RESEARCH ARTICLE

Multi-Turbine Dependency Reconstruction for Non-Stationary Wind Farm Operation Anomaly Detection

Anna Keller[1], Lukas Meier[2*]

School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland

*Corresponding author: l.meier@epfl.ch

Abstract

Wind farms generate large-scale operational time series from turbine speed, blade pitch angle, generator temperature, gearbox vibration, nacelle direction, wind speed, power output, and converter status. These signals are strongly affected by wind variability, wake interaction, seasonal weather, turbine aging, and control strategy changes, making anomaly detection under non-stationary conditions challenging. This study proposes a multi-turbine dependency reconstruction method for non-stationary wind farm operation anomaly detection. The method constructs a turbine dependency graph using spatial layout, wake influence, wind-direction alignment, and lagged power correlation. A temporal reconstruction network is then used to learn normal multi-turbine operating patterns under changing wind regimes. An adaptive residual envelope is introduced to detect abnormal deviation without using a fixed global threshold. Experiments are conducted on a wind farm dataset containing 132 turbines, 48 SCADA variables, and 5-minute observations collected over 32 months. The dataset includes 54 million operational records and 1,860 maintenance-confirmed abnormal events, including gearbox overheating, pitch-control failure, yaw misalignment, converter instability, and abnormal power curve deviation. The proposed method reduces median detection delay from 9.3 hours to 2.4 hours compared with an isolated-turbine LSTM baseline. False alerts decrease to 1.7 events per turbine per month. The Matthews correlation coefficient reaches 0.902 under seasonal wind-regime changes, while residual-envelope drift remains below 0.27 standard units across winter and summer conditions. Online inference processes 18,600 SCADA points per second, and one farm-level daily assessment completes in 5.9 minutes. The results demonstrate that multi-turbine dependency reconstruction can improve abnormal operation detection in non-stationary wind farm time series.

Keywords: Wind farm monitoring; non-stationary time series; operation anomaly detection; multi-turbine dependency; SCADA data; residual reconstruction; renewable energy systems.

Introduction

Wind farms have become an important component of renewable-energy systems, making reliable turbine monitoring essential for stable power generation, operational safety, and timely maintenance. Modern wind turbines continuously generate supervisory control and data acquisition data covering rotor speed, blade pitch angle, generator temperature, gearbox vibration, nacelle direction, wind speed, active power output, and converter status. These measurements jointly reflect turbine health, environmental conditions, and control-system behavior. Recent studies have applied autoencoders, recurrent neural networks, ensemble learning, and probabilistic models to identify turbine faults, performance degradation, and abnormal operating states [1,2]. These data-driven approaches reduce dependence on manually defined fault rules and provide a practical basis for condition monitoring in large wind farms [3,4]. Wind turbine SCADA data are inherently multivariate, and faults often influence several variables simultaneously. For example, gearbox degradation may gradually affect vibration, temperature, rotational speed, and power conversion efficiency, while converter abnormalities may produce coordinated changes in electrical output and control-related signals. Fine-grained multivariate anomaly detection based on causal inference has shown that directional dependencies among variables can help distinguish directly affected measurements from secondary or coincidental deviations [5]. This perspective is particularly relevant to wind turbine monitoring because abnormal behavior may be weak in individual SCADA variables but more evident in changes to their dependency structure. Probabilistic and uncertainty-aware approaches have also been developed to improve fault identification under noisy and incomplete operating data [6,7].

Despite these advances, anomaly detection in wind farms remains challenging because turbine signals are strongly affected by non-stationary environmental and operational conditions [8,9]. Wind speed, wind direction, turbulence intensity, air density, seasonal weather, turbine aging, and control-strategy updates can continuously alter the statistical distributions of SCADA variables [10]. A level of generator temperature or power fluctuation that is normal under one wind regime may appear abnormal under another. Fixed thresholds and globally trained reconstruction models may therefore generate excessive false alarms during normal changes in wind conditions [11,12]. Conversely, early-stage faults may remain undetected when their deviations are small, persistent, and partially masked by natural wind variability. Many existing methods also analyze turbines as independent monitoring units [13,14]. This treatment overlooks the fact that turbines within the same wind farm operate under spatially and aerodynamically related conditions [15,16]. Farm layout, prevailing wind direction, terrain, wake propagation, and turbine spacing create dependencies in power output and operating behavior. An upstream turbine may reduce the effective wind speed and increase turbulence for downstream turbines, causing delayed and direction-dependent changes in their SCADA signals [17,18]. Consequently, a deviation observed in one turbine cannot always be interpreted correctly without considering neighboring turbines and the current wind field. Independent-turbine models may incorrectly classify farm-wide wind-regime transitions as equipment faults or fail to recognize localized anomalies that disrupt the expected relationship among turbines [19,20]. Another limitation concerns the use of static turbine relationships. Wake influence changes with wind direction and operating state, while lagged power correlations may vary across seasons, control modes, and maintenance periods [21,22]. A dependency model constructed from long-term average correlations may therefore be unable to represent the current spatial interaction among turbines. Effective farm-level anomaly detection requires a dynamic representation that integrates physical layout, wake effects, wind-direction alignment, and data-derived temporal correlations [23,24]. It must also distinguish short disturbances from persistent degradation and adapt its decision boundary to different wind regimes without relying on a single global anomaly threshold.

To address these limitations, this study aims to develop a multi-turbine dependency reconstruction method for anomaly detection in non-stationary wind farm environments. The proposed method constructs a turbine dependency graph by integrating farm layout, wake influence, wind-direction alignment, and lagged power correlation, allowing inter-turbine relationships to vary with operating conditions. A temporal reconstruction network is used to learn normal multivariate patterns across connected turbines under changing wind regimes. An adaptive residual envelope is further introduced to estimate regime-dependent anomaly boundaries and identify weak but persistent deviations without using a fixed global threshold. The method is evaluated using SCADA data from 132 turbines, including 48 monitored variables, approximately 54 million timestamped records, and 1,860 maintenance-confirmed abnormal events collected over 32 months. Detection delay, false alert volume, fault-event coverage, and computational efficiency are jointly examined to assess both detection performance and practical deployment capability. By combining dynamic inter-turbine dependency modeling with adaptive residual evaluation, this study seeks to improve the separation of genuine equipment abnormalities from normal wind-induced variations, reduce unnecessary maintenance alerts, and support scalable farm-level condition monitoring and preventive maintenance.

Materials and Methods

2.1 Study samples and wind farm description

SCADA data were collected from a wind farm with 132 turbines. Each turbine recorded 48 variables, including turbine speed, blade pitch, generator temperature, gearbox vibration, nacelle direction, wind speed, power output, and converter status. Measurements were taken every 5 minutes over 32 months, yielding 54 million records. The farm included upstream, middle, and downstream turbines under varying wake conditions. The dataset contained 1,860 confirmed abnormal events, such as gearbox overheating, pitch-control failure, yaw misalignment, converter instability, and abnormal power curves. Data covered winter and summer, low and high wind, and different turbine ages.

2.2 Experimental design and control groups

The test group used the multi-turbine dependency reconstruction model, which combines spatial layout, wake influence, wind-direction alignment, and lagged power correlation. Two control groups were included. The first used isolated-turbine LSTMs analyzing each turbine separately. The second used a static correlation-based reconstruction method without updating turbine dependencies for changing wind conditions. All models were trained on normal operating periods and tested on periods with seasonal wind changes and confirmed abnormal events. This design tests whether incorporating turbine links, wake effects, and wind-regime adaptation improves anomaly detection.

2.3 Measurement methods and quality control

SCADA variables were collected by turbine control systems and reviewed before analysis. Duplicate timestamps, invalid values, and records during planned shutdowns were removed. Wind speed and power values outside operating limits were flagged. Short missing gaps (<3 records) were filled using linear interpolation; longer gaps were excluded. Maintenance logs, alarms, and inspection reports verified abnormal events. Grid curtailment and control actions were also checked to avoid false labeling. All turbines were time-aligned for consistent farm-level analysis.

2.4 Data processing and model formulation

Variables were aligned by turbine and timestamp. Each variable was normalized using training period statistics:

$$x'_{i,k,t} = \frac{x_{i,k,t} - \mu_{i,k}}{\sigma_{i,k} + \varepsilon}$$

where $x_{i,k,t}$ is the raw measurement, $\mu_{i,k}$ and $\sigma_{i,k}$ are the mean and standard deviation, and ϵ avoids division by zero. The turbine dependency graph was constructed using turbine distance, wind-direction alignment, wake relation, and lagged power correlation. The anomaly score for turbine i at time t was:

$$S_{i,t} = \sum_{k=1}^{K} w_k | x'_{i,k,t} - \hat{x}'_{i,k,t} |$$

where K is the number of SCADA variables, w_k is the variable weight, and $\hat{x}'_{i,k,t}$ is the reconstructed value. An event was recorded if $S_{i,t}$ exceeded the adaptive residual envelope for a continuous period.

2.5 Evaluation protocol and implementation

The dataset was divided into training, validation, and testing sets by time to prevent information leakage. Training learned normal turbine behavior, validation tuned the residual envelope, and testing evaluated event-level detection. Metrics included median detection delay, false alerts per turbine per month, Matthews correlation coefficient, residual-envelope drift, online inference speed, and daily farm-level assessment time. All control methods used the same periods. The proposed method processed 18,600 SCADA points per second, and a daily farm assessment completed in 5.9 minutes, demonstrating suitability for large wind farm monitoring.

Results and Discussion

3.1 Overall anomaly detection performance

The multi-turbine dependency reconstruction method improved anomaly detection under changing wind farm conditions. Compared with the isolated-turbine LSTM baseline, the median detection delay decreased from 9.3 h to 2.4 h. This shows that the method can report faults earlier, including gearbox overheating, pitch-control failure, yaw misalignment, converter instability, and abnormal power curve changes [25,26]. The false alert rate decreased to 1.7 events per turbine per month, which helps reduce unnecessary maintenance checks. The Matthews correlation coefficient reached 0.902 under seasonal wind-regime changes, indicating stable detection performance. In contrast, fixed thresholds and single-turbine models often fail to capture wake effects and farm-level operating changes. Fig. 1 shows a representative SCADA-based wind turbine anomaly detection framework with feature extraction and structured analysis.

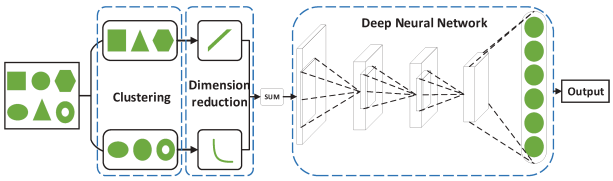

Fig. 1. *Framework for detecting wind turbine anomalies from SCADA data.*

3.2 Effect of multi-turbine dependency reconstruction

The turbine dependency graph improved the detection of faults involving several turbines. In wind farms, wake effects and wind-direction changes often cause downstream turbines to show delayed or coupled deviations. Isolated-turbine methods may miss these patterns because they analyze each turbine separately. The proposed method uses spatial layout, wake influence, wind-direction alignment, and lagged power correlation to describe turbine links. This helps separate true abnormal operation from normal wind variation. Compared with static correlation reconstruction, the proposed method better reflects physical relations among turbines [27,28]. It also improves detection of fault propagation across the farm.

3.3 Robustness under changing wind regimes

The adaptive residual envelope improved stability under different wind conditions. SCADA signals such as power output, vibration, generator temperature, and pitch angle vary with wind speed, turbulence, season, and control actions. These changes may occur during normal operation and should not always be treated as faults. Fixed thresholds are sensitive to such changes and may cause many false alerts. The adaptive envelope adjusts to recent operating states and reduces errors caused by normal regime shifts. Residual-envelope drift remained below 0.27 standard units across winter and summer conditions. Fig. 2 illustrates a SCADA-based anomaly detection workflow with preprocessing, feature extraction, and model evaluation, supporting the need for structured analysis in non-stationary turbine data.

Fig. 2. *Workflow for SCADA data cleaning, feature extraction, and fault evaluation.*

3.4 Comparison with existing methods and practical value

Compared with isolated machine learning models and static reconstruction methods, the proposed method provides faster detection, fewer false alerts, and better farm-level fault monitoring. Single-turbine models can capture local faults but often ignore wake interaction and shared wind conditions. Static reconstruction methods may also fail when wind regimes change. The proposed method addresses these issues by combining turbine dependency reconstruction with adaptive residual scoring. Online inference processed 18,600 SCADA points per second, and one daily farm-level assessment was completed in 5.9 min. This indicates that the method is suitable for continuous wind farm monitoring. However, the study used data from one wind farm only [29,30]. Future work should test more farm layouts and include turbulence intensity, atmospheric stability, and maintenance-cost data to improve fault explanation and practical use.

Conclusion

The multi-turbine dependency reconstruction method offers a practical way to detect abnormal wind farm operation under changing wind conditions. It combines turbine layout, wake effects, wind-direction links, lagged power correlation, and adaptive residual limits. This allows the method to capture both single-turbine faults and fault spread across the farm. The results show shorter detection delays, fewer false alerts, stable MCC values, and reliable residual changes across winter and summer regimes. The main contribution is the move from isolated-turbine monitoring to farm-level detection based on turbine links. This helps explain how abnormal signals spread among turbines under shared wind conditions. The method can support early fault warning, maintenance planning, and continuous SCADA-based monitoring. Limitations include testing on one wind farm and using mainly SCADA records. Future studies should test different farm layouts and include turbulence intensity, atmospheric stability, maintenance cost, and inspection records to improve fault explanation and practical use.

References

Wu, C., Zhang, F., Chen, H., & Zhu, J. (2025). Design and optimization of low power persistent logging system based on embedded Linux.

Karimi Dastgerdi, A., Mercorelli, P., & Nemati, H. (2026). Autoencoder-driven fault detection in drilling machines: a study of hybrid deep learning models. International Journal of Computer Integrated Manufacturing, 1-18.

Wang, Y., Chen, J., Arias, R., Wang, Y., & Yin, X. (2026). Development and Validation of a Patient-Friendly Digital Assessment Platform for Precision Screening of Oral Anti-Obesity Medications (AOMs).

Alagha, N., Khairuddin, A. S. M., Haitaamar, Z. N., Al-Khatib, O., & Kanesan, J. (2025). Artificial intelligence in wind turbine fault detection and diagnosis: Advances and perspectives. Energies, 18(7), 1680.

Chen, X., Xiao, H., Zeng, Z., Zhang, S., & Xiao, R. (2025). Fine-Grained Multivariate Time Series Anomaly Detection via Causal Inference. Knowledge-Based Systems, 114765.

Yin, J., Huang, Y., & Rao, H. (2026). A Study on the Dynamic Evolution of Learning Behavior and Outcome Prediction in Digital Educational Environments. Available at SSRN 6607139.

Haseeb, F., Peng, M., Arshad, F., & Ali, W. (2026). Uncertainty aware unsupervised fault diagnosis of PWR nuclear power plant using KNN and SHAP method. Progress in Nuclear Energy, 193, 106191.

Yang, J. (2026). Computational Analysis of How Digital Non-Clinical Communication Tools Influence Social Participation Among Older Adults in Community-Based Elderly Care. Available at SSRN 6682838.

Zhang, Z. (2026). A Study on Multimodal Product Understanding and Assembly Guidance Optimization in e-commerce for Complex Consumer Goods. Available at SSRN 6734559.

Gerbino, M. (2025). Data-Driven Diagnosis of Wind Turbines via SCADA-Based Anomaly Detection Methods (Doctoral dissertation, Politecnico di Torino).

Qi, C., & Qiao, X. (2026, May). Design Patterns for Multi-Agent Systems in Production. In 2026 International Conference on Intelligent Systems, Automation and Control (ISAC) (pp. 198-202). IEEE.

Su, D., & Shi, X. Optimizing Family–School Collaboration to Improve Educational and Social Outcomes for Children with Intellectual Disabilities in Inclusive Public Schools.

Sadeghi, R., Parenti, M., Memme, S., Fossa, M., & Morchio, S. (2025). A review and comparative analysis of solar tracking systems. Energies, 18(10), 2553.

Shao, W. (2026, March). Interpretable Ensemble Learning for Network Traffic Anomaly Detection: A SHAP-based Explainable AI Framework for Embedded Systems Security. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-4). IEEE.

Fu, Y., Gui, H., Li, W., & Wang, Z. (2020, August). Virtual Material Modeling and Vibration Reduction Design of Electron Beam Imaging System. In 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 1063-1070). IEEE.

Khan, M. A., Churchfield, M. J., & Watson, S. J. (2026). Dependence of wind-farm-induced gravity waves and wind farm performance on non-dimensional atmospheric parameters and simulation configuration. Wind Energy Science, 11(5), 1631-1652.

Chen, H., Ma, X., Mao, Y., & Ning, P. (2025). Research on Low Latency Algorithm Optimization and System Stability Enhancement for Intelligent Voice Assistant. Available at SSRN 5321721.

Wasilczuk, F., Kasprzak, P., Rybiński, M., Flaszyński, P., Dzierwa, P., Ziędalska, K., ... & Taler, D. (2025). Positive effect of wind farm control on energy production and revenue in European markets. Energy, 326, 136159.

Liang, R., Feifan, F. N. U., Liang, Y., & Ye, Z. (2025). Emotion-Aware Interface Adaptation in Mobile Applications Based on Color Psychology and Multimodal User State Recognition. Frontiers in Artificial Intelligence Research, 2(1), 51-57.

Yang, M., Wang, Y., Shi, J., & Tong, L. (2025). Reinforcement Learning Based Multi-Stage Ad Sorting and Personalized Recommendation System Design.

Bekbolatova, Z., Grigoryev, D., Minazhova, S., Bekbayev, A., Omarov, Z., & Kozhageldi, B. (2026). Wind Power Forecasting Based on Time Series Autoregressive Moving Average Model Taking into Account the Effect of Wind Speed Dispersion. ES Energy and Environment, 31, 2025.

Deng, H. C., Shu, Z. R., Chan, P. W., & He, X. H. (2026). Cross-height causal dependence of horizontal wind components in offshore environment: insights from transfer entropy analysis. Sustainable Energy Technologies and Assessments, 87, 104890.

Gao, G., Gao, R., Lu, C., Gao, R., & Kuang, Y. (2026, March). Security Governance Methods and Quantitative Evaluation for Enterprise SMS and Verification Code Systems. In 2026 International Conference on Generative Artificial Intelligence and Information Security (GAIIS) (pp. 455-458). IEEE.

Xu, T., Zhang, J., & Zhu, W. (2026). Fairness-Accuracy Trade-offs and Calibration Mechanisms in Machine Learning-Based Subprime Credit Scoring. Available at SSRN 6893759.

Gerbino, M. (2025). Data-Driven Diagnosis of Wind Turbines via SCADA-Based Anomaly Detection Methods (Doctoral dissertation, Politecnico di Torino).

Xie, Z., Ren, X., Zheng, T., Bai, J., Fan, W., Xu, B., ... & Song, Y. (2026). A Survey on AI Agent Harness. ResearchGate Preprint. DOI, 10.

Özkan, R., & Genç, M. S. (2025). Comparison of Numerical Results With Experiment Using Digital Image Correlation System of Multiobjective Structural Optimised Wind Turbine Blade. Wind Energy, 28(9), e70051.

Zhang, Z., Tong, Y., & Gao, Y. (2026). Causal Representation Learning for Explainable Early Warning in High-Stakes Decision Systems.

Al Hassan, A., & Dao, P. B. (2025). Bridging data and diagnostics: A systematic review and case study on integrating trend monitoring and change point detection for wind turbines. Energies, 18(19), 5166.

Qi, C., & Qiao, X. (2026). Building and Operating a Large Scale Multi-Agent System: A Case Study from Industry. Available at SSRN 6795198.