Preprints.org

Article

Not peer-reviewed version

# Survey of Data-Driven Methods for Power Systems

Rahul G *

Posted Date: 19 December 2024

doi: 10.20944/preprints202412.1697.v1

Keywords: data-driven; power systems; survey

*Article*

# Survey of Data-Driven Methods for Power Systems

**Rahul G.**

Independent Researcher; hronit245@protonmail.com

**Abstract:** The integration of renewable energy sources and distributed energy resources is transforming the electric power system, creating both opportunities for improved sustainability and challenges related to system complexity and uncertainty. While traditional model-based approaches have historically been used for power system analysis and control, the increasing availability of data from sensors, SCADA systems, and smart meters has spurred the development and application of data-driven methods. This survey paper provides a comprehensive overview of state-of-the-art data-driven techniques applied to power systems, addressing key areas such as predictive analytics for forecasting, state estimation for real-time monitoring, fault detection and diagnosis for enhanced reliability, control and optimization for improved efficiency, and cybersecurity for enhanced grid resilience. These data-driven approaches offer advantages over traditional methods, including data-driven decision making, enhanced system understanding, improved system performance, and adaptation to changing operating conditions. By leveraging data analysis, these methods enable more accurate insights into system behavior, uncover hidden patterns, and optimize operations for improved reliability, efficiency, and security of the modern power grid.

**Keywords:** data-driven; power systems; survey

---

## 1. Introduction

The electric power system, a critical infrastructure, is undergoing significant transformation due to the integration of renewable energy sources, distributed energy resources (DERs), and advanced technologies. This evolution presents both opportunities and challenges. On one hand, it promises a more sustainable and efficient power grid. On the other hand, it introduces complexities in terms of system dynamics, uncertainty, and security.

Traditional model-based approaches have been the mainstay for power system analysis and control. However, the increasing complexity and uncertainty associated with modern power systems have led to a growing interest in data-driven methods. These methods leverage large amounts of data collected from various sources, such as sensors, SCADA systems, and smart meters, to extract valuable insights and make informed decisions.

Data-driven methods offer a promising approach to address the challenges of the modern power system. By analyzing historical data and real-time measurements, these methods can enable more accurate forecasting, improved state estimation, efficient control strategies, and enhanced cybersecurity. This survey paper provides an overview of the state-of-the-art data-driven methods applied to power systems, including predictive analytics, state estimation, fault detection and diagnosis, control and optimization, and cybersecurity. These data-driven methods offer several advantages over traditional model-based approaches, including:

- **Data-Driven Decision Making:** By analyzing historical data and real-time measurements, data-driven methods can provide more accurate and timely insights into system behavior.
- **Enhanced System Understanding:** Data-driven methods can help uncover hidden patterns and correlations within the power system, leading to a deeper understanding of its dynamics.
- **Improved System Performance:** By optimizing system operations based on data-driven insights, it is possible to improve system efficiency, reliability, and security.
- **Adaptation to Changing Conditions:** Data-driven methods can adapt to changing conditions, such as variations in load demand, renewable energy generation, and system disturbances.

This survey paper provides a comprehensive overview of the state-of-the-art data-driven methods applied to power systems, including:

 doi:10.20944/preprints202412.1697.v1



1. **Predictive Analytics:** Techniques for forecasting future system behavior, such as load forecasting and renewable energy forecasting.
2. **State Estimation:** Methods for estimating the real-time state of the power system, including voltage magnitudes, angles, and power flows.
3. **Fault Detection and Diagnosis:** Techniques for identifying and locating faults in the power system, enabling rapid response and restoration.
4. **Control and Optimization:** Methods for optimizing power system operation, including control strategies and optimal power flow.
5. **Cybersecurity:** Techniques for detecting and mitigating cyber-attacks on power systems.

By understanding the principles and applications of these data-driven methods, power system engineers and operators can make informed decisions to ensure the reliable, efficient, and secure operation of the power grid.

## 2. Data-Driven Methods for Power Systems

Data-driven methods have the potential to revolutionize power system operations and control. Some of the key applications include:

### *2.1. Predictive Analytics*

#### 2.1.1. Load Forecasting

Accurate load forecasting is essential for efficient power system operation and planning. Data-driven methods, such as machine learning and time series analysis, can predict future load demand based on historical data and relevant factors like weather conditions, economic indicators, and social events. In [1], the authors used fuzzy logic to have the weather and temperature to forecast the future load on a short term basis, showing improved forecast based on the actual weather data. In [2], the authors developed a data-driven method to use advanced metering infrastructure (AMI) data to cluster consumers according to similar patterns, and perform load forecasting on the cluster level followed by aggregation at system level. This leads to improve forecasting accuracy. In [3], the importance of electrical vehicle (EV) load forecasting is highlighted. A neural network architecture is used to predict short term EV load using behavioral patterns in charging in North China, and an improved accuracy of forecasting is achieved compared to traditional forecasting methods.

#### 2.1.2. Renewable Energy Forecasting

Forecasting the intermittent nature of renewable energy sources, such as solar and wind power, is paramount for ensuring grid reliability and stability. Accurate predictions of their variability and uncertainty are essential for effective grid integration and operation. To address this challenge, data-driven methods have emerged as powerful tools for improving renewable energy generation forecasts. In [4], the authors propose a novel approach utilizing gradient boosting tree algorithms to leverage weather prediction data. By incorporating detailed meteorological information, this method aims to enhance the accuracy of solar and wind energy generation forecasts. Compared to traditional forecasting techniques, this approach has demonstrated a significant improvement of over 12%. A probabilistic forecasting model is presented in [5], which leverages historical hourly data to construct an annual eight-segment probabilistic model for wind and solar energy. This model employs a probabilistic approach to estimate hourly wind speeds and solar irradiance throughout the year, providing a more comprehensive understanding of the potential range of outcomes. In [6], three distinct methods of renewable probabilistic forecasting are explored as trading agents within a binary prediction market. By analyzing the equilibrium price in this market, the aggregated probability of renewable output can be elicited. Subsequently, a regression analysis is employed to extract the full cumulative distribution function of possible renewable output, offering a detailed probabilistic representation of the forecast.

 doi:10.20944/preprints202412.1697.v1



### 2.2. State Estimation

Data-driven methods can estimate the real-time state of the power system, including voltage magnitudes and angles, and real and reactive power flows. This information is essential for monitoring and control purposes. Data-driven methods can improve the robustness of state estimation algorithms by considering uncertainties and noise in the measurements. In [7], the authors proposed a robust and scalable Gaussian process regression (RS-GPR)-enabled distributed dynamic state estimation method. It allows accurate pseudo-measurement inference as well as robust state estimation with only a limited number of measurements, and perform superior compared to machine learning assisted state estimation methods. [8] uses an artificial neural network (ANN) to use sparse synchrophasor measurements (PMU data) to improve distribution system state estimation for unbalanced systems. The method is executed significantly faster than traditional algorithms using a parallelized architecture, which facilitates near real-time monitoring. In [9], the authors proposed a high temporal-spatial resolution state estimation (SE) method, leveraging graph convolutional network (GCN) and dense connectivity structure to estimate states of whole system at PMU reporting rate. This enabled the authors to perform online dynamic security assessment for transient stability analysis of the power grid.

In conclusion, data-driven methods have emerged as powerful tools for enhancing power system state estimation. By leveraging advanced techniques like Gaussian process regression, artificial neural networks, and graph convolutional networks, these methods offer significant advantages in terms of accuracy, robustness, and computational efficiency. These advancements enable more precise monitoring and control of the power grid, contributing to improved system reliability and security. As the power grid continues to evolve, the integration of data-driven methods will play a pivotal role in ensuring its efficient and resilient operation.

### 2.3. Control and Optimization

#### 2.3.1. Model-Free Control

Data-driven control methods, such as reinforcement learning and adaptive control, can optimize the control of power system components without relying on detailed system models. Traditionally, the controls were designed based off the nominal power system model. However, with the penetration of renewable energy and the constantly changing operating point, the nominal controllers might not perform optimally for fast changing operating points. Therefore, it's important to design controls which can adapt with the system in a data-driven manner. In [10], a data-driven method based on extreme learning machine (ELM)-based learning model is used to improve the frequency stability prediction based on a nominal model. This data-driven frequency prediction is used to create the load shedding plan. In [11], the authors proposed to design a linear wide-area damping control using synchrophasor measurements assuming availability of full system state. The work showed that how such an adaptive control method improves the damping performance of the system during off-nominal and stressed system conditions. In [12], a data-driven control strategy based on historical measurements is proposed to mitigate local power imbalances in control areas, using inverter-based resources in the areas.

#### 2.3.2. Optimal Power Flow

Data-driven methodologies are revolutionizing the approach to solving the complex optimal power flow (OPF) problem, promising significant improvements in grid efficiency, reliability, and economic operation by minimizing power losses, reducing operational costs, and ensuring adherence to stringent system constraints. For example, recognizing the inherent variability in renewable energy integration and load demand, researchers have developed a data-driven chance-constrained optimization framework for OPF [13]. This approach explicitly addresses uncertainties in available reserve capacity by incorporating probabilistic constraints learned from historical data and forecasts, ensuring a high probability of maintaining real-time balance between energy supply and demand, a crucial factor for grid stability with increasing penetration of intermittent renewable sources. Furthermore, the computational burden associated with OPF calculations, especially for large-scale distribution grids, has motivated the exploration of quantum computing. In this context, researchers have investigated the application of the Quantum Approximate Optimization Algorithm (QAOA) to accelerate power flow solutions [14]. A key advancement is the use of data-driven methods to determine the optimal parameters for QAOA, bypassing computationally expensive tuning processes and achieving performance comparable to meticulously tuned QAOA instances. This data-driven parameter selection significantly reduces the computational overhead associated with quantum OPF, making it a more practical solution for real-world applications. Finally, maintaining stable voltage levels across the grid is paramount for reliable power delivery. To address this, a data-driven optimization strategy for volt-var control has been proposed [15]. This method leverages real-time measurements from smart grid infrastructure to estimate volt-var sensitivities, which quantify the impact of reactive power adjustments on bus voltages. These dynamically estimated sensitivities are then used to design a more responsive and precise control strategy that effectively improves bus voltage profiles and enhances overall grid stability, adapting to constantly changing grid conditions. These diverse applications demonstrate the transformative potential of data-driven approaches in creating a more efficient, resilient, and sustainable future for power systems.

### *2.4. Cyber-Security*

#### 2.4.1. Intrusion Detection Systems (IDS)

Intrusion Detection Systems (IDSs) are a critical first line of defense in safeguarding power system network control from debilitating cyberattacks. These systems are essential for preventing unauthorized access to sensitive power system data and maintaining grid stability. The increasing integration of digital technologies in modern power grids, particularly the advent of smart grids (SGs), while offering substantial benefits like enhanced two-way communication and grid optimization, also introduces new security vulnerabilities. These vulnerabilities stem from the convergence with technologies like the Internet of Things (IoT) and Industrial Control Systems (ICS), which are inherently susceptible to cyber threats. Consequently, robust and intelligent intrusion detection mechanisms are paramount to mitigating these risks and ensuring the resilience of the power grid.

Several research efforts have focused on developing advanced IDS solutions tailored for the unique challenges of power systems. For instance, the research presented in [16] proposes a novel anomaly-based Intrusion Detection System specifically designed for smart grids. This approach leverages real-world power plant data and employs sophisticated machine learning and deep learning models, incorporating innovative feature engineering techniques. The study demonstrates the effectiveness of this IDS in accurately detecting a range of cyber threats targeting the SG infrastructure, highlighting the potential of data-driven approaches in enhancing grid security.

Another promising approach is presented in [17], which proposes a new automated, hybrid IDS. This system combines the strengths of data mining, specifically common path mining, to learn complex temporal patterns from diverse data sources, including synchrophasor measurements and audit logs. By analyzing these patterns, the IDS can effectively distinguish between normal operations, disturbances, and malicious cyberattacks. The researchers successfully tested a prototype of their system on a realistic two-line, three-bus transmission system, demonstrating its ability to accurately classify these events within a distance protection scheme. This hybrid approach offers a powerful combination of data-driven pattern recognition and expert knowledge, providing a robust defense against sophisticated attacks.

Addressing the challenges of data scarcity and adversarial manipulation in training effective IDS models, the work in [18] delves into the important areas of imbalanced learning and adversarial learning for training anomaly-based intelligent IDSs (AN-Intel-IDS). Through a rigorous qualitative

 doi:10.20944/preprints202412.1697.v1



study employing rapid review, structured reporting, and subgroup analysis, the research surveys and synthesizes various generative-based data augmentation techniques. These techniques aim to address the prevalent issue of uneven data distribution in cybersecurity datasets, where normal operation data significantly outweighs attack data. Furthermore, the study explores generative-based adversarial techniques for creating synthetic yet realistic data in an adversarial setting. This approach enhances the robustness of the IDS by training it to recognize and defend against evolving attack strategies, ultimately improving the overall security posture of the power system.

#### 2.4.2. Attack Detection and Response

Modern power system algorithms, crucial for real-time decision-making in functions like dispatch optimization and wide-area control, rely heavily on accurate data. However, the increasing sophistication of cyber-attacks poses a significant threat. Should an attack bypass intrusion detection systems (IDS) and compromise this data, the resulting erroneous decisions can severely degrade power system performance, potentially leading to instability and even catastrophic system collapse. This vulnerability underscores the urgent need for robust cyber-resilient control and protection strategies.

Several recent research efforts have explored data-driven approaches to enhance power system security and resilience. In [19], researchers leverage a deep learning framework to significantly improve the operational accuracy of protection devices across a wide range of system events. Their work specifically addresses persistent protection gaps related to challenging scenarios like transmission line high-impedance faults and transformer inter-turn faults. By employing data-driven techniques, they demonstrate a substantial improvement in accuracy compared to traditional protection methods, offering a promising avenue for enhanced fault detection and isolation.

Furthermore, the threat of false data injection attacks, specifically targeting wide-area control systems, has been addressed through innovative data-driven solutions. In [20], a semi data-driven method utilizing a deep learning framework is developed to detect these malicious data injections. Beyond detection, the method actively mitigates the impact of such attacks by dynamically adapting the wide-area control structure, effectively minimizing the influence of compromised wide-area network communication channels. Building upon this work, [21] presents a fully model-free approach, employing a Long Short-Term Memory (LSTM) architecture. This method not only detects the compromised communication channel but also replaces the corrupted data with information predicted by the LSTM network, effectively restoring the integrity of the control system. This proactive data restoration significantly enhances the system's resilience against cyber-attacks.

Expanding the scope to distributed generation, [22] proposes a novel method for detecting and diagnosing both cyber-attacks and physical faults (open/short circuits) within photovoltaic (PV) solar farms using a single waveform sensor. By employing advanced signal processing techniques, this approach can effectively distinguish between these events and even identify previously unseen cyber-attack patterns. Notably, the method's efficacy has been validated in a real-time simulation environment, demonstrating its potential for practical implementation in securing distributed energy resources. These advancements highlight the growing importance of data-driven methods in safeguarding the power grid against evolving cyber and physical threats.

## 3. Challenges and Future Directions

While data-driven methods offer significant potential, several challenges need to be addressed:

Data Quality and Quantity: The quality and quantity of data are crucial for the success of data-driven methods. Ensuring data accuracy, completeness, and consistency is essential. Model Development and Validation: Developing accurate and reliable data-driven models requires careful consideration of feature engineering, model selection, and validation techniques. Computational Efficiency: Data-driven methods often involve complex algorithms and large datasets, which can be computationally intensive. Developing efficient algorithms and leveraging advanced computing techniques is important. Integration with Existing Systems: Integrating data-driven methods with

 doi:10.20944/preprints202412.1697.v1



existing power system control and monitoring systems requires careful planning and coordination. Cyber-Security Risks: Protecting sensitive data and ensuring the security of data-driven systems is crucial to prevent cyber-attacks that could compromise the reliability and security of the power grid.

**References**


1. P. Mukhopadhyay, G. Mitra, S. Banerjee and G. Mukherjee, "Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter," 2017 7th International Conference on Power Systems (ICPS), Pune, India, 2017, pp. 812-819, doi: 10.1109/ICPES.2017.8387401.
2. F. L. Quilumba, W. -J. Lee, H. Huang, D. Y. Wang and R. L. Szabados, "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities," in IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 911-918, March 2015, doi: 10.1109/TSG.2014.2364233.
3. L. Guo, P. Shi, Y. Zhang, Z. Cao, Z. Liu and B. Feng, "Short-term EV Charging Load Forecasting Based on GA-GRU Model," 2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 2021, pp. 679-683, doi: 10.1109/AEEES51875.2021.9403141.
4. J. R. Andrade and R. J. Bessa, "Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions," in IEEE Transactions on Sustainable Energy, vol. 8, no. 4, pp. 1571-1580, Oct. 2017, doi: 10.1109/TSTE.2017.2694340.
5. K. E. Hagan, O. O. Oyebanjo, T. M. Masaud and R. Challoo, "A probabilistic forecasting model for accurate estimation of PV solar and wind power generation," 2016 IEEE Power and Energy Conference at Illinois (PECI), Urbana, IL, USA, 2016, pp. 1-5, doi: 10.1109/PECI.2016.7459241.
6. M. Shamsi and P. Cuffe, "Prediction Markets for Probabilistic Forecasting of Renewable Energy Sources," in IEEE Transactions on Sustainable Energy, vol. 13, no. 2, pp. 1244-1253, April 2022, doi: 10.1109/TSTE.2021.3112916.
7. J. Zhang, Y. Liu, A. Rouhani, J. Zhao, G. J. Warchol and K. Scott, "A Robust Data-Driven Gaussian Process Regression-Enabled Distribution System State Estimation using Partial Node Data," 2024 IEEE Power & Energy Society General Meeting (PESGM), Seattle, WA, USA, 2024, pp. 1-5, doi: 10.1109/PESGM51994.2024.10688998.
8. B. Zargar, A. Angioni, F. Ponci and A. Monti, "Multiarea Parallel Data-Driven Three-Phase Distribution System State Estimation Using Synchrophasor Measurements," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 9, pp. 6186-6202, Sept. 2020, doi: 10.1109/TIM.2020.2967512.
9. . Hu, Q. Wang, Y. Ye, Z. Wu and Y. Tang, "A High Temporal-Spatial Resolution Power System State Estimation Method for Online DSA," in IEEE Transactions on Power Systems, vol. 39, no. 1, pp. 877-889, Jan. 2024, doi: 10.1109/TPWRS.2023.3240826.
10. Q. Wang, F. Li, Y. Tang and Y. Xu, "Integrating Model-Driven and Data-Driven Methods for Power System Frequency Stability Assessment and Control," in IEEE Transactions on Power Systems, vol. 34, no. 6, pp. 4557-4568, Nov. 2019, doi: 10.1109/TPWRS.2019.2919522.
11. Kar J, Chakrabortty A., "Scalable design methods for online data-driven wide-area control of power systems", IET Gener. Transm. Distrib. 2021.
12. E. Ekomwenrenren, J. W. Simpson-Porco, E. Farantatos, M. Patel, A. Haddadi and L. Zhu, "Data-Driven Fast Frequency Control Using Inverter-Based Resources," in IEEE Transactions on Power Systems, vol. 39, no. 4, pp. 5755-5768, July 2024, doi: 10.1109/TPWRS.2023.3337011.
13. Y. Zhang, S. Shen and J. L. Mathieu, "Data-driven optimization approaches for optimal power flow with uncertain reserves from load control," 2015 American Control Conference (ACC), Chicago, IL, USA, 2015, pp. 3013-3018, doi: 10.1109/ACC.2015.7171795.
14. Jing, H., Wang, Y. & Li, Y. "Data-driven quantum approximate optimization algorithm for power systems", Commun Eng 2, 12 (2023).
15. J. Zhang, Z. Chen, C. He, Z. Jiang and L. Guan, "Data-Driven-Based Optimization for Power System Var-Voltage Sequential Control," in IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2136-2145, April 2019, doi: 10.1109/TII.2018.2856826.
16. G. Efstathopoulos et al., "Operational Data Based Intrusion Detection System for Smart Grid," 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Limassol, Cyprus, 2019, pp. 1-6, doi: 10.1109/CAMAD.2019.8858503.

17. S. Pan, T. Morris and U. Adhikari, “Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems," in IEEE Transactions on Smart Grid, vol. 6, no. 6, pp. 3104-3113, Nov. 2015, doi: 10.1109/TSG.2015.2409775.
18. Abdelmoumin, G.; Whitaker, J.; Rawat, D.B.; Rahman, A. “A Survey on Data-Driven Learning for Intelligent Network Intrusion Detection Systems", Electronics 2022, 11, 213.
19. R. Fan et. al., “New data-driven approach to bridging power system protection gaps with deep learning" in Electric Power Systems Research , Volume 208, July 2022.
20. J. Kar and A. Chakrabortty, “Neural Network-Assisted Resilient Wide-Area Control of Power Systems under Denial-of-Service Attacks," 2021 IEEE Conference on Control Technology and Applications (CCTA), San Diego, CA, USA, 2021, pp. 917-922, doi: 10.1109/CCTA48906.2021.9658660.
21. J. Kar and A. Chakrabortty, “LSTM based Denial-of-Service Resiliency for Wide-Area Control of Power Systems," 2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), Espoo, Finland, 2021, pp. 1-5, doi: 10.1109/ISGTEurope52324.2021.9640193.
22. L. Guo, J. Zhang, J. Ye, S. J. Coshatt and W. Song, “Data-Driven Cyber-Attack Detection for PV Farms via Time-Frequency Domain Features," in IEEE Transactions on Smart Grid, vol. 13, no. 2, pp. 1582-1597, March 2022, doi: 10.1109/TSG.2021.3136559.