 24–33. <https://doi.org/10.1109/MPRV.2003.1228524>
- Meinecke, S., Drauz, S., Klettke, A, et al. (2020). SimBench Documentation [M/OL], Germany. https://simbench.de/wpcontent/uploads/2021/09/simbench_documentation_en_1.1.0.pdf.
- Aprillia, H., Yang, H.-T., & Huang, C.-M. (2021). Statistical load forecasting using optimal quantile regression random forest and risk assessment index. *IEEE Transactions on Smart Grid*, 12(2), 1467–1480.
- SimBench. Complete list of SimBench Codes and their grid data[DB/OL]. <https://simbench.de/en/download/datasets/>
- Primadianto, A., & Lu, C. (2017). A review on distribution system state estimation. *IEEE Transactions on Power Systems*, 32(5), 3875–3883. <https://doi.org/10.1109/TPWRS.2016.2632156>
- Monticelli, A. (2000). Electric power system state estimation. *Proceedings of the IEEE*, 88(2), 262–282. <https://doi.org/10.1109/5.824004>
- Chen, B., Li, H., & Su, X. (2019). Dynamic state estimation of distribution network base on pseudo measurement modeling and UPF. *IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2019, 54–59. <https://doi.org/10.1109/ISGT-Asia.2019.8881190>
- Qiang, Q., Guoqiang, S., Wei, X., Minghui, Y., Zhinong, W., & Haixiang, Z. (2018). Distribution system state estimation based on pseudo measurement modeling using convolutional neural network. *China International Conference on Electricity Distribution (CICED)*, 2018, 2416–2420. <https://doi.org/10.1109/CICED.2018.8592565>
- Guo, Z., Li, S., Wang, X., et al. (2016). Distributed point-based gaussian approximation filtering for forecasting-aided state estimation in power systems. *IEEE Transactions on Power Systems*, 31(4), 2597–2608. <https://doi.org/10.1109/TPWRS.2015.2477285>
- Zhu, X., & Genton, M. G. (2012). Short-term wind speed forecasting for power system operations. *International Statistical Review*, 80(1), 2–23.
- Shi, H., Xu, M., & Li, R. (2017). Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Transactions on Smart Grid*, 9(5), 5271–5280.

Yaojian Wang (1999), male, received the bachelor's degree in electrical engineering in 2021 from Shanghai Jiao Tong University, Shanghai, where he is currently working toward the master degree with the Research Center for Big Data Engineering Technology, and majors in machine learning applications in power system, time series forecasting, uncertainty quantification and distribution network status estimation.

Jie Gu (1971), female, PHD and Associate Professor at Shanghai Jiao Tong University, Major in power system planning, power markets, smart grids, microgrids, and intelligent algorithms.

Lyuzerui Yuan (1995), female, received the B.E. degree in electrical engineering and automation from South China University of Technology, Guangdong, China, in 2017. She is currently pursuing the Ph.D. degree in electrical engineering at Shanghai Jiao Tong University. Her current research interests are in situational awareness, deep learning and edge computing, with applications to distribution power system.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)