https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Sensor-Error-Robust Normal Behavior Modeling for Wind Turbine Failure Prediction Using a Masked Autoencoder

Xavier Chesterman[1,2,3,4], Ann Nowé[1,2,3], and Jan Helsen[1,3,4]

[1]Acoustics Vibration Research Group, Vrije Universiteit Brussel, Brussels, Belgium
[2]AI lab, Vrije Universiteit Brussel, Brussels, Belgium
[3]Flanders Make@VUB, Flanders Make, Lommel, Belgium
[4]OWI, Vrije Universiteit Brussel, Brussels, Belgium

Correspondence: Xavier Chesterman (xavier.chesterman@vub.be)

Abstract. Condition monitoring and failure prediction in wind turbines has become an increasingly important research area due to its substantial economic impact. Accurate early detection of developing faults enables more efficient maintenance planning and minimizes costly downtime. However, predicting failures from operational wind farm data remains challenging. Real-world datasets are often affected by measurement noise, incomplete expert knowledge, and extraneous operating conditions, all of which complicate the identification and classification of emerging problems. This work presents a methodology designed to address one critical obstacle: measurement errors caused by faulty or unreliable sensors. Such errors can substantially degrade the performance of Normal Behavior Models (NBM), thereby hindering the detection of anomalies and incipient failures. To mitigate this issue, we introduce an approach based on masked autoencoders (MAE) that selectively suppresses signals deemed unreliable by domain experts or automated diagnostics. The proposed method is evaluated using four datasets from real operational wind farms. We analyze the impact of sensor-induced errors on NBM performance and demonstrate how the MAE framework improves robustness in the presence of corrupted measurements. The results highlight the potential of the method to improve the reliability of data-driven failure prediction systems in practical wind turbine applications.

1 Introduction

The wind energy sector has expanded rapidly in recent years, driven by substantial investments and ambitious renewable energy targets. In 2024, Europe installed 16.4 gigawatt (GW) of new wind power capacity, including 12.9 GW within the European Union, bringing the total installed capacity of the continent to 285 GW. Onshore installations accounted for 84% of this new capacity. Projections indicate that Europe will add approximately 31 GW of new wind capacity annually between 2025 and 2030, of which 23 GW is expected within the EU, resulting in a total installed capacity of roughly 351 GW by 2030 (Costanzo et al., 2025). These figures underscore the central and enduring role of wind power in the European energy system, and they highlight the significant economic importance of reducing the cost of wind energy.

Operation and maintenance (O&M) expenditures constitute a substantial share of the levelized cost of energy (LCOE) for wind turbines. For onshore turbines in the United States, O&M costs represent approximately 25% of the total LCOE (42 USD/megawatt-hour (MWh)). For offshore fixed-bottom turbines, the share increases to 26.9% (117 USD/MWh), while for

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

floating offshore turbines it is 17.8% (181 USD/MWh), the latter percentage being lower due to the larger proportion of capital expenditure in the total cost (Stehly et al., 2024). These values make clear that reducing O&M costs is critical for further lowering the cost of wind energy.

A key strategy for reducing O&M expenditures is improving maintenance planning through reliable failure prediction. By identifying component degradation early, operators can schedule interventions more efficiently and avoid costly unplanned downtime. As a result, failure prediction and condition monitoring have become active research areas, employing a wide range of modeling approaches and data sources (Kestel et al., 2025). However, accurate prediction remains challenging due to data quality issues, incomplete expert knowledge, and the scarcity of failure examples.

This paper addresses one specific but impactful data quality issue: sensor errors. Data-driven monitoring methods are highly sensitive to corrupted measurements, and sensor failures can degrade model performance from mildly inaccurate to entirely unreliable. For example, in Normal Behavior Models (NBM), a single distorted input signal can propagate errors through the model and significantly compromise anomaly detection. To mitigate this problem, we propose a methodology based on masked autoencoders (MAE) that enables selective deactivation of unreliable sensor signals while retaining the use of the full model without the need to train multiple parallel models.

The proposed approach is validated using data from four operational wind farms. The sensor failures investigated are real, not artificially generated. Through collaboration with industry experts, several characteristic sensor failure patterns were identified, enabling realistic simulation of field conditions. The validation assesses the effectiveness of the masking strategy using multiple performance metrics, evaluates the capability of the model as an NBM, and analyzes the impact of masking on model stability.

2 Related work

2.1 SCADA-based condition monitoring of wind turbines

Condition monitoring (CM) of wind turbines has emerged as a rapidly expanding research field, encompassing a wide range of methodological approaches. Based on the primary source of information, CM techniques can be categorized into three main groups: (i) Supervisory Control and Data Acquisition (SCADA)-based approaches, which rely on readily available low-frequency data from the SCADA system of the wind turbine; (ii) vibration-based approaches, which utilize high-frequency signals collected from dedicated accelerometers; and (iii) acoustic emission–based approaches, which analyze high-frequency acoustic signals generated by turbine components (Kestel et al., 2025). Each category exhibits distinct advantages and limitations and addresses specific needs within the CM domain.

This paper focuses on SCADA-based CM methods that exploit low-frequency operational data. The most commonly used temporal resolution for SCADA data is 10 minutes, although higher sampling frequencies have also been reported in the literature (Chesterman et al., 2023). Within the SCADA-based paradigm, Wang et al. (2026) distinguish between physical model–based approaches, which rely on explicit representations of turbine dynamics, and data-driven approaches, in which models are learned directly from historical SCADA data. This paper adopts the data-driven approach.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Data-driven CM encompasses a broad spectrum of methodologies. According to Wang et al. (2026), these methods can be divided into traditional data-driven techniques (e.g., trending analysis, statistical methods, and state-curve modeling) and machine learning– and deep learning–based approaches, including supervised, semi-supervised, and unsupervised methods. An alternative taxonomy is proposed by Tautz-Weinert and Watson (2017), who identify NBM, trending, clustering, damage modeling, alarm assessment, and expert systems as distinct categories. Comprehensive reviews of recent advances in wind turbine CM are provided in Wang et al. (2026), Kestel et al. (2025), Black et al. (2021), and Chatterjee and Dethlefs (2021).

This paper adopts the NBM approach, which has demonstrated strong performance in wind turbine health monitoring. A key advantage of NBM methods is that they learn a representation of normal system behaviour that can be used for prediction and anomaly detection. NBM techniques can themselves be subdivided into several groups, with statistical methods, shallow machine learning models, and deep learning architectures being the most prominent (see Kestel et al. (2025) for an extensive overview).

The statistical group includes conventional modelling and dimensionality-reduction techniques such as ordinary least squares regression (OLS) (Chesterman et al., 2022), principal component analysis (PCA) (Campoverde et al., 2022), and cointegration analysis (Dao, 2023).

Shallow machine learning models comprise widely used algorithms that are more flexible than statistical methods, particularly for capturing nonlinear relationships. Examples applied in wind turbine CM include random forests (RF) (Turnbull et al., 2021), gradient boosting machines (GBM) (Maron et al., 2022), and support vector machines (SVM) (Castellani et al., 2021). While these models can achieve higher predictive accuracy, they typically require larger training datasets and offer reduced interpretability compared to statistical approaches.

Deep learning models represent the most recent and rapidly expanding class of NBM techniques. These methods rely on neural network architectures with multiple layers and have been applied extensively in modern CM systems. Common examples include deep neural networks (Verma et al., 2022), convolutional neural networks (CNN) (Bermúdez et al., 2022), long short-term memory networks (LSTM) (Trizoglou et al., 2021), and autoencoders (AE) (Lee et al., 2024).

2.2 Sensor errors

This work addresses failure prediction in the presence of sensor errors, which are conditions under which corrupted sensor signals can significantly degrade predictive performance. Following the ISO definition, a measurement error is the difference between the measured value and the true value of the measurand, and can generally be decomposed into random and systematic components (BIPM et al., 2008). We define a sensor error as any condition causing a sensor to produce such measurement errors. In a similar vein, Balaban et al. (2009) describe a sensor error or fault as an unexpected deviation in the observed signal in the absence of any underlying anomalous system behavior.

A wide variety of sensor fault modes has been reported in the literature, and several taxonomies have been proposed. Balaban et al. (2009) identify the following major categories:

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

- Bias: A constant offset from the nominal signal, often caused by miscalibration or physical changes in the sensing element (e.g., a temperature sensor consistently overestimating by 10 °C).
- Drift: A time-varying offset from nominal behavior.
- Scaling: A multiplicative distortion of the signal magnitude while preserving waveform shape.
- Noise: Random fluctuations superimposed on the true signal.
- Hard faults: The signal becomes fixed at a particular value, including complete signal loss or a "stuck" sensor.

An alternative but partially overlapping taxonomy is provided by Teh et al. (2020), who introduce an additional outlier category representing abrupt, unstructured deviations beyond expected thresholds. They also identify specific subtypes such as "stuck-at-zero," which is a sub-category of the hard-fault class described in Balaban et al. (2009).

In practice, data acquired from complex industrial systems often contain multiple fault types simultaneously. This complicates downstream analysis, as some faults, such as bias, may be subtle and not visually apparent. Their detection typically requires models that capture relationships across multiple sensor signals. A wide range of such models has been explored. Two common strategies are data imputation and data fusion. Teh et al. (2020) review imputation methods for handling missing or noisy data, including association rule mining (ARM), clustering, k-nearest neighbors (KNN), singular value decomposition (SVD), and various hybrid techniques. Approaches for denoising include empirical mode decomposition, Savitzky–Golay filtering, and multivariate thresholding. A broader survey is provided by Trapani and Longo (2023), who group existing methods into categories such as machine learning, statistical techniques, PCA, Kalman filtering, SVM, digital twins, Markov models, fault-tree analysis, and others.

The low temporal resolution of SCADA data, combined with the often prolonged persistence of sensor faults (see the Methodology section), renders sensor errors a particularly critical issue for failure prediction in wind turbines using SCADA-based NBMs. Sensor faults in SCADA systems may lead to extended periods of unreliable or unusable data. Consequently, such faults can result in months of invalid NBM predictions, substantially reducing the practical value of the analysis. For this reason, the development of sensor-error-robust NBM models is of significant importance.

2.3 Masked autoencoder

Masked machine learning models form a prominent class of self-supervised learners. Their initial development was driven by natural language processing (NLP), with Bidirectional Encoder Representations from Transformers (BERT) being a well-known example. BERT employs a Masked Language Modeling (MLM) objective, in which portions of the input sequence are masked and subsequently predicted during training (Devlin et al., 2019).

MAEs have since been adopted in computer vision. For instance, He et al. (2021) introduce a MAE method for image data based on an asymmetric encoder–decoder architecture. This design differs from ours, which employs a strictly symmetric architecture. Their approach also uses a substantially higher masking ratio (75%), whereas the ratio in our study is considerably lower. Additionally, the data modality (images versus the tabular data used here) introduces fundamental differences in model

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

structure and learning dynamics. Several other works have extended MAE to computer vision tasks (Chen et al., 2024; Madan et al., 2024).

In contrast to vision-focused research, our work applies MAE to tabular sensor data, which requires important algorithmic modifications. Several recent papers have investigated this direction. Du et al. (2023) use MAE for missing-value imputation in tabular datasets. Similarly, Kim et al. (2025) develop a proportional masking strategy designed to compensate for biases that arise with naive random masking, and demonstrate that Multi-Layer Perceptron (MLP)-Mixed models often outperform transformer-based alternatives for tabular data. A general masked encoding methodology for tabular data is presented in Majmundar et al. (2022), with further specialization for time-series applications explored by Li et al. (2023), who use masking to learn stronger temporal representations.

To our knowledge, the methodology most closely related to the one proposed in this paper is that of Fu and Yan (2024), who apply a MAE to streamline the fault detection, isolation, and accommodation (FDIA) pipeline for offshore wind turbines. Despite the conceptual similarity, several important differences remain. First, their AE topology does not include a bottleneck layer, whereas our method explicitly enforces one, and they do not incorporate a feature-expansion layer. Their training noise level is also substantially lower (maximum 0.2) compared with ours (0.5). In terms of input structure, they use only masked SCADA data, while we provide both masked SCADA data and an explicit masking matrix. Their validation focuses on blade bending moments with simulated sensor faults, whereas our study evaluates generator and gearbox temperature sensors that exhibit real sensor failures. Furthermore, we additionally assess the ability of the model to detect wind-turbine component failures, while Fu and Yan (2024) concentrate solely on sensor-fault detection. As a result, the methodology developed in this paper is tailored to component-failure prediction under realistic sensor-fault conditions, aligning more closely with operational needs.

3 Methodology

3.1 Input data

This study uses 10-minute SCADA data from four operational offshore wind farms (WF1–WF4). The dataset spans three wind turbine types with rated capacities between 3 megawatt (MW) and 10 MW. Across the four wind farms, 136 turbines are included, representing an aggregated observation period of approximately 1,000 turbine-years. The SCADA data provide measurements of component temperatures as well as environmental and operational conditions. Only a subset of these signals is used in this work. Table 1 provides an aggregated overview, with detailed signal lists presented in the Tables A1, A2, A3, A4. The primary focus is on detecting abnormal behavior in temperature-related signals that may serve as precursors to component failures.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Signal type	WF1	WF2	WF3	WF4
generator bearing temp.	2	2	2	2
generator phase temp.	6	6	6	3
hydraulic oil temp.	1	1	0	0
gearbox main bearing temp.	2	2	2	0
gearbox oil temp.	2	5	5	2
gearbox bearing temp.	4	4	13	5
generator lubrication temp.	0	1	1	0
generator air outlet temp.	0	2	0	0
cooling water temp.	0	3	1	1
generator slipring temp.	0	0	0	1
nacelle temp.	1	1	1	1
Operation signals	6	7	6	5
Environment signals	5	5	5	5

Table 1. High-level overview of signals used by the NBM for the different wind farms.

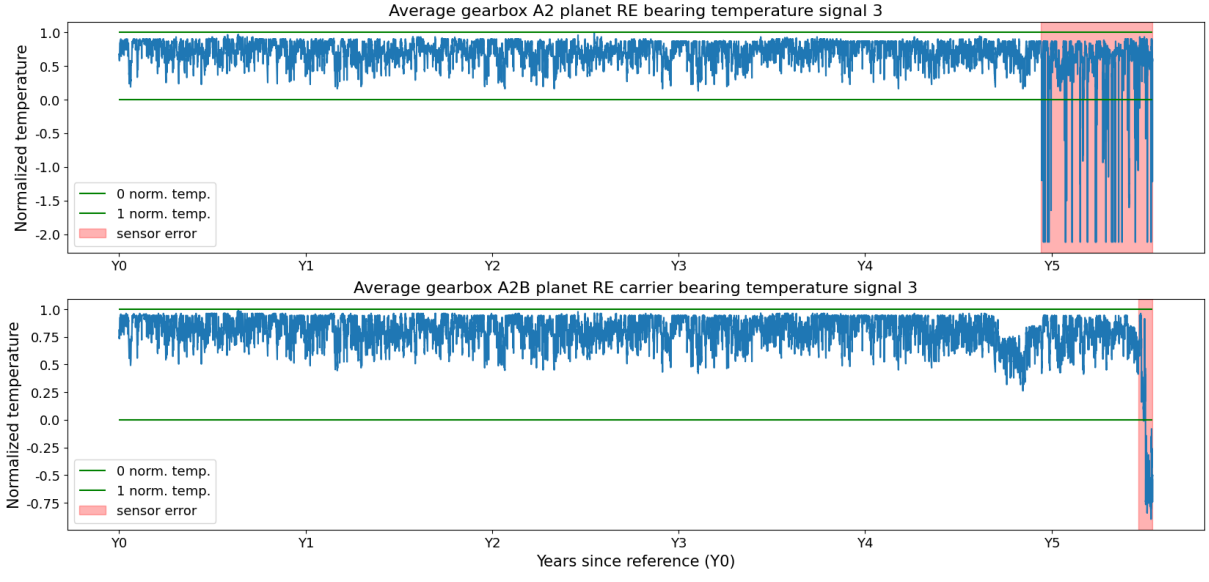


Figure 1. Sensor failures in two signals of turbine in WF1. The sensor failure is a combination of types, i.e. bias and scaling.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

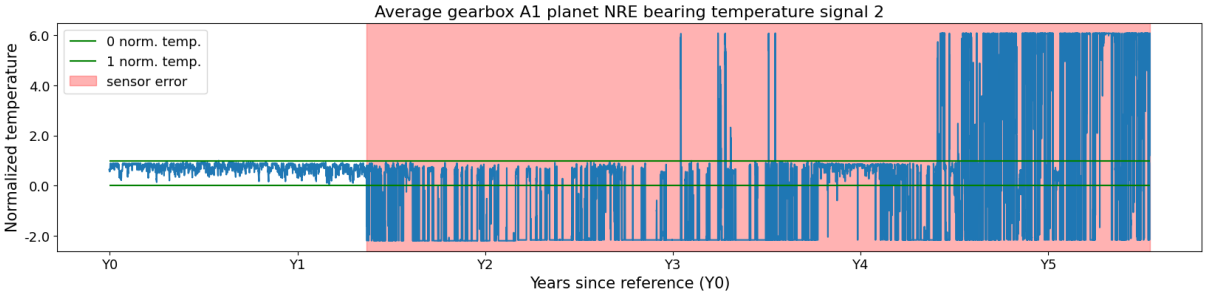


Figure 2. Sensor failure in temperature signal of turbine in WF3. The sensor failure is a combination of types, i.e. bias and scaling.

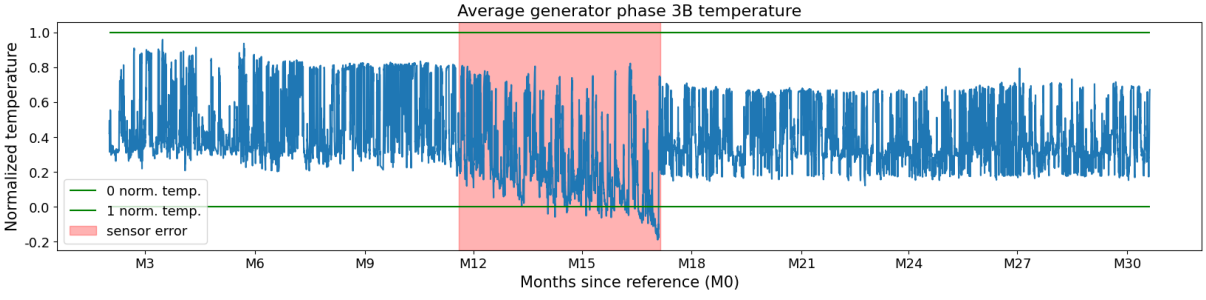


Figure 3. Sensor failure in temperature signal of turbine in WF2. The sensor failure is of the type drift.

3.2 Data preprocessing

Because anomaly detection and failure prediction for wind turbines must rely on operational SCADA data, which frequently contain gaps, sensor issues, and other imperfections, several preprocessing steps are required before model development. The following steps summarize the procedure adopted in this study:

- Data aggregation: The original 10-minute SCADA data are aggregated to a 1-hour resolution. This reduction in temporal resolution is acceptable because the degradation mechanisms of interest evolve over long time scales (days to months). Aggregation also reduces high-frequency noise, resulting in cleaner input for model training.
- Major failure identification: A major failure is defined as a turbine downtime exceeding seven days. Such prolonged outages typically indicate the failure of a major component requiring replacement. Given the size of these components and the logistical constraints associated with offshore maintenance, replacement can be time-consuming. Identifying major failures is essential, as post-replacement behavior (particularly temperature profiles) may differ from pre-failure behavior and may therefore require recalibration.
- Unhealthy data selection: SCADA data are unlabeled, and thus do not explicitly indicate whether the turbine is operating in a healthy state. For training a NBM, it is important to use data that reflect healthy operation as closely as possible.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

A pragmatic assumption is adopted: data occurring within 90 days prior to a major failure, or within 5 days after it, are considered unhealthy. All remaining data are treated as healthy.

- Train-validation-test split: To prevent overfitting and ensure generalizability, the dataset is partitioned into training, validation, and test sets. Training and validation data are drawn from a single 1-year window selected to include all seasons, as seasonal variability in SCADA signals tends to influence model output. Within this window, 80% of the observations are assigned to training and 20% to validation. All remaining data outside this 1-year window form the test set.
- Measurement error handling: Sensor failures occasionally occur in operational SCADA systems, producing unreliable measurements. To prevent such corrupt data from influencing model training, measurement errors are identified and removed (as well as possible) from the training and validation sets using expert-defined thresholds. These errors are not removed from the test set, as sensor-error robustness is an important practical requirement and a key objective of this study.
- Missing value handling: Missing values in SCADA data typically arise from sensor malfunctions or communication interruptions. Neither mechanism is expected to depend on the underlying operational state, suggesting that the missingness mechanism is likely Missing Completely at Random (MCAR). Under this assumption, the simplest and most transparent approach, i.e. removing observations containing at least one missing value, is adopted.
- Feature engineering: Several additional features are derived to enhance model performance. These include: (i) operating conditions following the IEC standard, (ii) wind vector components derived from wind speed and direction, (iii) nacelle–wind direction offset, which may indicate increased mechanical loading when large, (iv) blade-angle differences (when available), as large offsets may also correspond to increased drive-train stresses.

3.3 Normal behavior modeling

A wide range of techniques for anomaly detection and failure prediction have been proposed in the literature, including trending approaches, clustering methods, NBM, physics-based damage models, alarm-based systems, and expert-rule systems (Tautz-Weinert and Watson, 2017). In this work, we adopt the NBM approach, which is widely used and has demonstrated strong performance in industrial monitoring applications.

NBM consists of two steps. First, a model is trained exclusively on data representing healthy system behavior. When successful, the model learns the normal relationships among the variables. In the second step, the model is applied to new data to generate predictions that represent the expected (healthy) behavior. Deviations between the predictions and the measured values are interpreted as indicators of abnormal or degraded operation: small deviations correspond to nominal behavior, whereas large deviations suggest emerging faults.

In this study, a MAE is employed as the NBM. This architecture is specifically chosen for its ability to handle measurement errors and missing values caused by sensor failures. By incorporating an explicit mask, the AE can learn normal inter-signal relationships even when some sensor values are unavailable or unreliable.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

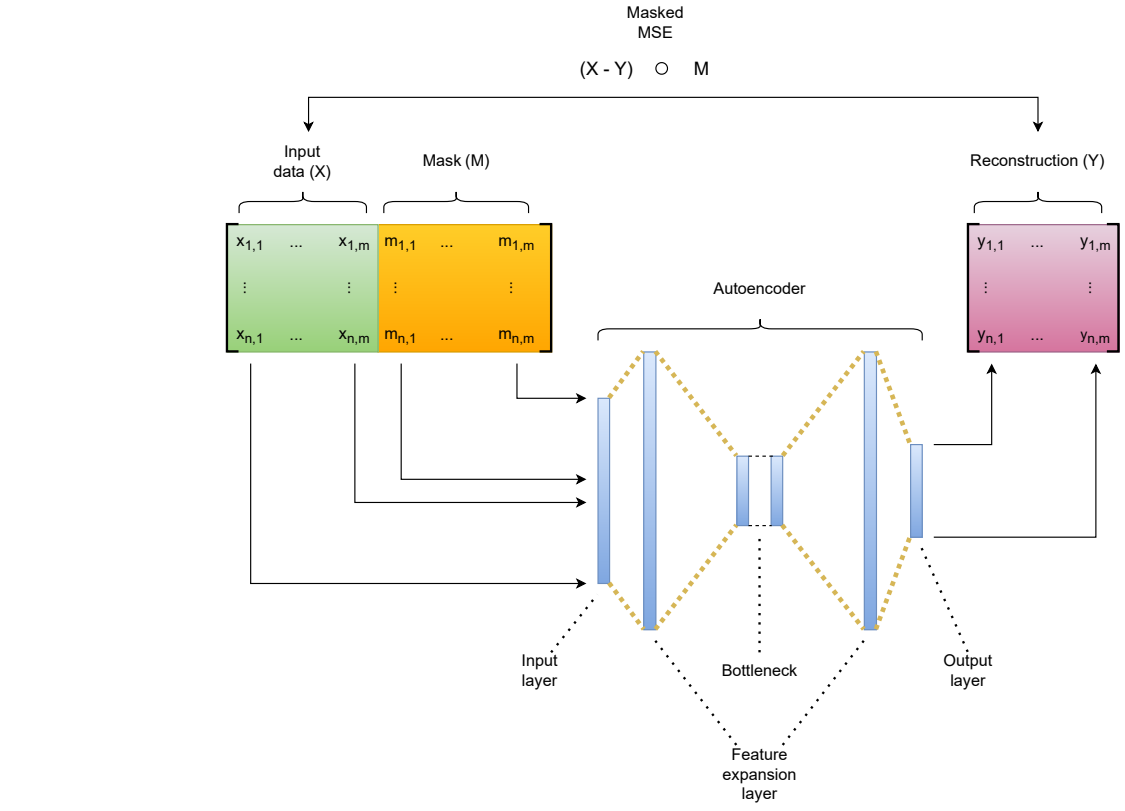


Figure 4. Schematic overview of the masked autoencoder methodology.

3.3.1 Masked autoencoder training and prediction

We develop a variant of a MAE tailored to the structure of SCADA data. Let the input matrix X contain n observations and m features, where $x_{i,j}$ denotes the value of feature j in observation i. Each value has an associated mask entry $m_{i,j} \in \{0,1\}$, forming the mask matrix M. 0 indicates that the value in X is masked and 1 otherwise. If a value in X is masked its value is changed to the masking value 0 ($m_{i,j} = 0 \implies x_{i,j} = 0$). Before being fed into the AE, X and M are concatenated, enabling the network to distinguish between true zeros and masked entries without requiring out-of-range mask values. The AE comprises of an encoder ($f_\theta : \mathbb{R}^{2m} \to \mathbb{R}^d$) and a decoder ($g_\phi : \mathbb{R}^d \to \mathbb{R}^m$), producing the reconstruction $Y = g_\phi(f_\theta(X, M))$. Training minimizes a masked means squared error (MSE) loss $l(Y, X, M) = ||(Y - X) \odot M||$, ensuring that only unmasked elements contribute to the loss.

Model architecture and training parameters are optimized using the Keras Tuner Hyperband algorithm, which allocates computational resources via a multi-armed bandit strategy (Li et al., 2018). Separate training and validation sets are used to avoid overfitting. Several architectural constraints are imposed to limit the hyperparameter search space:

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Farm	Model partition	Layers	Units per layer
WF1	Encoder	3	[68, 249, 24]
WF1	Decoder	3	[24, 249, 34]
WF2	Encoder	3	[84, 309, 30]
WF2	Decoder	3	[30, 309, 42]
WF3	Encoder	3	[88, 298, 31]
WF3	Decoder	3	[31, 298, 44]
WF4	Encoder	4	[56, 158, 54, 19]
WF4	Decoder	4	[19, 54, 158, 28]

Table 2. Overview of the number of layers and number of units per layer for the best model for WF1, WF2, WF3 and WF4

- Symmetric architecture: The encoder and decoder are mirror images, except that the encoder input layer has dimensionality $2m$ due to concatenation of X and M, while the decoder outputs m reconstructed features.
- Feature expansion layer: The first hidden layer of the encoder expands the dimensionality, with the number of units constrained to $[2.5m, 8m]$.
- Bottleneck layer: The latent dimension d is required to satisfy $d \in [0.25m, 0.75m]$ ensuring a compressed representation.
- Intermediate layers: Units in intermediate layers follow an exponential decay pattern determined by the expansion-layer width $2m$, the bottleneck width d, and the number of layers h between them: $r = (\frac{d}{2m})^{\frac{1}{h}}$.

Training is capped at 200 epochs with early stopping (patience = 10). For MAE, it is important that candidate models are allowed sufficient training time before elimination. Accordingly, the Hyperband reduction factor is set to 2, slowing down the elimination schedule, and the hyperband_iterations parameter is set to 3 to encourage adequate exploration.

For three of the four datasets, the best models selected through hyperparameter tuning are MAE architectures whose encoders and decoders each comprise three layers. For the WF4 dataset, both the encoder and decoder consist of four layers. In all cases, the bottleneck induces a substantial reduction in dimensionality. An overview of these architectures is provided in Table 2.

A single model provides only point estimates and does not quantify prediction variability. To obtain stability estimates, we apply bootstrapping (Efron, 1979). The training dataset is resampled with replacement 200 times, generating 200 datasets of equal size. Because the autoencoder relies on masking during training, a new random mask with noise ratio 0.5 is generated for each resampled dataset.

For each bootstrap sample, a model with the best-performing hyperparameters is trained from scratch with randomized weights, again for up to 200 epochs with early stopping. This procedure yields 200 independently trained AEs. During prediction, each model produces its own reconstruction errors, resulting in an ensemble of 200 reconstruction-error time series that collectively characterize both expected behavior and predictive uncertainty.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

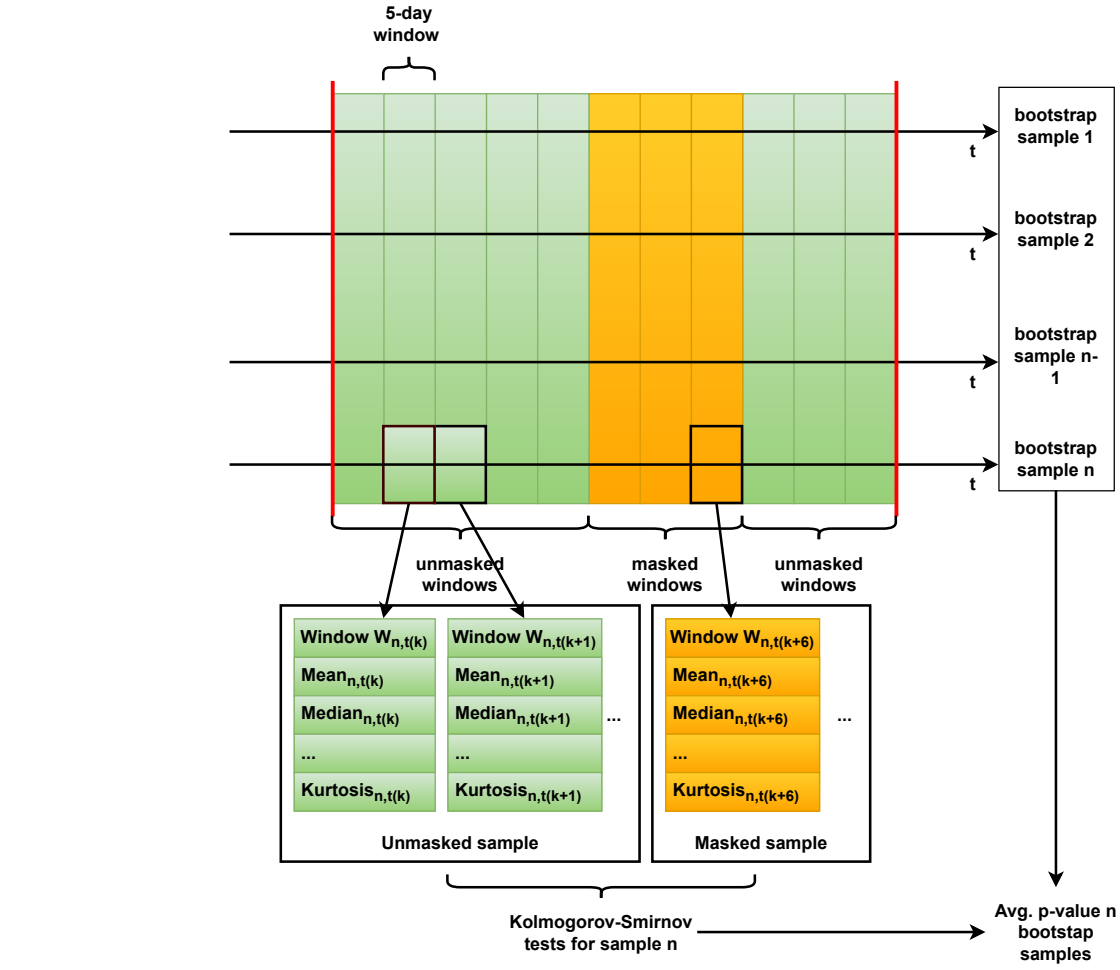


Figure 5. Schematic overview of the methodology used for determining the similarity between the reconstruction error distributions of masked data and nominal data.

3.3.2 Evaluation of the sensor error solution

The bootstrapped models are used to generate predictions on the test data in two ways: first without masking measurement errors, and then with masking applied. This enables an assessment of the impact of masking. When masking is applied, the mask for the affected signal(s) in the relevant time window is set to 0, and the corresponding signal values are set to 0 as well. This reflects a realistic operational scenario. The procedure is carried out for all sensor error cases described in the Input data section. Two aspects of the methodology are evaluated: (i) the extent to which masking corrects the prediction artefacts caused by sensor failures, and (ii) the ability of the MAE to identify an approaching failure.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

The first aspect is assessed by comparing the reconstruction errors of each signal inside a masked window to those outside. If the MAE functions correctly, these reconstruction errors should be statistically similar. The comparison is based on several distributional properties: mean, median, standard deviation, interquartile range, skewness, and kurtosis. To ensure fair comparison, only data from the same inter-failure data window (IFDW), which is a continuous interval between two consecutive major component failures/replacements or between turbine commissioning and the first major failure, is used for each sensor error case. Since major replacements can substantially alter the relationships between signals, comparing reconstruction errors across different IFDWs would lead to misleading conclusions.

The IFDW data are segmented into blocks of five days, and the metrics listed above are computed for each block. Blocks are then labeled as masked or unmasked. To determine whether the distributions differ between the two groups, the two-sided Kolmogorov–Smirnov test (95% significance level) is applied. For each signal and each bootstrapped model, one p-value is obtained, yielding 200 p-values per signal (one for each resampled dataset); their mean is then computed. The MAE is considered to perform adequately if, for all signals except the one affected by the sensor error, the average p-value exceeds 0.025. Because the validation is conducted in operational rather than simulated or laboratory settings, uncontrolled variables may introduce unforeseen effects. Nevertheless, aside from minor discrepancies, the results for most signals should be robust and unambiguous.

The second aspect is evaluated by examining how well the model differentiates between healthy and unhealthy data using the abnormal behavior sensitivity metric (ABSM). This is also based on IFDWs. This ensures that healthy and unhealthy data are both generated under consistent component configurations. Unhealthy data are defined as observations within 90 days preceding a major failure; healthy data are defined as the first 30 days of the corresponding IFDW. For each signal, the ratio is computed between (i) the fraction of unhealthy observations for which the 2.5th percentile of the reconstruction error exceeds zero and (ii) the corresponding fraction for healthy observations.

Since the failure logs specify only whether a failure concerns the generator or the gearbox, the results are aggregated per component by taking the maximum ratio among all signals associated with that component. This approach reflects practical usage, where an analyst inspects multiple signals and determines which component shows the strongest deviation. For the method to be useful as an anomaly detector, this ratio should be substantially greater than 1, indicating clear differentiation between healthy and unhealthy behavior. A ratio above 2 is considered strong evidence, values between 1.25 and 2 indicate marginal identification, and values below 1.25 are classified as failures to identify.

3.4 Sensor errors

The input data contain numerous sensor failures that manifest as measurement errors. For validation, 22 representative cases were selected from WF1, WF2, and WF3. These cases include several of the failure modes discussed in the Related work section. Some sensor errors are readily identifiable because they produce physically impossible values, as illustrated in Figures 1 and 2. These typically correspond to biased sensors or hard faults. Other errors are more subtle, arising from gradual sensor drift. Such drift appears as a slow upward or downward trend and may only produce implausible values after an extended

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

period, if at all (see Figure 3). These cases are more challenging to detect, especially because temperature-related signals may also deviate due to legitimate causes such as component wear or damage.

To automate the detection of sensor errors, the MAE is first applied in a non-masking mode, meaning that no observations or signals are switched off. In this configuration, sensor failures manifest as characteristic anomalies in the reconstruction errors. In some instances, the failures appear as isolated outliers, but more commonly they result in prolonged periods during which the reconstruction error deviates consistently from the expected pattern, reflecting biased readings or hard faults, as visible in Figures 1, 2, and 3. These extended error periods arise because sensor issues are often not noticed immediately, and even once identified, repairs or replacements typically occur only during scheduled maintenance. In some cases, the sensor errors persist for several years before the issue is resolved (see Figure 2). These patterns are identified and logged (i.e. turbine, signal, start and end). The log is then used to switch off signals during the run in masking mode.

The sensor errors occur across a broad range of sensors, including hydraulic oil temperature sensors; gearbox oil sump temperature sensors; gearbox A2B planet non–rotor-end (NRE) carrier bearing temperature sensors; gearbox A2B planet rotor-end (RE) carrier bearing temperature sensors; generator phase 3B temperature sensors; generator air outlet 1 temperature sensors; gearbox A1 planet NRE bearing temperature sensors; gearbox A2 planet RE bearing temperature sensors; gearbox B planet NRE carrier bearing temperature sensors; the nacelle temperature sensor; and generator slip-ring temperature sensors.

4 Results

4.1 Distributional similarity of the reconstruction errors under masked sensor error and nominal conditions

The first question examined in this study is whether the reconstruction errors produced by the MAE when at least one sensor has failed remain comparable to those obtained under fully healthy sensing conditions. Establishing this equivalence is critical: if reconstruction errors differ substantially between these two regimes, it would indicate that the AE has not adequately learned the underlying inter-sensor relationships in the presence of missing or corrupted inputs. In such a case, the reconstruction errors would become unreliable for anomaly detection precisely when sensor faults occur, thereby undermining the purpose of the method.

Figures 6 and 7 provide qualitative insight into the behavior of the algorithm. In each figure, the interval affected by the sensor fault is highlighted in green. In Figure 6, the failure occurs in the gearbox oil sump temperature sensor (bottom panel). When the failure is not masked, the corrupted measurement propagates through the learned correlations and induces distortions in the reconstruction errors of other signals, for example, the temperature measured at the gearbox rotor end carrier bearing of planetary stage A1. These distortions render the anomaly detections for these signals unreliable. Specifically, the failed sensor exhibits a sudden downward bias, which in turn produces an inverted (upward) bias in the reconstructed estimates of correlated signals. Because the AE models the normal relationships among these variables, an anomalously high value in one channel prompts overestimation of the corresponding “normal” values of others, leading to negatively biased reconstruction errors.

An analogous but reversed pattern is observed in Figure 7. Here, the affected sensor shows a pronounced downward bias during the failure window, resulting in upward-biased reconstruction errors in several correlated signals, again exemplified by

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

Figure 6. Sensor failure (bottom plot) and its impact (top plot) on the reconstruction error of a different signal without (orange curve) and with (blue curve) masking for turbine 1 of WF1.

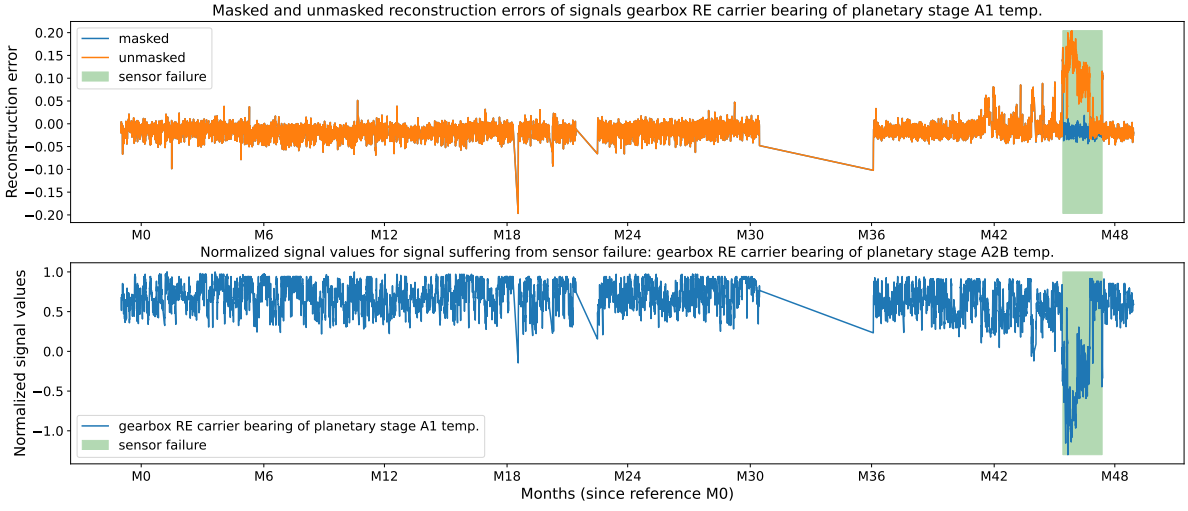
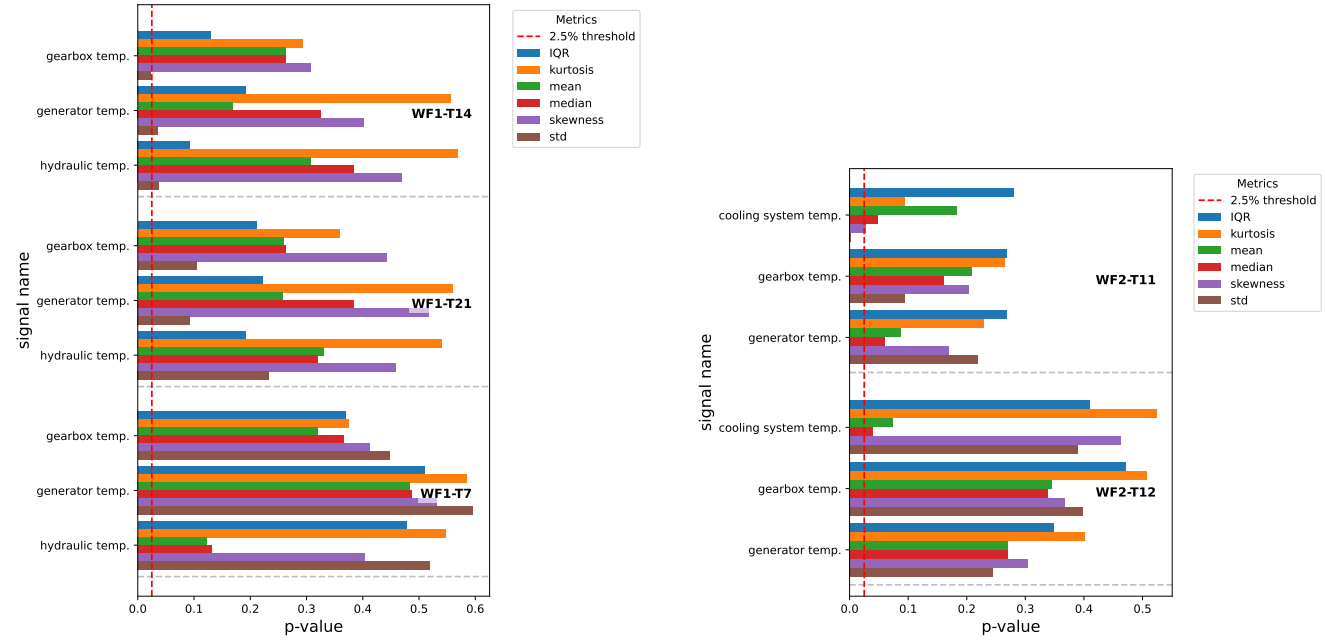


Figure 7. Sensor failure (bottom plot) and its impact (top plot) on the reconstruction error of a different signal without (orange curve) and with (blue curve) masking for turbine 21 of WF1.

the gearbox rotor end carrier bearing temperature (top panel, orange trace). When the faulty sensor is masked, this spurious upward bias is eliminated, and the reconstruction errors align with the surrounding baseline values.

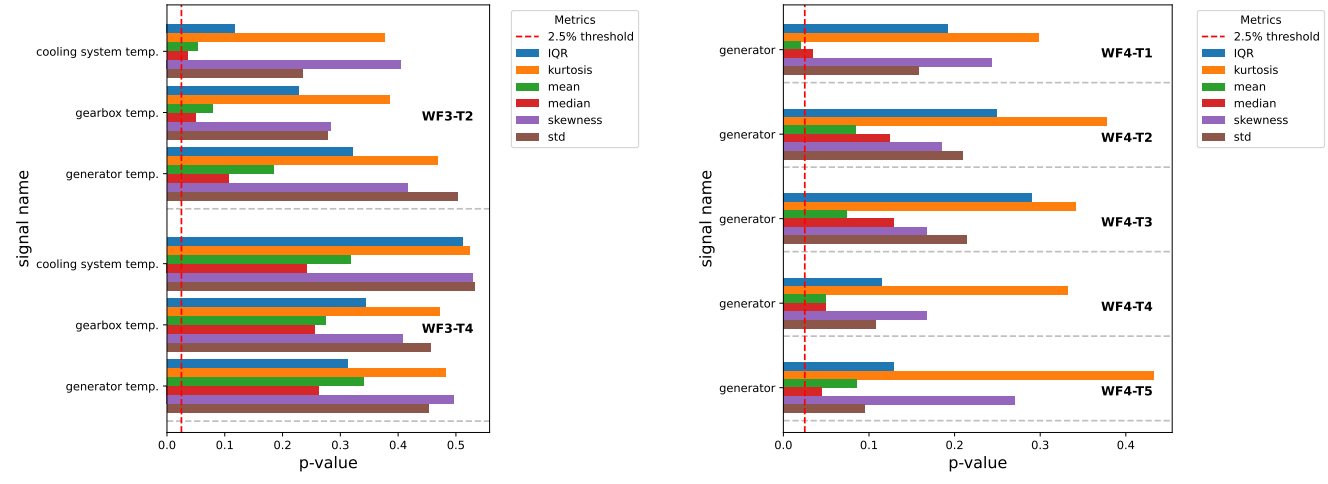
https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025



(a) Wind farm 1

(b) Wind farm 2

Figure 8. Mean of Kolmogorov-Smirnov p-values for several wind turbine components.



(a) Wind farm 3

(b) Wind farm 4

Figure 9. Mean of Kolmogorov-Smirnov p-values for several wind turbine components.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Because of the large size of the validation dataset, exhaustive visual inspection is not feasible. Therefore, we employ an automated and formal evaluation based on several statistical metrics (see Methodology section). Figures 8a–9b present the average p-values of the Kolmogorov–Smirnov (KS) tests for different subsystems, i.e. generator, gearbox, cooling system, and hydraulics, and for multiple statistical descriptors of the reconstruction error distributions (i.e. mean, median, standard deviation, interquartile range, skewness, and kurtosis) across the four wind farms. In general, the average p-values exceed the threshold of 0.025, indicating that the null hypothesis (i.e., no difference between the reconstruction error distributions with and without the sensor-error mask) cannot be rejected. Overall, this suggests that there is insufficient statistical evidence to conclude that reconstruction errors differ between masked and fully nominal conditions.

In some cases, however, p-values approach or fall below the 0.025 threshold. This is expected given that the data originate from operational wind farms rather than controlled laboratory settings, where unforeseen operational events may influence the measurements. A second, farm-specific explanation applies to wind farm 4, which exhibits particularly complex sensor-error scenarios. In this farm, all gearbox-related signals are missing over an extended period of more than two years, likely due to a persistent communication failure, although the root cause cannot be confirmed. Consequently, only generator-related results are presented in Figure 9b. For turbine WF4-T1, the p-value for the mean is below the significance threshold, and the p-value for the median is close to it. The overall average p-values are also somewhat lower than for the other turbines. This reduction can be attributed to the difficulty of the case: the model must operate with approximately half of its input signals missing. Despite this challenging condition, the AE still produces reconstruction errors for the generator signals that remain broadly representative, although with reduced robustness.

4.2 Discriminative performance of model in distinguishing healthy and unhealthy data

For failure prediction, it is not sufficient that the model effectively masks sensor faults; it must also reliably distinguish between healthy and degraded operating conditions. Evaluating this capability is considerably more challenging than assessing sensor-error robustness. Real-world data often contain uncertainties regarding the precise failure modes and root causes, as well as additional unrecorded events that may influence the signals. Through detailed data inspection and expert consultation, the failure modes were identified as accurately as possible, and several confounding cases were removed. Nonetheless, some unknown factors may still affect the results.

Figures 10 and 11 illustrate examples of correct detections. In Figure 10, a gradual degradation trend is visible, with reconstruction errors increasing over approximately 18 months prior to the failure, which occurred at year 6.25. In Figure 11, temperatures in the gearbox rotor end main bearing begin to rise around year 5.5, followed by failure at year 5.75. The detection lead time varies substantially across failures, influenced by many factors that are not always observable. In both of these cases, the model identifies the degradation well in advance. The figure also includes several long-duration downtimes unrelated to the gearbox; these are not detected when using gearbox signals, as the associated failure modes are not gearbox-related.

To obtain a general assessment of the value of the model as NBM, we validate it on historical failure data from operational offshore wind farms. For three of the four wind farms described in the Methodology section, industrial partners provided lists containing the turbine ID, timestamp, and failure type. Validation is performed using the ABSM metric (see Methodology

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

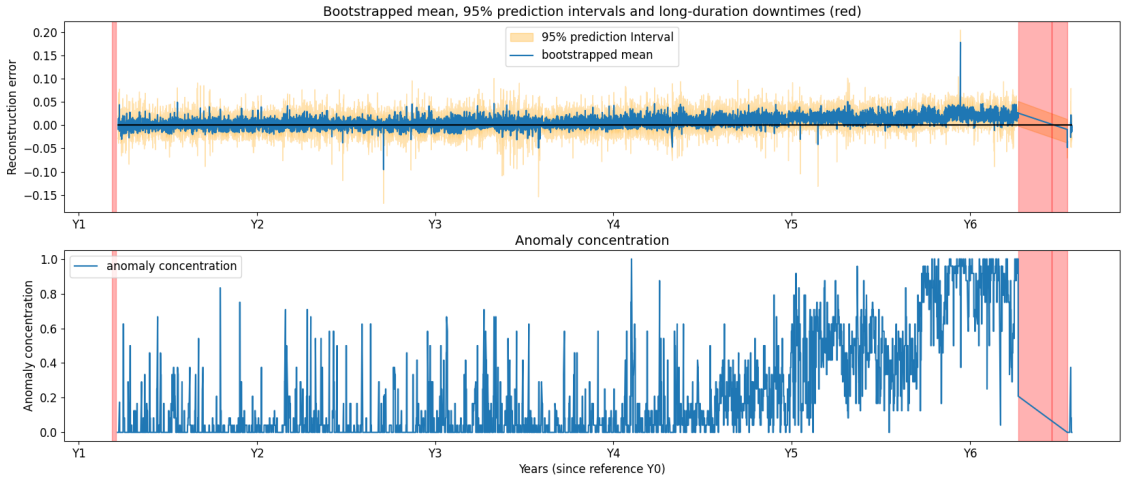


Figure 10. Generator failure for turbine 22 of wind farm 1. The top plot shows the reconstruction error for the generator phase 2 temperature. The bottom plot shows the anomaly concentration score.

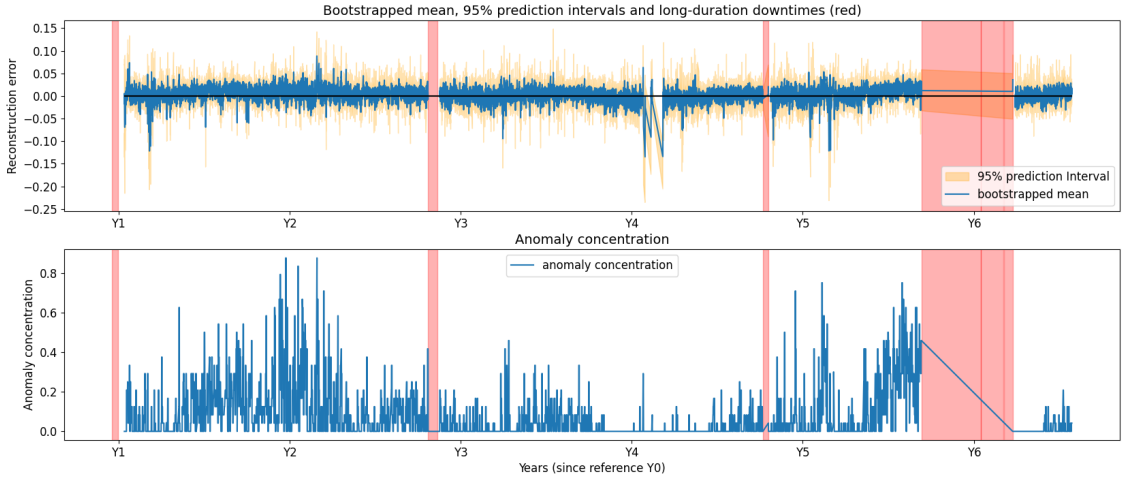


Figure 11. Gearbox failure for turbine 5 of wind farm 1. The top plot shows the reconstruction error for the generator phase 2 temperature. The bottom plot shows the anomaly concentration score.

section). A strong NBM should yield ABSM values substantially greater than 1 for the component that ultimately fails. In total, 22 drive-train-related failures were evaluated: nine gearbox failures and thirteen generator failures. The results for wind farm 1 and for wind farms 2 and 3 are presented in Tables 3 and 4, respectively.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Farm	Turbine	Failure	ABSM
1	1	gearbox	2.90
1	2	gearbox	4.70
1	3	generator	43.59
1	8	gearbox	4.94
1	11	generator	1.44
1	13	generator	13.26
1	16	gearbox	1.49
1	17	gearbox	10.54
1	17	generator	3.85
1	18	generator	0.98
1	23	generator	1.05

Table 3. ABSM results for wind farm 1.

Farm	Turbine	Failure	ABSM
3	8	generator	1.44
3	9	generator	2.41
3	12	generator	1.88
3	19	gearbox	37.13
4	6	gearbox	1.91
4	12	generator	2.39
4	18	generator	3.46
4	30	gearbox	3.10
4	35	gearbox	20.10
4	41	generator	9.46
4	52	generator	2.19

Table 4. ABSM results for wind farms 3 and 4.

Detection performance is categorized as follows: ABSM > 2 indicates a strong detection, $1.25 <$ ABSM ≤ 2 a marginal detection, and ABSM ≤ 1.25 a missed detection. Across all 22 failures, 15 (68.18%) are strong detections, 5 (22.72%) are marginal detections, and 2 (9%) are missed detections, yielding an overall detection rate of 20 out of 22 cases (90.91%). When split by failure type, the detection rate is 11 of 13 (84.62%) for generator failures and 9 of 9 (100%) for gearbox failures.

Overall, the model demonstrates strong discriminative ability. It reliably separates healthy from unhealthy behavior in most cases, with performance somewhat higher for gearbox failures than for generator failures. This difference is expected: several generator failures, particularly in wind farms 1 and 3, stem from electrical issues that are only indirectly reflected in generator-phase temperatures, making early detection more difficult. The two missed detections correspond to generator failures in wind farm 1.

5 Conclusions

This paper presented a sensor-error-robust methodology for failure prediction in wind turbine drive-train components. Because sensors inevitably degrade or fail during the operational lifetime of a turbine (and because modern condition-monitoring systems rely on large numbers of sensors) the likelihood of encountering faulty measurements is high. The validation data examined in this work contained a substantial number of sensor failures, many of which persisted for extended periods. If left untreated, such failures can significantly distort reconstruction errors and degrade anomaly-detection performance. Several examples in this study illustrate how unmasked sensor faults propagate bias into the reconstruction errors of correlated signals, underscoring the necessity of a method that remains reliable in the presence of sensor errors.

The MAE proposed here enables the selective exclusion of corrupted signals during prediction by leveraging a mask matrix that identifies missing or unreliable inputs. This robustness was achieved by training the model on data in which 50% of

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

the inputs were randomly masked, allowing it to learn inter-signal relationships even under substantial data loss. The results demonstrate that the model successfully maintains consistent reconstruction-error distributions when one or more sensors fail. Even in the extreme scenario where all gearbox signals were absent for more than two years, the model produced generator-signal reconstruction errors of acceptable quality, although with some signs of performance degradation.

In addition, the MAE proved capable of reliably distinguishing healthy from degraded operating conditions, even when sensor errors were present. This indicates that the approach is suitable as a NBM. Validation on three operational offshore wind farms, covering both generator and gearbox failures, yielded an overall detection rate of 90.91%. As expected, detection performance was slightly lower for electrically driven generator failures, which manifest only indirectly in temperature signals.

Several avenues for future work remain. First, the influence of the masking (noise) ratio during training warrants further investigation, as previous studies report effective ratios both lower and higher than the 50% used here. Exploring this parameter may yield improvements in robustness or detection sensitivity. Second, a more detailed assessment of the stability of the model and uncertainty is desirable. While the current methodology incorporates bootstrapping to characterize stability, further work is needed to translate this stability into calibrated uncertainty estimates and to determine whether additional recalibration methods are required. Finally, evaluating the method on different turbine types and a broader range of failure modes would help establish its generalizability.

Appendix A

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Signal name	Component
generator bearing 1 temp.	generator
generator bearing 2 temp.	generator
generator phase 1 temp.	generator
generator phase 2 temp.	generator
generator phase 3 temp.	generator
hydraulic oil temp.	hydraulic
gearbox main NRE bearing temp.	gearbox
gearbox main RE bearing temp.	gearbox
gearbox oil inlet temp.	gearbox
gearbox oil sump temp.	gearbox
generator phase 1B temp.	generator
generator phase 2B temp.	generator
generator phase 3B temp.	generator
gearbox NRE carrier bearing of B planetary stage temp.	gearbox
gearbox NRE carrier bearing of A2B planetary stage temp.	gearbox
gearbox RE carrier bearing of A2B planetary stage temp.	gearbox
gearbox RE carrier bearing of A1 planetary stage temp.	gearbox

Table A1. Component temperature signals for turbines in WF1

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Signal name	Component
generator bearing 1 temp.	generator
generator phase 1 temp.	generator
generator phase 2 temp.	generator
generator phase 3 temp.	generator
generator bearing 2 temp.	generator
hydraulic oil temp.	hydraulic
gearbox main NRE bearing temp.	gearbox
gearbox main RE bearing temp.	gearbox
gearbox oil inlet temp.	gearbox
gearbox main tank oil temp.	gearbox
gearbox main bearing distributor oil temp.	gearbox
generator lubrication intlet temp.	generator
generator phase 1B temp.	generator
generator phase 2B temp.	generator
generator phase 3B temp.	generator
power train hydraulic water temp.	cooling system
converter generator water temp.	cooling system
water temp.	cooling system
gearbox LSS oil sump temp.	gearbox
gearbox RE planet carrier of LSS temp.	gearbox
gearbox NRE planet carrier of IMS temp.	gearbox
gearbox NRE planet carrier of LSS temp.	gearbox
gearbox IMS oil sump temp.	gearbox
gearbox thrust NRE bearing of IMS temp.	gearbox
generator air outlet 1 temp.	generator
generator air outlet 2 temp.	generator

Table A2. Component temperature signals for turbines in WF2

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Signal name	Component
gearbox oil inlet temp.	gearbox
gearbox main NRE bearing temp.	gearbox
gearbox main RE bearing temp.	gearbox
gearbox main tank oil temp.	gearbox
gearbox main bearing distributor oil temp.	gearbox
gearbox A1A2 oil sump temp.	gearbox
gearbox oil sump temp.	gearbox
gearbox RE carrier bearing of A1 planetary stage temp.	gearbox
gearbox NRE carrier bearing of A1 planetary stage sensor 1 temp.	gearbox
gearbox NRE carrier bearing of A1 planetary stage sensor 2 temp.	gearbox
gearbox NRE carrier bearing of A1 planetary stage sensor 3 temp.	gearbox
gearbox RE carrier bearing of A2 planetary stage sensor 1 temp.	gearbox
gearbox RE carrier bearing of A2 planetary stage sensor 2 temp.	gearbox
gearbox RE carrier bearing of A2 planetary stage sensor 3 temp.	gearbox
gearbox NRE bearing of A2 planetary stage sensor 3 temp.	gearbox
gearbox NRE bearing of A2 planetary stage sensor 2 temp.	gearbox
gearbox NRE bearing of A2 planetary stage sensor 1 temp.	gearbox
gearbox NRE carrier bearing of A2B planetary stage temp.	gearbox
gearbox RE carrier bearing of A2B planetary stage temp.	gearbox
gearbox NRE carrier bearing of B planetary stage temp.	gearbox
generator lubrication inlet temp.	generator
generator bearing 1 temp.	generator
generator bearing 2 temp.	generator
generator phase 1 temp.	generator
generator phase 2 temp.	generator
generator phase 3 temp.	generator
generator phase 1B temp.	generator
generator phase 2B temp.	generator
generator phase 3B temp.	generator
converter and generator water temp.	cooling system

Table A3. Component temperature signals for turbines in WF3

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Signal name	Component
generator bearing 1 temp.	generator
generator phase 1 temp.	generator
generator phase 2 temp.	generator
generator phase 3 temp.	generator
generator slipring temp.	generator
generator bearing 2 temp.	generator
gearbox oil basis temp.	gearbox
gearbox oil level 1 temp.	gearbox
gearbox high speed rotor end bearing temp.	gearbox
gearbox high speed generator end bearing temp.	gearbox
gearbox high speed middle bearing temp.	gearbox
gearbox hollow shaft rotor end bearing temp.	gearbox
gearbox hollow shaft generator end bearing temp.	gearbox
generator cooling water temp.	generator

Table A4. Component temperature signals for turbines in WF4

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Author contributions. XC did the conceptualization; XC developed the methodology, did the data curation and the formal analysis, and wrote the paper; XC, JH reviewed and edited the paper; JH, AN did the funding acquisition.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Xavier Chesterman, Ann Nowé, and Jan Helsen received funding from the Flemish Government (AI Research Program). The research presented in this paper is partly financed by VLAIO/De Blauwe Cluster through ICON project Supersized 5.0 and SBO project Core.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

References

Balaban, E., Saxena, A., Bansal, P., Goebel, K. F., and Curran, S.: Modeling, Detection, and Disambiguation of Sensor Faults for Aerospace Applications, IEEE Sens. J., 9, 1907–1917, https://doi.org/10.1109/JSEN.2009.2030284, 2009.

Bermúdez, K., Ortiz-Holguin, E., Tutivén, C., Vidal, Y., and Benalcázar-Parra, C.: Wind Turbine Main Bearing Failure Prediction using a Hybrid Neural Network, J. Phys.: Conf. Ser., 2265, 032 090, https://doi.org/10.1088/1742-6596/2265/3/032090, 2022.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML: Evaluation of measurement data — Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology, JCGM 100:2008, https://doi.org/10.59161/JCGM100-2008E, 2008.

Black, I. M., Richmond, M., and Kolios, A.: Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management, Int. J. Sustain. Energy, 40, 923–946, https://doi.org/10.1080/14786451.2021.1890736, 2021.

Campoverde, L., Tutivén, C., Vidal, Y., and Benaláazar-Parra, C.: SCADA Data-Driven Wind Turbine Main Bearing Fault Prognosis Based on Principal Component Analysis, J. Phys.: Conf. Ser., 2265, 032 107, https://doi.org/10.1088/1742-6596/2265/3/032107, 2022.

Castellani, F., Astolfi, D., and Natili, F.: SCADA Data Analysis Methods for Diagnosis of Electrical Faults to Wind Turbine Generators, Appl. Sci., 11, 1–14, https://doi.org/10.3390/app11083307, 2021.

Chatterjee, J. and Dethlefs, N.: Scientometric review of artificial intelligence for operations maintenance of wind turbines: The past, present and future, Renew. Sustain. Energy Rev., 144, 111 051, https://doi.org/10.1016/j.rser.2021.111051, 2021.

Chen, H., Zhang, W., Wang, Y., and Yang, X.: Improving Masked Autoencoders by Learning Where to Mask, arXiv, https://doi.org/10.48550/arXiv.2303.06583, 2024.

Chesterman, X., Verstraeten, T., Daems, P.-J., Sanjines, F. P., Nowé, A., and Helsen, J.: The detection of generator bearing failures on wind turbines using machine learning based anomaly detection, J. Phys.: Conf. Ser., 2265, 032 066, https://doi.org/10.1088/1742-6596/2265/3/032066, 2022.

Chesterman, X., Verstraeten, T., Daems, P.-J., Nowé, A., and Helsen, J.: Overview of normal behavior modeling approaches for SCADA-based wind turbine condition monitoring demonstrated on data from operational wind farms, Wind Energy Sci., 8, 893–924, https://doi.org/10.5194/wes-8-893-2023, 2023.

Costanzo, G., Brindley, B., and Tardieu, P.: Wind energy in Europe: 2024 Statistics and the outlook for 2025-2030, Wind Europe, Brussels, Belgium, 2025.

Dao, P. B.: On Cointegration Analysis for Condition Monitoring and Fault Detection of Wind Turbines Using SCADA Data, Energies, 16, 2352, https://doi.org/10.3390/en16052352, 2023.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL-HLT, 1, https://doi.org/10.18653/v1/N19-1423, 2019.

Du, T., Melis, L., and Wang, T.: ReMasker: Imputing Tabular Data with Masked Autoencoding, arXiv, https://doi.org/10.48550/arXiv.2309.13793, 2023.

Efron, B.: Bootstrap Methods: Another Look at the Jackknife, Ann. Statist., 7, 1–26, https://doi.org/10.1214/aos/1176344552, 1979.

Fu, Y. and Yan, W.: One Masked Model is All You Need for Sensor Fault Detection, Isolation and Accommodation, arXiv, https://doi.org/10.48550/arXiv.2403.16153, 2024.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R.: Masked Autoencoders Are Scalable Vision Learners, arXiv, https://doi.org/10.48550/arXiv.2111.06377, 2021.

https://doi.org/10.5194/wes-2025-280
Preprint. Discussion started: 30 December 2025

Kestel, K., Chesterman, X., Zappalá, D., Watson, S., Li, M., Hart, E., Carroll, J., Vidal, Y., Nejad, A. R., Sheng, S., Guo, Y., Stammler, M., Wirsing, F., Saleh, A., Gregarek, N., Baszenski, T., Decker, T., Knops, M., Jacobs, G., Lehmann, B., König, F., Pereira, I., Daems, P.-J., Peeters, C., and Helsen, J.: Condition monitoring of wind turbine drivetrains: State-of-the-art technologies, recent trends, and future outlook, WESD, 2025, 1–85, https://doi.org/10.5194/wes-2025-168, 2025.

Kim, J., Lee, K., and Park, T.: To predict or not to predict? proportionally masked autoencoders for tabular data imputation, in: Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI'25/IAAI'25/EAAI'25, AAAI Press, ISBN 978-1-57735-897-8, https://doi.org/10.1609/aaai.v39i17.33967, 2025.

Lee, Y., Park, C., Kim, N., Ahn, J., and Jeong, J.: LSTM-Autoencoder Based Anomaly Detection Using Vibration Data of Wind Turbines, Sensors, 24, 2833, https://doi.org/10.3390/s24092833, 2024.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A.: Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization, J. Mach. Learn. Res., 18, 1–52, https://doi.org/10.5555/3122009.3242042, 2018.

Li, Z., Rao, Z., Pan, L., Wang, P., and Xu, Z.: Ti-MAE: Self-Supervised Masked Time Series Autoencoders, arXiv, https://doi.org/10.48550/arXiv.2301.08871, 2023.

Madan, N., Ristea, N.-C., Nasrollahi, K., Moeslund, T. B., and Ionescu, R. T.: CL-MAE: Curriculum-Learned Masked Autoencoders, arXiv, https://doi.org/10.48550/arXiv.2308.16572, 2024.

Majmundar, K., Goyal, S., Netrapalli, P., and Jain, P.: MET: Masked Encoding for Tabular Data, arXiv, https://doi.org/10.48550/arXiv.2206.08564, 2022.

Maron, J., Anagnostos, D., Brodbeck, B., and Meyer, A.: Artificial intelligence-based condition monitoring and predictive maintenance framework for wind turbines, J. Phys.: Conf. Ser., 2151, 1–9, https://doi.org/10.1088/1742-6596/2151/1/012007, 2022.

Stehly, T., Duffy, P., and D., H.: Cost of Wind Energy Review: 2024 Edition, National Renewable Energy Laboratory, Denver, USA, 2024.

Tautz-Weinert, J. and Watson, S. J.: Using SCADA data for wind turbine condition monitoring – a review, IET Renew. Power Gener., 11, 382–394, https://doi.org/10.1049/iet-rpg.2016.0248, 2017.

Teh, H. Y., Kempa-Liehr, A. W., and Wang, K. I.-K.: Sensor data quality: a systematic review, J. of Big Data, 7, https://doi.org/10.1186/s40537-020-0285-1, 2020.

Trapani, N. and Longo, L.: Fault Detection and Diagnosis Methods for Sensors Systems: a Scientific Literature Review, IFAC-PapersOnLine, 56, 1253–1263, https://doi.org/10.1016/j.ifacol.2023.10.1749, 22nd IFAC World Congress, 2023.

Trizoglou, P., Liu, X., and Lin, Z.: Fault detection by an ensemble framework of Extreme Gradient Boosting (XGBoost) in the operation of offshore wind turbines, Renew. Energy, 179, 945–962, https://doi.org/10.1016/j.renene.2021.07.085, 2021.

Turnbull, A., Carroll, J., and McDonald, A.: Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure, Wind Energy, 24, 197–211, https://doi.org/10.1002/we.2567, 2021.

Verma, A., Zappalá, D., Sheng, S., and Watson, S. J.: Wind turbine gearbox fault prognosis using high-frequency SCADA data, J. Phys.: Conf. Ser., 2265, 032 067, https://doi.org/10.1088/1742-6596/2265/3/032067, 2022.

Wang, S., Vidal, Y., and Pozo, F.: Recent advances in wind turbine condition monitoring using SCADA data: A state-of-the-art review, Reliab. Eng. Syst. Safety., 267, 111 838, https://doi.org/10.1016/j.ress.2025.111838, 2026.