

Data-driven Digital Twin Construction and Terminal Collaborative Control Method of Distribution Network

Feng Wang^{1,*}, Kunkun Wang^{1,*}, Pengbo Zhao², Xikui Li²

¹State Grid Shangqiu Power Supply Company, Shangqiu, Henan, China

²Zhengzhou Jingyun Technology Co., Ltd., Zhengzhou, Henan, China

*Corresponding Author

Abstract: This paper proposes a data-driven new distribution network digital twin construction and terminal collaborative control method to solve the problem of new distribution network operation control in the context of high proportion of renewable energy access. The research constructs a four-layer digital twin architecture including the physical perception layer, data fusion layer, model construction layer and application service layer, and establishes a technical framework for deep integration of physical systems and virtual space. On this basis, a hybrid modeling method that combines mechanism models and data-driven models is proposed. Through multi-source data fusion and dynamic update mechanisms, high-fidelity digital representation of distribution network operating characteristics is achieved. We further designed a collaborative control mechanism based on digital twin simulation deduction, using a virtual sandbox for control strategy pre-verification and safety verification, achieving closed-loop collaboration of centralized optimization and distributed execution. Research shows that this method can improve the distribution network status sensing accuracy, control decision-making foresight and terminal collaboration efficiency, provide theoretical basis and technical support for enhancing the power grid's ability to absorb distributed energy, operational safety and economy, and has important application value in promoting the construction of new power systems.

Keywords: New Distribution Network; Digital Twin; Data-Driven; Collaborative Control; Hybrid Modeling

1. Introduction

With the continuous expansion of the power system scale and the continuous development of the power network, the traditional power system management and optimization methods face many challenges. As the end link of the power system, the distribution network is undergoing a structural transformation from a traditional one-way power supply network to an active interactive complex system. While this shift promotes clean energy consumption and improves system flexibility, it also raises many technical challenges. The randomness and fluctuation of distributed power generation significantly increase the complexity of system power balance and voltage control [1]. At the same time, the deployment of massive intelligent terminals and sensing equipment not only generates multi-source heterogeneous data, but also puts forward new requirements for the system's real-time perception capabilities, collaborative decision-making, and rapid response efficiency [2].

By constructing a virtual image that maps and dynamically interacts with physical entities, digital twins can realize real-time mapping, historical retrospective, future deduction, and strategy rehearsal of physical system operations in the digital space, thereby supporting more accurate and forward-looking decision-making [3]. In the field of power systems, the research and application of digital twins is extending from the equipment level to the system level. Existing research mainly focuses on the technical architecture design and key enabling methods. In terms of architecture, scholars generally follow the design concept of hierarchical decoupling. For example, the five-dimensional model proposed by Onaji et al. provides a theoretical framework for the construction of general digital twins, which has a wide influence in the industrial field

[4]. Pacheco et al. designed a four-layer architecture that includes physical entities, data support, model construction, and application interaction based on the characteristics of distribution networks, emphasizing the core role of data drive [5]. In terms of model construction methods, the research shows a trend of developing from a single physical model to multi-model fusion and data-driven. El et al. discussed the hybrid modeling technology path that combines the physical mechanism of the power grid with operational data and artificial intelligence models [6]. In view of the strong uncertainty of new energy power generation, Huang et al. explored data-driven methods using deep learning algorithms such as long short-term memory networks to improve the accuracy of digital representation of photovoltaic output, which effectively compensated for the shortcomings of mechanistic models [7]. At the specific application level, the existing work mostly focuses on analysis links such as state perception and fault diagnosis. For example, Shen et al. used digital twin technology to achieve real-time, high-precision estimation of voltage at key nodes in the distribution network [8]. Yang et al. built a fault simulation and self-healing framework based on digital twins, which improved the O&M response efficiency [9]. However, through a systematic review of the existing literature, it can be found that there are still obvious shortcomings in the current research, and most studies either take the construction of digital twins themselves as the endpoint, or focus on the realization of a single analysis function, and the constructed virtual models fail to form a deep closed loop with control decision-making, which limits the core value of digital twins in the complete chain of "prediction-optimization-control" [10]. In addition, when dealing with the problem of massive distributed terminal control, existing strategies rely on centralized optimization and face computing and communication bottlenecks [5,9]. How to achieve effective collaboration is the core scientific problem that this paper focuses on. In view of the above research shortcomings, this paper proposes a data-driven new digital twin construction and terminal collaborative control method for distribution networks.

This paper designs a four-layer digital twin architecture comprising physical perception layer, data fusion layer, model construction layer and application service layer, and clarifies the functional positioning and data interaction logic of each level.

2. Digital Twin-Driven Distribution Network Collaborative Control Framework

2.1 Characteristics Analysis and Technical Requirements of New Distribution Networks

The new distribution network presents significant technical characteristics that distinguish it from traditional power grids. Its core characteristics are reflected in its highly distributed characteristics and two-way interaction. From the perspective of physical composition, the distribution network is connected to a large number of renewable energy generation units such as distributed photovoltaic and wind power, as well as controllable resources such as electric vehicle charging piles, energy storage systems, and flexible loads. These devices are distributed in terms of geographical location, access capacity, and operating characteristics, forming a typical active network structure. From the perspective of operation mode, the power flow has changed from the traditional one-way radial flow to a multi-directional and dynamic complex mode, and the voltage distribution and power balance face significant challenges. From the perspective of time scale, the random fluctuation of distributed power output and the rapid change of load demand require the control system to have a response ability of minutes or even seconds. From the perspective of information interaction, a large number of terminal devices such as smart meters, environmental sensors, and smart switches continue to generate multi-source heterogeneous monitoring data, which puts forward higher requirements for real-time information collection, transmission, processing, and decision-making.

In the face of the above characteristics, the collaborative control of new distribution networks needs to meet multi-dimensional technical requirements. First, it is necessary to establish comprehensive, accurate and real-time system status perception capabilities, and break through the limitations of traditional monitoring methods in data dimension, update frequency

and fusion accuracy. Second, it is necessary to build a refined dynamic model that can be compatible with physical mechanisms and data laws to accurately reflect the complex interaction behavior of the system under the access of new components. Third, it is necessary to design a collaborative control mechanism that can coordinate the overall optimization goal and local autonomous execution, solve the bottleneck of centralized control in computing and communication, and avoid the local optimization or operational conflict caused by fully distributed control. These requirements constitute the logical starting point for building a new generation of distribution network operation control system.

2.2 Overall Architecture Design of Digital Twin System

To meet the above technical requirements, this paper proposes a four-layer collaborative control architecture driven by digital twins (Figure 1), which aims to realize intelligent perception, accurate mapping, and collaborative regulation of the distribution network through the deep integration and closed-loop interaction between physical systems and virtual spaces. The architecture is logically connected from top to bottom and includes cross-level data and instruction flow in both directions.

The core levels and functions of the architecture are divided as follows:

The physical perception layer is the foundation of the architecture, which consists of all real physical equipment in the distribution network and its supporting sensing, measurement and execution units. This layer is responsible for real-time collection of electrical energy data (such as voltage, current, power), equipment status data (such as switching position, energy storage SOC), and environmental data (such as light and temperature), and receives control commands from the upper layer to drive the action of actuators such as circuit breakers, inverters, and controllable loads.

The data fusion layer plays the role of an information hub. It receives multi-source, heterogeneous, and high-frequency raw data streams from the physical perception layer, and eliminates outliers and noise through preprocessing operations such as data cleaning, format standardization, and time series alignment to ensure data quality and consistency. Subsequently, data correlation

analysis and feature extraction technology are used to fuse data from different devices and spatiotemporal dimensions to form a standardized data set that can comprehensively reflect the overall operating status of the system, providing high-quality data input for upper-level model construction and control decision-making.

The core task of the model building layer is to build and maintain a virtual image that is highly consistent with the physical distribution network. The layer contains two core model libraries: one is a mechanism model based on grid topology, line parameters, and physical laws, which is used to describe the basic connection relationship of the network and the physical processes of steady-state and transient states; and the other is a device characteristic model and behavior prediction model based on data-driven methods (such as machine learning algorithms) to characterize the complex nonlinear dynamic characteristics of distributed power sources, loads, and other components. The two are deeply integrated through hybrid modeling technology to form a high-fidelity virtual distribution network system that can evolve and update in real time.

The application service layer is the key to the realization of the value of digital twins, which is directly oriented to the specific business requirements of distribution network operation control. It provides a series of core service functions based on the virtual system provided by the model building layer, including but not limited to real-time monitoring and visualization of system status, advanced deduction of operation situation and safety warning, generation of collaborative control strategies based on global optimization algorithms, and virtual simulation verification and effect evaluation of control instructions. The optimization policies or control boundaries generated by this layer will descend to the physical perception layer through the data fusion layer to drive the execution of actual terminal devices, thus forming a closed loop of "perception-decision-execution-evaluation".

2.3 System Operation Process

The overall operation of the architecture follows a continuous iterative closed-loop process. The starting point of the process is the comprehensive collection of grid operation status and environmental information by the

physical perception layer, and the original data is synchronized into the model construction layer after standardized processing and fusion by the data fusion layer. The model construction layer uses this real-time data to calibrate parameters and update the status of the virtual model to ensure that the digital twin is synchronized with the physical entity. The application service layer performs analysis, deduction, and optimization calculations based on the high-fidelity virtual model of the current moment to generate specific collaborative control schemes. After simulation verification and security verification in the virtual space, the scheme is transformed into a specific control instruction set. The instruction set is sent to the corresponding execution terminal of the physical perception layer through the data fusion layer to change the operating state of the physical system. The new state change is again sensed and fed back to the digital space, starting the next round of iterative updates. This closed-loop process realizes the continuous interaction and co-evolution of the physical system and the digital space, making the control system have the potential for adaptation, self-learning, and self-optimization.

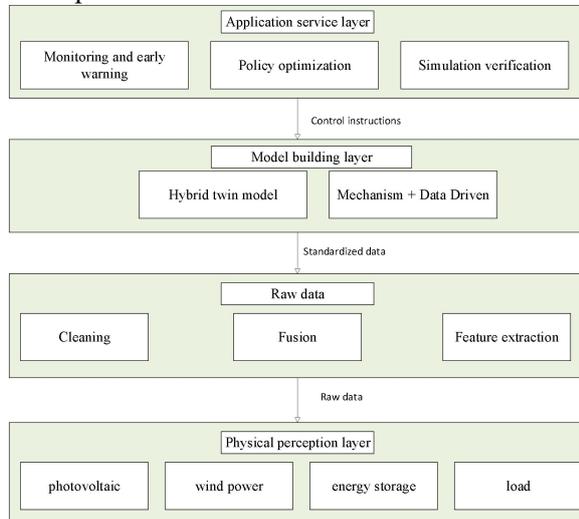


Figure 1. Four-Layer Architecture Diagram of a Digital Twin System

3. Data-Driven Digital Twin Modeling Method for Distribution Networks

3.1 Multi-Source Heterogeneous Data Collection and Fusion Processing

In the new distribution network, the data source mainly includes three dimensions: first, the power grid operation data, which comes from

smart meters, distribution automation terminals, synchronous phasor measurement devices, etc., covering electrical measurement information such as voltage, current, power, frequency, as well as status signals such as switch position and protection action; second, equipment status data, including the operating parameters, health status and alarm information of power electronic equipment such as photovoltaic inverters, energy storage converters, and charging piles; and third, environmental and external data, such as irradiance, wind speed, and temperature information provided by meteorological stations, network topology and equipment geographical location data provided by geographic information systems, and load prediction and adjustable resource information provided by demand-side management systems. These data have significant differences in spatiotemporal scale, sampling frequency, data format and signal-to-noise ratio, and must go through a systematic fusion preprocessing process. The preprocessing process first performs data cleaning, identifies and eliminates outliers and missing values caused by communication interruptions and equipment failures, and timestamps aligns the data in different sections. Subsequently, the data is normalized to convert the data from different measurement systems into a unified physical reference and numerical range. On this basis, the data correlation analysis technology is used to establish the correlation matrix between the data of different monitoring points according to the topological connection relationship and electrical coupling characteristics of the power grid, so as to identify data contradictions and improve the redundancy of state estimation. Finally, through the feature extraction method, the key index sets that can characterize the operating characteristics of the system are condensed from the massive time series data, forming a standardized and high-quality fusion data set for the construction of the upper model.

3.2 Mechanism and Data-Driven Hybrid Modeling Technology

The mechanism model part is mainly responsible for constructing the physical skeleton of the distribution network. This part is based on the known basic data such as grid topology, line impedance parameters, transformer ratio, etc., and uses Kirchhoff's law and other physical principles to establish a

mathematical model describing the basic electrical characteristics of the network, such as power balance, voltage distribution and fault current. This model has clear physical significance, which can ensure the theoretical correctness of the simulation results under steady-state and partial transient conditions, and provide an interpretable basis for the whole hybrid model.

The data-driven model part focuses on describing complex dynamic characteristics and uncertainties that are difficult to accurately describe in mechanism models. Aiming at the strong randomness and intermittency of distributed power outputs such as photovoltaic and wind turbines, deep learning algorithms such as long short-term memory network and temporal convolutional network are used to construct a high-order nonlinear mapping relationship between their output, meteorological factors, and historical output sequences. For flexible resources such as electric vehicle charging load and temperature control load, cluster analysis and reinforcement learning are used to model their aggregate response characteristics and behavior patterns. For slow time-varying parameters such as equipment aging and contact resistance change, online parameter identification algorithms are used for dynamic tracking.

Hybrid modeling is realized through two methods: structural fusion and result fusion. Structural fusion refers to embedding data-driven submodules in specific links or equations of the mechanism model, such as using neural networks to replace traditional empirical formulas to describe the extradynamic characteristics of the inverter. Result fusion adopts a parallel architecture to integrate the output of the mechanism model with the output of the data-driven model through weighting, voting, or meta-learning to comprehensively utilize the advantages of the two. The model continuously connects to the real-time fusion data stream, uses online learning technology to dynamically update the parameters of the data-driven part, and checks the key parameters of the mechanism model according to the measured values to ensure that the virtual model can adaptively track the actual evolution of the physical system.

3.3 Model Dynamic Update and Accuracy Maintenance Mechanism

The model accuracy real-time monitoring module continuously compares the twin simulation output with the actual measurement data to calculate the error indicators of key state variables (such as node voltage and branch power). When the error exceeds the preset threshold or detects a significant switch in the system operating mode, the model evaluation process is triggered. The evaluation process not only analyzes the size of the error but also diagnoses the source of the error, distinguishing whether it is due to model parameter drift, unmodeled dynamics, or data quality issues.

Based on the evaluation results, a differentiated update strategy is executed. For errors caused by slow changes in equipment parameters, the parameter automatic calibration algorithm is started to re-estimate the line resistance, transformer leakage resistance, and other parameters based on the latest measurement data. When new equipment is put into operation or topology changes, the preset component template in the model library is called to automatically expand or reconstruct the network model. For the performance degradation of the data-driven model due to changes in source and load characteristics, the incremental learning process is started to fine-tune the existing model using new data to avoid catastrophic forgetting. All update operations are verified in an independent simulation sandbox environment, and can be synchronized to the online twin model after confirming that it is correct. Through this series of mechanisms, the digital twin model can achieve self-evolution and continuous optimization, providing a reliable analysis and prediction basis for collaborative control.

4. Terminal Collaborative Control Mechanism Based on Digital Twin

4.1 Collaborative Control Architecture and Operation Objectives

The collaborative control mechanism based on digital twins aims to build a closed-loop regulation system of "centralized decision-making-distributed execution-real-time feedback". The system takes the digital twin platform as the core hub and connects the upper-level centralized optimization decision-making with the lower-level distributed terminal execution. At the top level, the digital twin provides global situational awareness,

forward-looking deduction and optimization computing services based on panoramic data and high-fidelity models, and generates system-level coordinated control strategies or safe operation boundaries. At the bottom layer, various intelligent terminals perform rapid and autonomous adjustment actions based on receiving superior instructions or boundary constraints, combined with local real-time information. Information exchange between the two layers through standardized data interfaces and communication protocols forms a two-way closed loop.

The collaborative control mechanism pursues multi-objective comprehensive optimization. The primary goal is to ensure the safe and stable operation of the power grid, maintain key electrical parameters within the safe allowable range, especially to suppress the voltage overrun caused by distributed power fluctuations. The second goal is to improve the system operation economy, reduce network losses by optimizing power flow distribution, and guide controllable resources to participate in demand response to shave peaks and fill valleys. The third goal is to enhance the consumption capacity of renewable energy and stabilize power fluctuations through the collaborative interaction between source, grid, load and storage. The fourth goal is to improve the reliability of power supply, and in the event of a fault, it can be supported by rapid simulation support based on the digital twin to achieve fault isolation and rapid restoration of power supply in non-fault areas. These goals are coordinated and compromised through the multi-objective optimization algorithm in the digital twin platform.

4.2 Centralized Optimization Decision-Making Based on Digital Twins

In the centralized optimization decision-making process, the digital twin platform acts as a repeatable and risk-free "simulation sandbox". The decision-making process begins with accurate perception of the current system operating state and prediction of future ultra-short-term trends. Based on the hybrid model updated to the latest state, the platform can perform multi-time continuous power flow calculation and safety verification on the system operation scenario in the next 5-15 minutes in minutes or even seconds.

When possible risks such as voltage out-limit

and branch overload are predicted, the optimization decision-making module is activated. The module constructs a mathematical optimization model with constraints, and the decision variables include the active/reactive output set points of each controllable distributed power source, the charging and discharging power of the energy storage system, and the switching or power adjustment amount of the flexible load. The objective function is dynamically adjusted according to the operating mode, which may be the lowest network loss, the smallest voltage deviation, or the lowest adjustment cost. The constraints include the power flow equation, the upper and lower limits of equipment operation, and the operation rate limit.

4.3 Autonomous and Collaborative Execution of Distributed Terminals

After centralized optimization and simulation verification, the control instructions are issued to each region or each type of terminal aggregation node. The command form can be a specific set value, or a more flexible constraint boundary or objective function parameter. On this basis, the distributed terminal carries out fast and autonomous collaborative execution based on local real-time measurement information.

For voltage reactive power control, the consistency algorithm can be used to realize distributed collaboration. Each node installed with reactive power compensation equipment (such as photovoltaic inverter and stationary reactive power generator) acts as an agent, which independently adjusts the reactive power output according to the local voltage measurement value and the information from neighboring nodes, according to the preset consistency protocol, so that the voltage coordination of all nodes in the area is close to the target value or target range given by the optimization instruction. This process does not require continuous intervention from a centralized controller, and communication only occurs between neighboring nodes, which has good scalability and robustness.

For active power balance, the sag control based on local measurement combined with target correction can be adopted. The terminals of energy storage systems and controllable loads can independently adjust the output according to the preset sag coefficient according to the

local frequency or voltage change to quickly stabilize the power fluctuation. At the same time, the digital twin platform periodically issues correction instructions based on global optimization calculations to fine-tune the reference points of the sag characteristics of each terminal to eliminate the steady-state deviation that may exist in pure distributed control and ensure the overall economic optimization.

The execution status and effect data of all terminals are collected in real time and fed back to the digital twin platform. The platform evaluates the accuracy of the hybrid model, the applicability of the optimization algorithm, and the effectiveness of the control instructions by comparing the expected control effect with the actual feedback data. If the deviation persists, the model parameter calibration or control strategy re-optimization process is triggered, so as to realize the complete closed-loop and iterative optimization of "perception-decision-execution-evaluation", and ensure that the collaborative control mechanism can adapt to the dynamic changes of the power grid.

5. Conclusion

In view of the operation control challenges faced by new distribution networks, this paper systematically proposes a set of data-driven digital twin construction and terminal collaborative control methods for new distribution networks. This study provides a feasible technical path to improve the carrying capacity, operation safety and economy of distribution networks for a high proportion of distributed energy. However, this study mainly focuses on the discussion of the method framework and mechanism, and still faces practical challenges such as difficulty in ensuring multi-source data quality, limited edge computing resources, and insufficient standardization of heterogeneous terminal communication protocols at the engineering practice level. In the future, it is necessary to focus on the deployment of lightweight models, the optimization of edge-cloud collaborative computing architecture, communication security and standard formulation, and explore the integrated application value of this method in a wider range of scenarios such as distribution network planning, fault self-healing, and market transactions.

References

- [1] Habbak, H., Mahmoud, M., Metwally, K., et al. (2023). Load forecasting techniques and their applications in smart grids. *Energies*, 16(3), 1480.
- [2] Huang, J., Chen, C., Sun, C., et al. (2023). An integrated risk assessment model for the multi-perspective vulnerability of distribution networks under multi-source heterogeneous data distributions. *International Journal of Electrical Power & Energy Systems*, 153, 109397.
- [3] Jia, W., Wang, W., Zhang, Z. (2022). From simple digital twin to complex digital twin Part I: A novel modeling method for multi-scale and multi-scenario digital twin. *Advanced Engineering Informatics*, 53, 101706.
- [4] Onaji, I., Tiwari, D., Soulatiantork, P., et al. (2022). Digital twin in manufacturing: conceptual framework and case studies. *International journal of computer integrated manufacturing*, 35(8), 831-858.
- [5] Pacheco-Cherrez, D. S., Guillen, D., Mayo-Maldonado, J. C., et al. (2024). Data-driven optimal voltage performance index tracking in active distribution networks. *IEEE Transactions on Smart Grid*, 15(5), 4804-4818.
- [6] El Rhatri, A., Bouihi, B., Mestari, M. (2024). Challenges and limitations of artificial intelligence implementation in modern power grid. *Procedia Computer Science*, 236, 83-92.
- [7] Huang, C., Yang, M. (2023). Memory long and short term time series network for ultra-short-term photovoltaic power forecasting. *Energy*, 279, 127961.
- [8] Shen, J., Hu, L., Yang, Y., et al. (2025). Real-time update algorithms for digital twin models of distribution network equipment under internet of things and optical imaging technology. *Scientific Reports*, 15(1), 5910.
- [9] Yang, C., Cai, B., Wu, Q., et al. (2023). Digital twin-driven fault diagnosis method for composite faults by combining virtual and real data. *Journal of Industrial Information Integration*, 33, 100469.
- [10] Wu, Q., Wang, W., Wang, Q., et al. (2022). Digital twin approach for degradation parameters identification of a single-phase DC-AC inverter. In *2022 IEEE Applied Power Electronics Conference and Exposition (APEC)* (pp. 1725-1730). IEEE.