electronics

MDPI

Article

Digital Twin for Modern Distribution Networks by Improved State Estimation with Consideration of Bad Date Identification

Huiqiang Zhi [1], Rui Mao [1], Longfei Hao [1], Xiao Chang [1], Xiangyu Guo [1] and Liang Ji [2,*]

[1] Electric Power Research Institute, State Grid Shanxi Electric Power Company, Taiyuan 730087, China; zhihuiqiang@sx.sgcc.com.cn (H.Z.); maorui@sx.sgcc.com.cn (R.M.); haolongfei@sx.sgcc.com.cn (L.H.); changxiao@sx.sgcc.com.cn (X.C.); guoxiangyu@sx.sgcc.com.cn (X.G.)

[2] College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China

* Correspondence: jiliang@shiep.edu.cn

Abstract: With the rapid development of modern power systems, the structure and operation of distribution networks are becoming increasingly complex, demanding higher levels of intelligence and digitization. Digital twin, as a virtual cutting-edge technique, can effectively reflect the operational status of distribution networks, offering new possibilities for real-time monitoring, optimization and other functions for distribution networks. Building efficient and accurate models is the foundation of enabling a digital twin of distribution networks. This paper proposes a digital twin operating system for distribution networks with renewable energy based on robust state estimation and deep learning-based renewable energy prediction. Furthermore, the identification and correction of possible bad or missing data based on deep learning are also included to purify the input data for the digital twin system. A digital twin test platform is also proposed in the paper. A case study and evaluations based on a real-time digital simulator are carried out to verify the accuracy and real-time performance of the established digital twin system. In general, the proposed method can provide the basis and foundation for distribution network management and operation, as well as intelligent power system operation.

Keywords: digital twin; state estimation; bad data identification; power system modeling

check for updates

Citation: Zhi, H.; Mao, R.; Hao, L.; Chang, X.; Guo, X.; Ji, L. Digital Twin for Modern Distribution Networks by Improved State Estimation with Consideration of Bad Date Identification. *Electronics* **2024**, *13*, 3613. https://doi.org/10.3390/electronics13183613

Academic Editor: Ahmed Abu-Siada

Received: 5 August 2024
Revised: 9 September 2024
Accepted: 9 September 2024
Published: 11 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The transformation of the traditional power grid into a smart grid, along with the continuous development of distributed energy resources, has gradually evolved the traditional distribution network into an active distribution network with bi-directional power flow [1]. This has led to an increase in the complexity of the operation state and control mode. With the advent of Industry 4.0, the introduction of intelligent devices and systems has made the intelligent transformation of distribution networks possible and provided a pathway for various emerging digital technologies. Digital twins (DTs), as a part of Industry 4.0, play an important role. By utilizing advanced IoT, big data analytics, and AI technologies, DT enhances production efficiency and product quality while providing strong support for predictive maintenance and optimized decision-making [2]. The application of this technology is not limited to manufacturing; the use of DT in power systems marks a significant step towards smarter, more efficient, and more resilient energy infrastructure. By leveraging real-time data, advanced analytics, and predictive capabilities, DT technology transforms the management and operation of power systems, paving the way for a more sustainable energy future.

A DT is a virtual mapping of physical entities. It constructs models of physical objects using a data-model hybrid-driven approach based on measured data to achieve functions such as real-time state perception, future state projection, closed-loop control, and other functions. DTs have been gradually applied in practice in many fields such as manufacturing, aeronautics and space and urban management [3]. In the field of

power systems, DT technology has been the subject of initial research and subsequent application. Reference [4] develops a second response grid DT online analysis system, which achieves millisecond data processing in the grid. Reference [5] gives predominant focus to a DT-based longitudinal protection method for DC grids. This method compares real measurements with virtual measurements through a dynamic state estimation method. Reference [6] offers a DT-driven multi-agent coordinated optimization control strategy of the smart microgrid. This enhances the ability of microgrids to be aware, predict and adapt. Reference [7] establishes a digital twin model of the steam turbine system in a thermal power plant, optimizing the system through the indicators provided by the model and effectively improving efficiency. Reference [8] achieves the precise localization of real-time cyber attacks by establishing a digital twin reference model for the distribution network. Reference [9] proposes a digital twin-based distributed energy coordination control method, minimizing the need for real-time communication and achieving the overall coordination of distributed energy resources.

The DT relies on the real-time input of distribution network data for modeling and analysis, so the accuracy of the measurement data is the main factor affecting the modeling of the DT distribution network. In practice, the measurement data are not completely real, and a small amount of dirty data enters the measurement data set due to sensor damage, data attacks, etc., which seriously affects the accuracy of distribution network modeling. Large-scale access to renewable energy makes distribution network measurement more complex and diverse, with randomness and volatility, which also poses challenges to the identification of bad data. Traditional bad data identification methods such as the residual search identification method [10], non-quadratic criterion identification method [11], and estimation identification method [12] are used to identify bad data after state estimation operations, which are inefficient and prone to identify fluctuating or sudden changes in the normal measurement data as bad data, contributing to inaccuracies in the DT distribution network and influencing the subsequent dispatching and coordinated control of the distribution network.

State estimation (SE) is a crucial part of state perception and secure and stable operation of distribution network and it is also the foundation for the DT distribution network modeling. Commonly used distribution network SE algorithms are a power branch method based on weighted least square (WLS) or weighted least absolute value (WLAV) [13], as well as SE algorithms that consider robustness [14–16]. These algorithms are not applicable to the complex situation of multiple sources of real measurement data and the computation process is complicated and time-consuming making it fail to meet the needs for real-time awareness of DT distribution networks. In [17], the authors combine WLS and WLAV to make the SE model more robust, but the computation is slow. In [18], the branch current state estimation (BCSE) is used, which sets the measurement and state variables as current data, to make the SE linear and speed up the iteration, but it still needs to go through flow calculation to solve the voltage state quantity. Reference [19] applied linear Bayesian theory to SE, analyzing the impact of measurement errors on the algorithm and improving the performance of state estimation. With the development of artificial intelligence technology, many scholars have started to adopt machine learning (ML) methods to estimate the state of distribution networks, such as References [20–22], etc. However, due to the complexity of the distribution network model, it is difficult to establish an accurate mapping by the learning method, and it requires a large amount of historical data support.

This paper adopts SE as the method for the DT modeling of distribution networks with a high proportion of renewable energy, capturing the real-time state of the physical distribution network and identifies and corrects the distribution network measurement data prior to SE. The contributions of this paper are as follows:

- A novel bad data identification and correction method is proposed, which identifies bad data based on temporal correlations and completes the data using a neural network training approach.

- A linear SE model incorporating photovoltaic (PV) integration is established. The PV output is predicted using a neural network, and complete data for the distribution network are obtained through a linear SE algorithm based on multi-source data fusion.
- A DT-based SE model and database for the distribution network were established on the server side and were run synchronously with the simulated physical model of the real-time digital simulator (RTDS), verifying the accuracy and real-time performance of the DT SE model.

2. DT for Distribution Network

The power system digital twin (PSDT) is an emerging technology driven by the increasing complexity of power system models, the dramatic growth of data, and the gradual improvement of DT technology, as illustrated in Figure 1. In contrast to traditional model-based simulation software and cyber physical systems, PSDT focuses more on real-time situational awareness and super-real-time virtual testing through data-driven or hybrid data-model-driven approaches to support power system operation and regulatory decision making. Especially in the smart grid domain, PSDT can provide more precise and real-time system state monitoring and analysis, enabling the smart grid to achieve self-optimization and intelligent decision-making, thereby improving grid reliability.

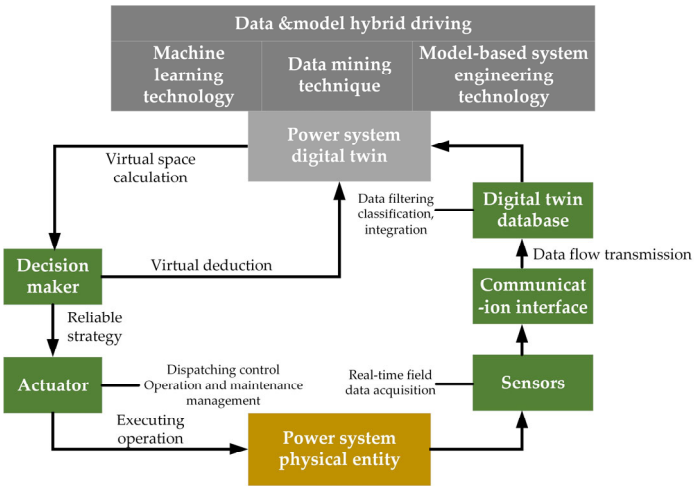


Figure 1. Power system digital twin.

The DT architecture for distribution networks can be divided into four layers: the physical distribution network layer, the data processing layer, the twin distribution network layer and the twin application layer [3]. The schematic diagram of the distribution network DT hierarchical architecture is shown in Figure 2. The physical distribution layer is the information source of the DT distribution network and is responsible for metering based on multi-source metering devices and transmission of distribution network data; Based on the data received from the physical distribution network layer, the data processing layer performs multi-source data fusion processing; the twin distribution layer implements real-time sensing of the physical distribution network, as well as renewable energy on the network, and continuously updates the twin model to reflect operational status in a timely manner; the twin application layer provides diversified solutions for various application scenarios in the distribution network based on the DT distribution network model. The accurate DT model is the groundwork for the DT distribution network; thus, the data-processing layer and the twin distribution network layer are crucial. The effectiveness of data processing and model establishment directly impacts the accuracy and real-time nature of the DT model, which is also the focus of this study.

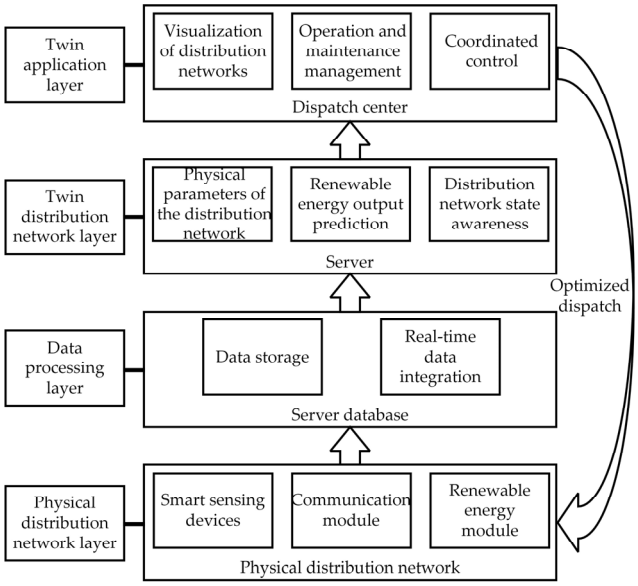


Figure 2. The schematic diagram of the distribution network DT hierarchical architecture.

3. Identification and Correction of Bad Measurement Data Using BILSTM

3.1. Bad Data Identification

Distributed energy access makes the distribution network measurement data fluctuate, and it is difficult to distinguish bad data from the sudden change data only by measuring autocovariance identification, and it is easy to have omissions and misjudgments. Therefore, it is necessary to identify the bad data in real-time measurement data sets by combining the historical pattern of distribution network measurement data and using the inertia of distribution network operation.

Due to the topology of the physical distribution network, there is a temporal correlation between different measurement data in the same area of the distribution network. A time series of measurements of length T can be defined as $[z_{k+2-T}, z_{k+3-T}, \cdots, z_{k+1}]$. It is formed by extracting from the DT database a period of historical measurements phases of length $T-1$ before the current moment, together with real-time measurements of phases (z_{k+1}) at that moment. Calculating the Pearson correlation coefficient for the time series of this measure and creating a matrix of Pearson correlation coefficients based on the correlation coefficients of each two measures are as follows:

$$\rho_{ij} = \begin{cases} \dfrac{\sum_{t=1}^{T}(z_{k+2-t}(i)-\overline{z}(i))(z_{k+2-t}(j)-\overline{z}(j))}{\sqrt{\sum_{t=1}^{T}(z_{k+2-t}(i)-\overline{z}(i))^2}\sqrt{\sum_{t=1}^{T}(z_{k+2-t}(j)-\overline{z}(j))^2}} & i \neq j \\ 1 & i = j \end{cases} \tag{1}$$

where ρ_{ij} is the Pearson correlation coefficient, and $-1 < \rho_{ij} < 1$. ρ_{ij} indicate the temporal correlation between i and j, and the matrix diagonal element is 1. The larger the value of $|\rho_{ij}|$ is, the stronger the temporal correlation between i and j is. $\overline{z}(\cdot)$ is the average value of time series measurement sequence and $z_{k+2-t}(\cdot)$ is the measurement of $k+2-t$ moment. Calculate the average of the measured correlation coefficients $\overline{\rho}_i$:

$$\overline{\rho}_i = \frac{1}{N-1}\sum_{j=1, j\neq i}^{N}|\rho_{ij}| \tag{2}$$

It is verified by simulation that when there are fluctuations in the distribution network, due to a strong correlation, the average value of the correlation coefficient of the normal measurement data is much higher than that of the bad data. The maximum value $\overline{\rho}_{\min}$ is

selected in the historical database when bad data are generated, the minimum value $\overline{\rho}_{\max}$ when there are no bad data, and the threshold value $\alpha \in (\overline{\rho}_{\min}, \overline{\rho}_{\max})$.

If $\overline{\rho}_i < \alpha$, these data are then flagged as bad and rejected.

3.2. Bad Data Correction Based on BILSTM

3.2.1. Structure of BILSTM

Once the bad data have been identified, they need to be corrected by replacing the bad data to obtain the complete measurement data set to ensure the observability of the distribution network. In this paper, the historical measurement data in the DT database are trained to generate replacement data to correct the bad data.

Long short-term memory (LSTM) is a deep learning model for sequential data processing. It efficiently captures and maintains long-term dependencies in sequential data through gating mechanisms including input gate, forgetting gate and output gate. LSTM has a wide range of applications in time series forecasting, natural language processing, etc. It is able to handle long time series data effectively and achieve excellent performance in many tasks. Figure 3 shows the structure of the LSTM.

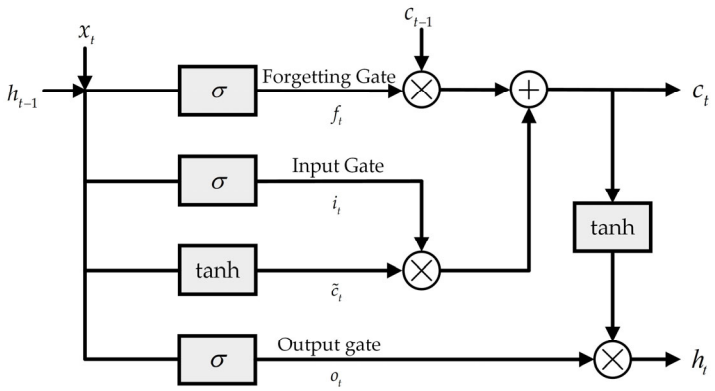


Figure 3. LSTM structure.

As new energy access or load fluctuations lead to fluctuations in distribution network measurement data, the traditional LSTM is unable to capture the correlation on the measurement time scale and does not have better predictive performance. Bidirectional long short-term memory (BILSTM), an extension of LSTMs, combines forward and reverse information flow for modeling sequential data. BILSTM includes two main parts: one LSTM model processing past input features in the forward direction, and another LSTM handling future input features in the reverse direction. And BILSTM effectively captures the contextual dependencies in the sequence by splicing the hidden states to obtain a bidirectional representation of each time step. BILSTM's wide range of applications in natural language processing, time series prediction and speech recognition can enhance the model's ability to understand and characterize sequence data. The problem of quantitative prediction can be solved by using BILSTM to better capture the properties of bidirectional data dependence. The structure of the BILSTM is in Figure 4:

The hidden layer of the BILSTM has to store two values, the forward parameter h_t and the reverse parameter h_k, which refer to the parameter h_{t-1} on the left side and the parameter h_{k-1} on the right side, respectively. The final output of the BILSTM y_t depends on the forward parameter and the reverse parameter, and the key equations are as follows:

$$\begin{cases} y_t = g(Kh_t + K'h_k) \\ A' = L(W'h_{k-1} + U'x_t) \\ A = L(Wh_{t-1} + Ux_t) \end{cases} \tag{3}$$

where y_t is the output at time t; x_t is the input at time t; $g(\cdot)$ and $L(\cdot)$ are both sigmoid activation function; the parameters K, K', W, W', U and U' are weight values corresponding

to the input values; and h_{t-1} and h_{k-1} are LSTM-processed values determined by the state at the previous instant.

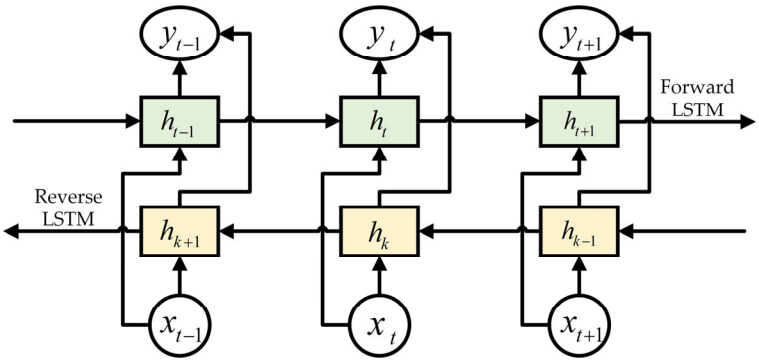


Figure 4. BILSTM structure.

3.2.2. BILSTM Bad Data Correction Process

Firstly, the historical data of distribution network measurements, including distribution network branch power and current, node injection power, voltage amplitude and phase angle measurements, are processed in 10 consecutive time point cycles for data dimensional reconstruction. For example, measurement sample $[z_1, z_2, z_3, ..., z_{10}]^T$ is converted to a new measurement sample Z_1, measurement sample $[z_2, z_3, z_4, ..., z_{11}]^T$ is converted to a new measurement sample Z_2, and so on.

Secondly, the reconstructed sample data set is fed into the BILSTM model to obtain the trained BILSTM model. The parameters of the model are as follows: the number of layers of the BILSTM network is 2; the number of neurons is 64 and the activation function is sigmoid; the model optimizer is RMSprop, and the learning rate is 0.001; the loss function is MSE, the number of iterations is 150, and the batch size is 64.

Finally, the measurement samples $[z_{k+1-T}, z_{k+2-T}, \cdots, z_k]$ from time k + T − 1 to time k are reconstructed and fed into the BILSTM model to obtain the predicted measurement vector z_{k+1} at time k + 1. Replacing the rejected bad data with the predicted values completes the bad data correction. The flowchart for BILSTM-based bad data correction is depicted in Figure 5.

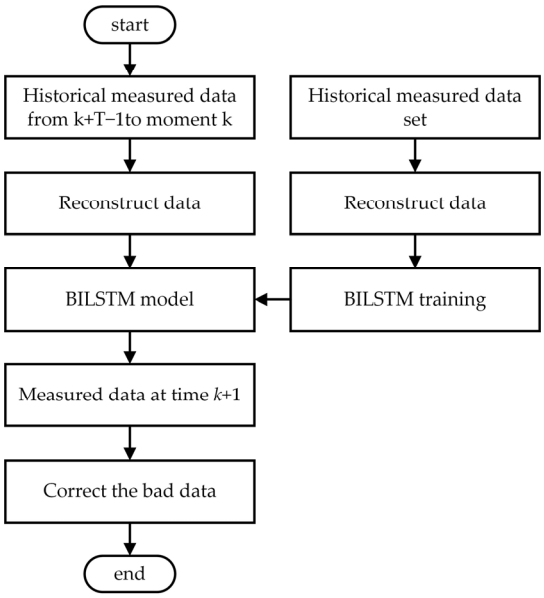


Figure 5. Flowchart for BILSTM-based bad data correction.

4. DT Modeling Using State Estimation with PV Forecasting

In the actual distribution network, there are three main types of measurement devices: phasor measurement units (PMUs), supervisory control and data acquisition (SCADA) systems, and new energy pseudo measurement devices. The type and accuracy of measurement data from different devices can vary significantly. In order to apply multisource measurement data in a reasonable manner with the objective of increasing SE redundancy and ensuring network observability, this paper employs a neural network to predict PV power in order to establish a pseudo measurement model of the PV system. Thereafter, the different measurement data are computed after linear transformation aiming to establish an efficient and precise state model of the distribution network.

4.1. PV Power Generation Prediction Method Based on BILSTM

Due to external environmental factors and other influences, the output power of PV power plants is subject to significant randomness and uncertainty, which can lead to a non-negligible error in the SE. Conventional new energy prediction methodologies [23–25] are constrained in their ability to provide accurate real-time output prediction data in the presence of more complex fluctuations. In contrast, deep learning-based PV prediction methodologies are capable of achieving satisfactory prediction outcomes in a diverse range of scenarios. However, the traditional artificial neural network (ANN), back propagation neural network (BPNN), is not optimal for handling time-series problems and may not be able to effectively capture the long-term trends and periodicity in PV power generation. Consequently, this paper employs the BILSTM neural network, as detailed in Section 2, to predict the real-time PV output. Figure 6 illustrates this process.

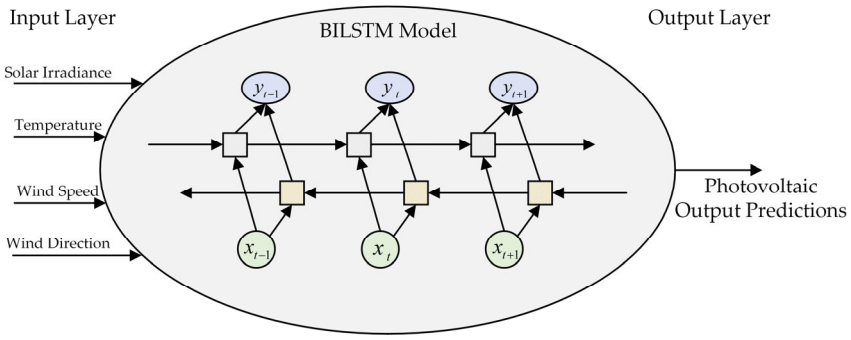


Figure 6. Schematic diagram of BILSTM PV power system output prediction.

The method utilizes meteorological statistics of PV solar irradiance, temperature, wind speed and wind direction as inputs, with corresponding historical output data serving as outputs and sets the input time series step at 10 in order to train the BILSTM neural network. Given that the meteorological data units and sizes of the input layers differ, a normalization process is required. This is achieved through the application of the following equation:

$$\widetilde{D} = \frac{D - D_{min}}{D_{max} - D_{min}}, \tag{4}$$

where $\widetilde{D}$ represents the normalized number. D represents the original measurement data, including all PV solar irradiance, temperature, wind speed, and wind direction data. D_{max} and D_{min} denote the maximum and minimum values of the data, respectively. All normalized data are in the range of 0 to 1. Using normalized data as input can accelerate the training process of neural networks, improve prediction accuracy, and prevent gradient vanishing or exploding problems.

The Input of real-time meteorological data into the trained model will result in the generation of the real-time PV system output pseudo-measurement model, which is required for the SE.

4.2. Linear Transformation of Measured Data

This paper uses a measurement transformation method for converting the initial quantity measurements of the distribution network uniformly into voltage and current measurements that are separated into real and imaginary parts in a rectangular cartesian coordinate system. In this instance, the PMU is responsible for measuring the node voltage phasors and branch current phasors, and subsequently performing a linear transformation of these quantity measurements using the following equation.

$$U_{i-re} + jU_{i-im} = \left|\dot{U}_{i-P}\right| \cos\theta_{i-P} + j\left|\dot{U}_{i-P}\right| \sin\theta_{i-P}, \tag{5}$$

$$I_{ij-re} + jI_{ij-im} = \left|\dot{I}_{ij-P}\right| \cos\varphi_{ij-P} + j\left|\dot{I}_{ij-P}\right| \sin\varphi_{ij-P}, \tag{6}$$

where $\left|\dot{U}_{i-P}\right|$ and θ_{i-P} are the voltage phasors measured by the PMU; $\left|\dot{I}_{ij-P}\right|$ and φ_{ij-P} are the current phasors measured by the PMU.

In a similar manner, the SCADA measurements are converted. The voltage and current amplitudes measured by SCADA are calculated based on Equations (5) and (6). It is possible to convert the more frequent branch power measurements in SCADA measurements to equivalent branch current measurements as follows.

$$I_{ij-re} = \frac{P_{ij-S}V_{i-re} + Q_{ij-S}V_{i-im}}{(V_{i-re})^2 + (V_{i-im})^2}, \tag{7}$$

$$I_{ij-im} = \frac{P_{ij-S}V_{i-re} - Q_{ij-S}V_{i-im}}{(V_{i-re})^2 + (V_{i-im})^2}, \tag{8}$$

where P_{ij-S} and Q_{ij-S} are the branch active and reactive power measured by SCADA, respectively. V_{i-re} and V_{i-im} are the real part and imaginary part of the node voltage obtained at each iteration, respectively.

In addition, the node injection power measurements in SCADA measurements and new energy pseudo measurements can be converted into equivalent node injection current measurements as:

$$I_{i-re} = \frac{P_i V_{i-re} + Q_i V_{i-im}}{(V_{i-re})^2 + (V_{i-im})^2}, \tag{9}$$

$$I_{i-im} = \frac{P_i V_{i-re} - Q_i V_{i-im}}{(V_{i-re})^2 + (V_{i-im})^2}, \tag{10}$$

where the values of node injection power, both actual and pseudo-measured, are included in both P_i and Q_i.

4.3. Linear State Estimation Model for Distribution Network

After the above series of measurement transformations, the SE measurement equation is constructed as:

$$z = h(x) + v, \tag{11}$$

where z is the measurement vector. The measurement vector function is represented by $h(x)$. And v is the error vector for z and $h(x)$.

The state variable x is set to the real and imaginary parts of the voltages of the n nodes of the distribution network, where:

$$x = [V_{1,re}, V_{1,im}, ..., V_{n,re}, V_{n,im}]^T. \tag{12}$$

The z is shown below:

$$z = [z_{U_{i-\mathrm{re}}}, z_{U_{i-im}}, z_{I_{i-re}}, z_{I_{i-im}}, z_{I_{ij-re}}, z_{I_{ij-im}}]^T, \tag{13}$$

where $z_{U_{i-\mathrm{re}}}$ and $z_{U_{i-im}}$ are the converted real part and imaginary part voltage measurement. $z_{I_{i-re}}$ and $z_{I_{i-im}}$ are the converted node injection current real part and imaginary part measurement. $z_{I_{ij-re}}$ and $z_{I_{ij-im}}$ are the converted branch current real part and imaginary part measurement.

Once the distribution network is connected to the PV system, the corresponding node injects active and reactive power. However, the inverter connected to the PV power supply generates reactive power at a high cost, which means that the PV system usually only outputs active power. Consequently, the reactive power $Q_{PV}(x)$ emitted by the PV system in the distribution network is zero. The SE model with equation constraints can be constructed by using the weighted least squares method and considering the reactive power injection constraints at PV nodes. And this model can be described by the following equation.

$$\begin{cases} \min\left\{J(x) = (z - h(x))^{\mathrm{T}} R^{-1}(z - h(x))\right\} \\ s.t.c(x) = Q_{PV}(x) = 0 \end{cases} \tag{14}$$

where $c(x)$ represents the zero injection node constraint function, and $J(x)$ denotes the least squares estimation objective function.

The Lagrange multiplier method is used for the solution of the above mathematical model. The Lagrangian extremal function is written here as

$$L(x, \lambda) \equiv \frac{1}{2} J(x) - \lambda^T c(x), \tag{15}$$

where $L(x, \lambda)$ is the objective function.

The partial derivatives for the state variable x and the constrained phase λ are as follows

$$\begin{cases} L_x \equiv -\partial L/\partial x = H^T R_M^{-1}[z - h(x)] + C^T \lambda = 0 \\ L_\lambda \equiv -\partial L/\partial \lambda = c(x) = 0 \end{cases} \tag{16}$$

where C is the zero injection measurement constrained Jacobi matrix, and R_M is the converted fused covariance matrix.

Following the implementation of a linear transformation on a measurement, the Jacobi matrix H can be expressed as

$$\boldsymbol{H}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \\ \boldsymbol{G} & -\boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{G} \end{bmatrix} \tag{17}$$

where $\boldsymbol{G}$ is the branch conductance. $\boldsymbol{B}$ is the branch susceptance.

It is known that the Jacobi matrix H is a constant coefficient matrix that remains constant during the iteration process, which speeds up the computation of SE. Newton's method is employed to resolve Equation (16) in order to ascertain the state variable x^{k+1} for the k + 1st iteration.

If the state variables satisfy $\left|x^{k+1} - x^k\right|_{\max} < \varepsilon$, the value of x^{k+1} is output as the final estimation and the iteration ends.

4.4. DT Modeling Process Based on Distribution Network State Estimation

The specific procedure of DT modeling based on the distribution network SE can be summarized as follows:

1. Collect measurement data from the physical distribution network and transfer the data to the DT database in the server via communication methods;
2. Store the historical data in the database and identify the bad data from the real-time in-coming measurements and correct the bad data employing the predicted measurements obtained from the training of the historical data;

3. Predict the output of PV power plants based on real-time-measured meteorological data and historical meteorological data;
4. Use the linear SE method to iteratively solve the distribution network states with consideration of the effect of PV node power constraints;
5. Obtain a DT model that reflects the real-time state of the distribution network.

According to the description above, the specific flowchart of DT modeling based on distribution network SE is depicted in Figure 7.

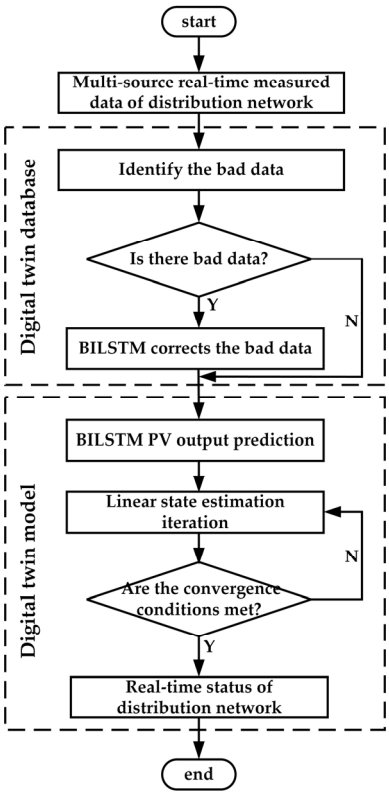


Figure 7. Flowchart of DT modeling based on distribution network state estimation.

5. Case Study

To simulate the real operating conditions of the distribution network and verify the accuracy of the DT model established in this paper, a hardware-in-the-loop (HIL) simulation method is used. The distribution network model is built in RTDS using RSCAD software to simulate the operation of the actual physical distribution network. The DT mathematical model is established on the server side. Real-time simulation in RTDS communicates with the server via an Ethernet switch, exchanging information using the user datagram protocol (UDP). Compared to the transmission control protocol (TCP), UDP has lower communication latency, enabling faster data transmission from RTDS and ensuring the timeliness of the DT model. Various measurement modules in RTDS measure the distribution network data, which is output from RTDS through communication boards, passed through the Ethernet switch, and then input into the DT database established on the server based on MySQL 8.0.20 (Oracle Corporation, Austin, TX, USA). The DT mathematical model extracts data from the database for calculations, establishes an accurate DT model, and stores the calculation results in the historical database. The DT test platform is shown in Figure 8.

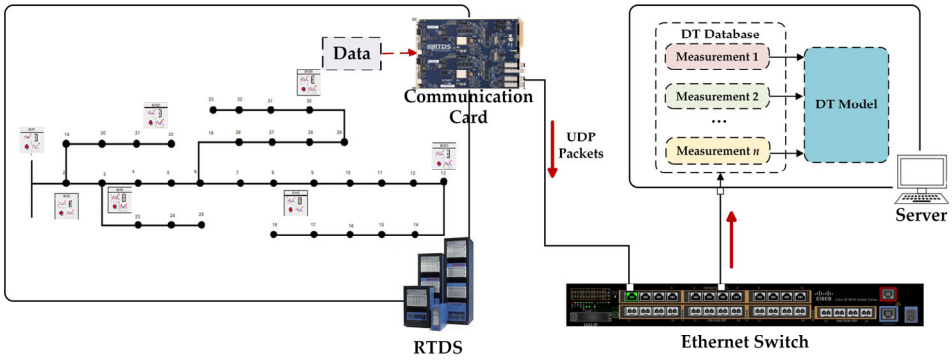


Figure 8. DT test platform.

The IEEE33 node 12.66KV distribution network model is built in RTDS, and the topology is shown in Figure 9. In order to simulate system fluctuations during real distribution network operation, the simulation accesses real-time fluctuating solar photovoltaic power generation equipment at Node 8, 16, 22 and 33, with an installed capacity of 1 MW, and adjusts the load data of other nodes (except generator Node 1) according to real load fluctuations. PMUs are installed at nodes 1, 3, 6, 11, 15, 21 and 29, and PMU measurements are highly accurate, with voltage amplitude and phase angle measurement errors of $\pm 0.05\%$ and ± 0.005 rad, respectively. All branch circuits are equipped with SCADA meters to measure branch circuit power data, with a measurement error of $\pm 1\%$. Additionally, the SCADA system measures power injection at nodes 4, 7, 10, 20, 21, 25, 27, and 31 to increase redundancy, with a measurement error of $\pm 0.5\%$.

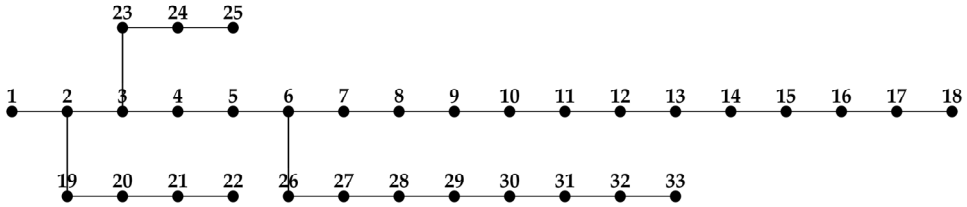


Figure 9. Topology of the 33-node distribution network.

The construction of the server-side DT model of the distribution network is based on the C++ code in the Visual Studio 2022 software (Microsoft, Redmond, WA, USA) platform, where part of the C++ code is automatically generated or modified based on the code written in the Matlab 2021b software. The RTDS simulation of the distribution network model is built on RSCAD version 5.014.1, with the RTDS equipment version being NovaCor 2.0. Testing was performed in follow environment:

The computer CPU was Core i5-9300H, the master frequency was 2.40 GHz, the RAM was 16 GB, and the GPU was NVIDIA GTX 1660Ti.

5.1. Data Evaluation Index

In this paper, we utilize three statistical measures, namely the mean absolute percentage error (MAPE), the root mean square error (RMSE) and the coefficient of determination R^2, in order to assess the deviation between the predicted value, the estimated value and the true value. MAPE, RMSE, and R^2 can be calculated as follows:

$$\mathrm{MAPE}(k) = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\hat{X}_{k,t} - X_{k,t}}{X_{k,t}}\right|, \tag{18}$$

$$\mathrm{RMSE}(k) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left|\hat{X}_{k,t} - X_{k,t}\right|}, \tag{19}$$

$$R^2(k) = 1 - \frac{\sum_{t=1}^{N} \left(X_{k,t} - \hat{X}_{k,t}\right)^2}{\sum_{t=1}^{N} \left(X_{k,t} - \overline{X}_{k,t}\right)^2}, \tag{20}$$

where $\hat{X}_{k,t}$ is the predicted or estimated value of node k at moment t. $X_{k,t}$ is the true value of node k at moment t. $\overline{X}_{k,t}$ is the average value of k node over time. N represents the number of consecutive time section. MAPE represents the relative error between the predicted and actual values. RMSE expresses the absolute difference between predicted and actual values. R^2 stands for the goodness of fit of the model.

5.2. Identification of Bad Data of DT Database

Measurement data from 2000 time sections in the DT database were selected as the test set, and 5%, 10% and 15% of bad data were subsequently incorporated into the aforementioned test set for the purpose of validation. The bad data of voltage-phase quantity and current-phase quantity exhibited a 20% to 30% increase or decrease in relation to the original value. Similarly, the bad data of active and reactive power exhibited a 40% to 50% increase or decrease in relation to the original value. The measurement time series length T was set to 10, and the $T-1$ time series measurement data prior to the current measurement moment was designated as the corrected measurement data. In order to facilitate a comparison between the method proposed in this paper and alternative approaches, the residual search method and DBSCAN clustering method were selected for analysis. The results of this comparison are presented in Table 1.

Table 1. Comparison of the effectiveness of different bad data identification methods.

Bad Data Ratio/%	Residual Search/%		DBSCAN/%		Proposed/%	
	Missing Detection Rate	False Detection Rate	Missing Detection Rate	False Detection Rate	Missing Detection Rate	False Detection Rate
5	8.6	11.5	3.9	5.1	0.7	1.8
10	16.7	19.2	6.5	12.8	1.6	2.1
15	22.9	27.6	13.7	19.6	2.2	2.9

As can be observed in Table 1, the traditional residual search method and clustering method fail to account for the temporal correlation of fluctuating data. Consequently, as the proportion of bad data increases, the missing detection rate and false detection rate of these two methods escalate significantly, whereas the method proposed in this paper maintains a high degree of discrimination in all cases, with a missing detection rate and false detection rate that are consistently lower than those of the residual search method and clustering method. This enables the accurate identification of bad data.

5.3. BILSTM Measurement Data Prediction Accuracy

SCADA measurements account for the largest proportion of measurement data, and the proportion of bad data is also higher. Consequently, this section mainly focuses on the verification of the prediction accuracy of branch power based on SCADA measurements. In this section, the BP neural network [26], LSTM [27], and BILSTM methods are employed for comparison, respectively. Measurement data from 2500 consecutive time sections in the DT database were selected for experimentation, with 80% of these utilized as the training set and 20% as the test set. The predicted values of active and reactive power for branches 3–4 in the 500 time sections of the test set were selected for comparison. Figures 10 and 11 demonstrate that the predicted values of BILSTM have the best fit to the true values.

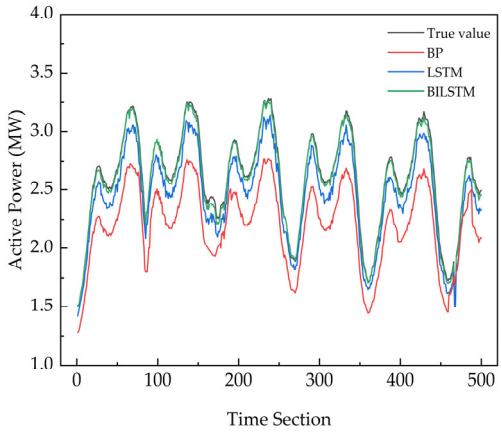


Figure 10. Comparison of branch circuit active power prediction results.

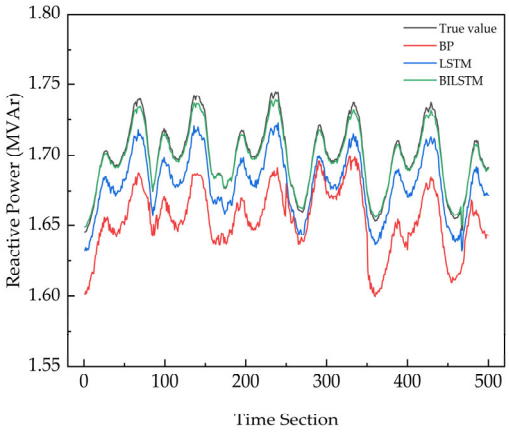


Figure 11. Comparison of branch circuit reactive power prediction results.

In Table 2, the predictive accuracy of each method is evaluated by calculating the mean of the MAPE, RMSE and R^2 values for all predicted branch power measurements. As can be noted from Table 2, BP neural networks have lower accuracy for time-series problems. In contrast, LSTM networks demonstrate greater efficacy than BP networks for time-series or natural language data. BILSTM networks combine forward and reverse LSTM structures to more comprehensively capture and understand the contexts and dependencies in the time series data, and reveal a higher prediction accuracy for bad data correction.

Table 2. Comparison of the predictive performance of distinct algorithms.

Algorithm	MAPE/%	RMSE/%	R^2/%
BP	11.84	13.68	81.22
LSTM	6.44	7.65	90.26
BILSTM	2.04	3.92	97.42

5.4. DT Model Validation Based on State Estimation

5.4.1. Accuracy of PV Output Prediction

Using the method proposed in this paper to predict the output of the PV power station, the predicted values of the 22-node PV power station at 300 consecutive time intervals were selected for comparison. Compare the ANN, BP, and BILSTM methods proposed in this paper for predicting photovoltaic output. Real-time meteorological data and historical meteorological data were input into trained ANN, BP, and BILSTM models, and the photovoltaic output prediction results are shown in the following figure, Figure 12.

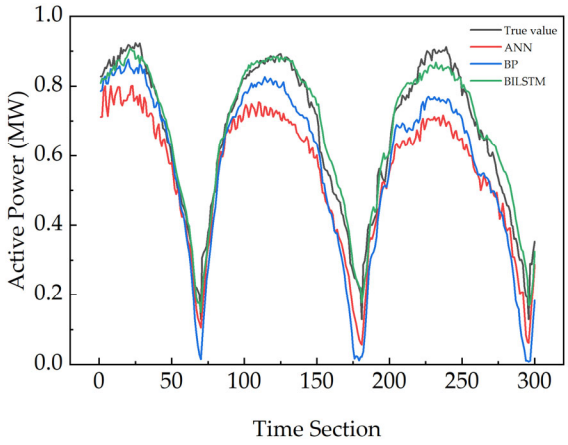


Figure 12. Comparison of predicted output curves of PV power plant at node 22.

From the figure, it can be seen that the proposed BILSTM method fits the photovoltaic actual output better than the ANN and BP methods. The calculated MAPE for the BILSTM predicted data is 16.95%, and the RMSE is 4.1%. It can be concluded that the BILSTM-based photovoltaic output prediction proposed in this paper exhibits good forecasting performance, which can serve as pseudo-measurements to enhance measurement redundancy in state estimation.

5.4.2. Analysis of DT Model Accuracy

This section introduces a comparative analysis of the WLS SE method and the WLAV SE method commonly used in distribution networks for constructing DT models, with the DT model proposed in this paper. A measurement dataset comprising 500 consecutive measurement time sections was constructed, with 10% of the bad data added. This process was conducted to simulate the presence of bad data in a real-world measurement scenario. Data for PV nodes were predicted using the method proposed in this paper. The actual value of the voltage phase quantity of a randomly selected time section was compared with the estimated value, and the results are presented in Figure 13. Node 10 was selected at random, and the state estimation results for the time sections 0–50 were obtained, as illustrated in Figure 14. Figures 13 and 14 demonstrate that the DT model presented in this paper exhibits enhanced accuracy and alignment with the real distribution network model. It is capable of tracking the changes in the physical distribution network model in a synchronized manner.

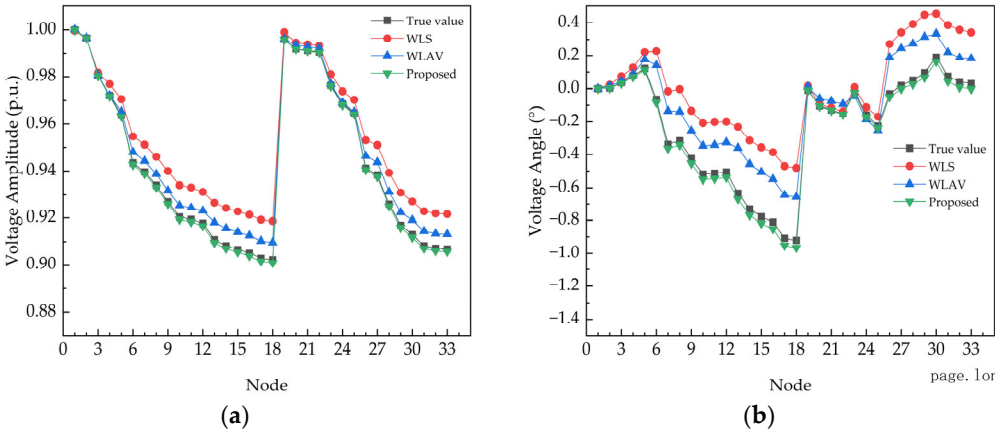


Figure 13. Comparison of the SE accuracy of a section at a given time. (**a**) Comparison of node voltage amplitude estimation accuracy. (**b**) Comparison of node voltage angle estimation accuracy.

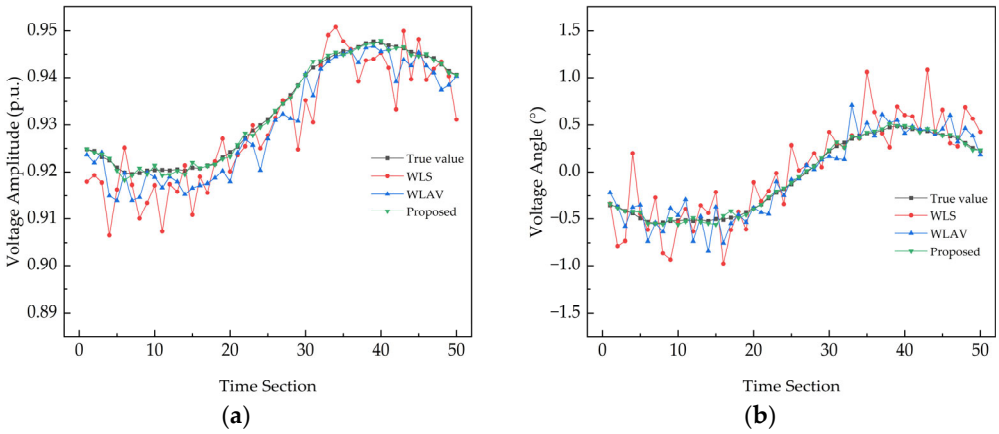


Figure 14. Comparison of node 10 SE accuracy for continuous time sections. (**a**) Comparison of node 10 voltage amplitude estimation accuracy. (**b**) Comparison of node 10 voltage angle estimation accuracy.

The total error in SE was calculated for all nodes over a period of 500 time sections, as demonstrated in Table 3 Following the comparison in Table 3, it can be observed that the traditional WLS SE model is reliant on data redundancy to enhance the accuracy of SE. However, it lacks the capacity to identify and rectify bad data, resulting in a considerable error. In contrast, the WLAV SE model is capable of automatically reducing the weight of bad measurement data, exhibiting robust performance. Nevertheless, it is unable to address the issue of bad data, leading to a lack of redundancy in the measurement. The method proposed in this paper demonstrates enhanced robustness against bad data and the capacity to rectify such data, thereby satisfying the SE redundancy requirement, which further improves the accuracy of the SE.

Table 3. Performance comparison of different SE algorithms.

State Estimation Model	Voltage Amplitude/%		Voltage Phase Angle/%	
	MAPE	RMSE	MAPE	RMSE
WLS	1.206	1.352	1.872	2.339
WLAV	0.215	0.341	0.785	1.058
Proposed	0.086	0.080	0.104	0.121

5.4.3. Analysis of DT Model Efficiency

In order to fulfil the real-time requirements of the DT model of the distribution network, it is necessary to ensure that the SE process is capable of providing the real-time distribution network state in an efficient manner. This section presents an estimation of the distribution networks and a comparison of the SE speeds of different SE models for a specific time period within each distribution network calculation. Table 4 displays the operational lengths of the various SE models.

Table 4. Comparison of computational efficiency among different state estimation models.

Distribution Network Model	Time/s		
	WLS	WLAV	Proposed
IEEE33	0.0182	0.0420	0.0014
IEEEE57	0.0516	0.1506	0.0025
IEEEE118	0.2385	0.4728	0.0102

As found in Table 4, the WLS method necessitates an increased number of iterations when confronted with complex measurement scenarios. Each iteration necessitates the recalculation of the Jacobi matrix, which is slower to converge and less computationally

efficient. In contrast, the WLAV employs an absolute value loss function, which necessitates the computation of the absolute value of each residual in order to minimize the error and the optimization problem is more complex. This presents a challenge in meeting real-time demands, despite the method's robust performance. This paper adopts an efficient method for identifying and correcting bad data, which is then used to preprocess measurement data. This process makes the SE more robust to differences. Furthermore, the linearization method reduces the Jacobi matrix to a constant matrix, which speeds up the iteration time while reducing the computational memory and greatly accelerating the computation speed. This approach is suitable for real-time demands of the DT model.

6. Conclusions

This paper addresses the challenge of developing an accurate DT model for distribution networks. It begins by adopting a method that considers the temporal correlation of measurement data to identify and exclude bad data. This is followed by the use of a BILSTM neural network training method to address the issue of missing measurement data, thereby ensuring the observability of the distribution network. Secondly, meteorological data and the BILSTM method are used to predict the real-time output of PV power plants. Subsequently, a linear SE algorithm is employed to model the DT of the distribution network in a rapid and efficacious manner. Finally, the constructed DT model is validated through the real-time synchronous operation of the RTDS and server models. Following verification of the simulation, the method for identifying bad data presented in this paper has a low missing and false detection rate. Moreover, the BILSTM neural network exhibits a higher degree of prediction accuracy than other neural networks in the prediction of measurement data. The linear SE method with PV integration proposed in this paper ensures the redundancy of SE measurements, improves the accuracy and iteration speed of the SE, and provides a guarantee for the accuracy and real-time performance of the DT model. The DT modeling method proposed in this paper provides new approaches and insights for the construction of smart grids. Based on efficient and precise real-time monitoring and state awareness, it facilitates the efficient operation and intelligent management of smart grids, thereby driving further optimization and enhancement of grid systems. However, the digital twin model established in this paper cannot track changes in the distribution network topology in real time, and the state estimation results do not vary with changes in the topology. In the future, we will further investigate digital twin modeling that can track changes in the distribution network topology. Based on the established DT model, we aim to achieve functions such as state prediction, fault location, and coordinated control, and explore its applicability and performance optimization in different grid scenarios.

Author Contributions: Conceptualization, H.Z. and R.M.; methodology, H.Z. and L.H.; software, H.Z. and R.M.; validation, H.Z. and X.C.; formal analysis, H.Z.; investigation, L.J.; resources, X.G.; data curation, L.H.; writing—original draft preparation, H.Z. and X.C.; writing—review and editing, R.M. and X.G.; visualization, H.Z.; supervision, L.H.; project administration, L.J.; funding acquisition, R.M and X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by State Grid Shanxi Electric Power Company Science and Technology Project Research (52053023000V).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wu, X.Q.; Xia, D. Review on intelligent planning and decision-making technology for the new active distribution network. *Integr. Intell. Energy* **2023**, *5*, 1–7.
2. Javaid, M.; Haleem, A.; Suman, R. Digital twin applications toward Industry 4.0: A review. *Cogn. Robot.* **2023**, *3*, 71–92. [CrossRef]
3. Shen, C.; Cao, Q.; Jia, M.; Chen, Y.; Huang, S. Concepts, characteristics, and prospects of application of digital twin in power system. *Proc. CSEE* **2022**, *42*, 487–499.

4. Zhou, M.; Feng, D.; Yan, J.; Zhou, X. A software platform for second-order responsiveness power grid online analysis. *Power Syst. Technol.* **2020**, *44*, 3474–3480.
5. Li, M.; Nie, M.; He, J.; Chen, K.; Wang, X.; Xu, Y. Pilot protection of flexible DC grid based on digital twin. *Proc. CSEE* **2022**, *42*, 1773–1782.
6. Gao, Y.; He, X.; Ai, Q. Multi agent coordinated optimal control strategy for smart microgrid based on digital twin drive. *Power Syst. Technol.* **2021**, *45*, 2483–2491.
7. Chen, C.; Liu, M.; Li, M.; Wang, Y.; Wang, C.; Yan, J. Digital twin modeling and operation optimization of the steam turbine system of thermal power plants. *Energy* **2024**, *290*, 129969. [CrossRef]
8. Khan, M.M.S.; Giraldo, J.; Parvania, M. Real-time cyber attack localization in distribution systems using digital twin reference model. *IEEE Trans. Power Deliv.* **2023**, *38*, 3238–3249. [CrossRef]
9. Han, J.; Hong, Q.; Syed, M.H.; Khan, M.A.U.; Yang, G.; Burt, G.; Booth, C. Cloud-edge hosted digital twins for coordinated control of distributed energy resources. *IEEE Trans. Cloud Comput.* **2022**, *11*, 1242–1256. [CrossRef]
10. Dobakhshari, A.S.; Terzija, V.; Azizi, S. Normalized deleted residual test for identifying interacting bad data in power system state estimation. *IEEE Trans. Power Syst.* **2022**, *37*, 4006–4016. [CrossRef]
11. Gu, Y.; Yu, Z.; Diao, R.; Shi, D. Doubly-Fed Deep Learning Method for Bad Data Identification in Linear State Estimation. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1140–1150. [CrossRef]
12. Lu, D.; Ma, L. Mixed Bad Data Diagnosis and Parameter Identification Based on Augmented State Estimation. *Electr. Power Eng. Technol.* **2019**, *38*, 99–104.
13. Jiao, Z.; Wu, R. A new method to improve fault location accuracy in transmission line based on fuzzy multi-sensor data fusion. *IEEE Trans. Smart Grid* **2018**, *10*, 4211–4220. [CrossRef]
14. Zheng, W.; Wu, W.; Shi, X.; Zhang, B.; Yang, J. A robust bilinear three-phase state estimation method for power systems. In Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, USA, 17–21 July 2016; IEEE: New York, NY, USA, 2016; pp. 1–4.
15. Chen, T.; Ren, H.; Li, P.; Amaratunga, G.A.J. A robust dynamic state estimation method for power systems using exponential absolute value-based estimator. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–10. [CrossRef]
16. Lin, C.; Wu, W.; Guo, Y. Decentralized robust state estimation of active distribution grids incorporating microgrids based on PMU measurements. *IEEE Trans. Smart Grid* **2019**, *11*, 810–820. [CrossRef]
17. Yuan, Z.; Chen, J.; Yan, M.; Xu, X. Fast Estimation of Power Grid Operation State Based on Multi-source Data. In Proceedings of the 2022 IEEE 5th International Electrical and Energy Conference (CIEEC), Nangjing, China, 27–29 May 2022; IEEE: New York, NY, USA, 2022; pp. 3781–3786.
18. Kandenkavil, S.V.; Padmanbahan, N. Performance Analysis of Optimization Based Static Distribution State Estimation Techniques. In Proceedings of the 2020 International Conference on Power, Instrumentation, Control and Computing (PICC), Thrissur, India, 17–19 December 2020; IEEE: New York, NY, USA, 2020; pp. 1–6.
19. Song, W.; He, J.; Lin, J.; Ye, H.; Ling, Z.; Lu, C. Bias analysis of PMU-based state estimation and its linear Bayesian improvement. *IEEE Trans. Ind. Inform.* **2023**, *20*, 1607–1617. [CrossRef]
20. Bhusal, N.; Shukla, R.M.; Gautam, M.; Benidris, M.; Sengupta, S. Deep Ensemble Learning-Based Approach to Real-Time Power System State Estimation. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106806. [CrossRef]
21. Ngo, Q.-H.; Nguyen, B.L.H.; Vu, T.V.; Zhang, J.; Ngo, T. Physics-informed graphical neural network for power system state estimation. *Appl. Energy* **2024**, *358*, 122602. [CrossRef]
22. Azimian, B.; Moshtagh, S.; Pal, A.; Ma, S. Analytical verification of performance of deep neural network based time-synchronized distribution system state estimation. *J. Mod. Power Syst. Clean Energy* **2024**, *12*, 1126–1134.
23. Lorenz, E.; Hurka, J.; Heinemann, D.; Beyer, H.G. Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2009**, *2*, 2–10. [CrossRef]
24. Wang, Q.; Ji, S.; Hu, M.; Li, W.; Liu, F.; Zhu, L. Short-Term Photovoltaic Power Generation Combination Forecasting Method Based on Similar Day and Cross Entropy Theory. *Int. J. Photoenergy* **2018**, *2018*, 6973297. [CrossRef]
25. Yang, X.; Li, Q.; Yuan, X.; Wei, Z.; Sun, G. Active distribution system state estimation considering the characteristics of DGs. In Proceedings of the 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), Changsha, China, 26–29 November 2015; IEEE: New York, NY, USA, 2015; pp. 2700–2705.
26. Wang, W.; Liu, W.; Chen, H. Information Granules-Based BP Neural Network for Long-Term Prediction of Time Series. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 2975–2987. [CrossRef]
27. Tan, M.; Yuan, S.; Li, S.; Su, Y.; Li, H.; He, F. Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning. *IEEE Trans. Power Syst.* **2020**, *35*, 2937–2948. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.