# A Probabilistic Baseline Learning Framework for SCADA-Based Wind Turbine Aging Diagnosis and Multi-Scale Performance Degradation Analysis

Qianling Wang[1,2,*], and Yolanda Vidal[1,3]

[1] *Control, Data, and Artificial Intelligence (CoDAlab), Department of Mathematics, Universitat Politècnica de Catalunya · BarcelonaTech (UPC), Av. Eduard Maristany, 16, 08019 Barcelona, Spain*
*qianling.wang@upc.edu*
*yolanda.vidal@upc.edu*

[2] *School of Civil Engineering, Chongqing Jiaotong University, Chongqing 400074, China*

[3] *Institute of Mathematics (IMTech), Universitat Politècnica de Catalunya · BarcelonaTech (UPC), Barcelona, 08028, Spain*

## ABSTRACT

Accurate diagnosis of wind turbine aging from SCADA data is essential for reliable long-term operation, yet conventional wind-speed-binning methods often fail to capture the nonlinear and condition-dependent nature of power degradation. To address this issue, this study proposes a probabilistic baseline learning framework for SCADA-based wind turbine aging diagnosis and multi-scale degradation analysis. A two-stage mean–variance XGBoost model with uncertainty calibration is developed to estimate both the healthy-reference power output and its predictive uncertainty under varying operating conditions. The deviation between measured power and the probabilistic healthy baseline is then used to quantify degradation across temporal, wind-speed, directional, and joint wind-speed–direction dimensions. The results show that turbine aging is cumulative but strongly condition-dependent, with the most severe degradation concentrated in specific operating sectors rather than uniformly distributed across all inflow states. Furthermore, some directional degradation hotspots appear to correspond to potential upstream turbine positions, suggesting a possible wake-related contribution, although further verification is still required. More importantly, the proposed framework not only improves aging diagnosis accuracy, but also provides wind farm operators with interpretable information on the conditions under which degradation is most severe, thereby supporting targeted inspection, maintenance planning, and future evaluation of wake-aware mitigation strategies such as active wake steering.



## 1. INTRODUCTION

With the continued expansion of wind energy deployment, an increasing number of wind turbines are entering long-term service, making aging-related performance degradation a growing concern for wind farm operation and asset management. Even in the absence of severe faults, gradual deterioration in aerodynamic efficiency, drivetrain behavior, or control performance can still lead to persistent power loss over time (Astolfi, Pandit, Celesti, Lombardi, & Terzi, 2022; S. Wang, Vidal, & Pozo, 2026). Such losses are economically significant because they accumulate throughout the service life of a turbine and directly affect energy yield, maintenance planning, and lifetime extension decisions. An analysis of long-term data from 282 wind farms in the United Kingdom reported an average annual reduction in energy production of $1.6 \pm 0.2\%$ associated with turbine aging (Staffell & Green, 2014). Reliable wind turbine aging diagnosis is therefore essential not only for quantifying long-term energy loss, but also for supporting condition-aware maintenance, operational optimization, and long-term asset management.

Existing wind turbine aging diagnosis methods are often based on aggregated indicators or conventional wind-speed-binned power curve analysis (Kim & Kim, 2025). Although such approaches are intuitive and easy to implement, they are typically built on deterministic comparisons and simplified operating assumptions, which limits their ability to characterize the nonlinear and condition-dependent nature of turbine degradation. In practice, performance deterioration does not occur uniformly across all inflow states, but is often concentrated in specific wind-speed ranges, directional sectors, or operating environments. Turbine power output is influenced not only by

wind speed, but also by short-term fluctuations, thermal conditions, directional inflow, and possible wake-related effects. Under these circumstances, deterministic baselines may confound true aging-related deviations with normal environmental and operational variability, thereby falsely attributing noise-induced fluctuations to degradation (Mathew, Kandukuri, & Omlin, 2022; Ulmer, Jarlskog, Pizza, Manninen, & Goren Huber, 2020; Singh, 2024). As a result, traditional methods provide limited information about predictive reliability and are insufficient for reliable condition-dependent aging diagnosis. These limitations motivate the development of an uncertainty-aware framework capable of separating persistent degradation signals from expected variability under diverse operating conditions.

To address this challenge, this study proposes a probabilistic baseline learning framework for SCADA-based wind turbine aging diagnosis and multi-scale degradation analysis. Instead of relying on deterministic baseline comparison, the proposed method formulates aging diagnosis as a deviation-learning task relative to a healthy operational reference. Specifically, a two-stage mean–variance XGBoost model is developed to estimate the healthy-reference power output and its input-dependent predictive uncertainty using multi-source SCADA and ERA5 variables (Q. Wang et al., 2024; Q. Wang, Yao, & Hou, 2025; Astolfi, Castellani, Lombardi, & Terzi, 2021). Based on the deviation between measured power and the probabilistic healthy baseline, turbine degradation is further quantified across temporal, wind-speed, directional, and joint wind-speed–direction perspectives. The main contributions of this study are threefold. First, a probabilistic healthy-baseline framework is established for wind turbine aging diagnosis. Second, input-dependent predictive uncertainty is explicitly incorporated to reduce false aging indications caused by environmental and operational variability. Third, a multidimensional degradation analysis strategy is developed to characterize long-term performance deterioration under diverse operating conditions. The present study shows that turbine aging is cumulative yet highly condition-dependent, highlighting the importance of uncertainty-aware and multidimensional diagnosis for reliable identification of wind turbine aging.

## 2. METHODOLOGY

### 2.1. Problem formulation

The objective of this study is to assess wind turbine aging by constructing a probabilistic healthy baseline model using historical operational and environmental data collected during an early-stage healthy baseline period. The baseline model is trained to learn the relationship between turbine operating conditions, inflow-related features, and the corresponding healthy power output, while also quantifying predictive uncertainty. The trained model is then applied to later operating periods to estimate the power output expected under the healthy baseline condition. Deviations between the measured power and the predicted healthy baseline are interpreted as degradation signals and used to characterize aging-related performance loss. These deviations are further analyzed across multiple dimensions, including temporal evolution, wind-speed dependence, directional variation, and joint wind-speed–direction conditions, so as to identify the operating regimes in which turbine aging is most pronounced.

### 2.2. Brief description of the dataset

This study is based on the Hill of Towie Wind Farm Open Dataset, which contains high-resolution 10-minute Supervisory Control and Data Acquisition (SCADA) records collected from 21 Siemens SWT-2.3-VS-82 onshore wind turbines (Clerc & Lingkan, 2025). Covering the period from January 2016 to August 2024, the dataset provides a rich basis for long-term degradation analysis and counterfactual performance assessment. As illustrated in Fig. 1, the turbines are distributed across the wind farm with distinct spatial arrangements, which is important for understanding heterogeneous operating conditions and potential wake interactions.

The dataset can be broadly divided into three components: (i) turbine-level operational measurements, including active power output, nacelle-measured wind speed, and related statistical descriptors; (ii) turbine metadata, including geographic coordinates and technical specifications; and (iii) maintenance and retrofit records, including downtime events and efficiency-improvement interventions.

To mitigate the limitations of turbine-local SCADA signals, which may be affected by local control strategies, wake-induced flow distortion, and sensor uncertainty, the internal measurements are further complemented by ERA5 reanalysis data. These external atmospheric variables provide a more stable mesoscale meteorological background, including ambient temperature, relative humidity, surface pressure, and wind information at 10 m and 100 m heights. By integrating turbine-level observations with reanalysis-based environmental descriptors, the dataset supports a more robust reconstruction of healthy operational baselines under varying atmospheric conditions.

This multi-source data foundation is particularly valuable for separating environmentally driven power fluctuations from structural or operational aging effects. As a result, it enables a more reliable quantification of long-term degradation patterns and provides a stronger basis for subsequent upgrade-effect evaluation.

### 2.3. Data processing and feature selection

The raw SCADA records are first subjected to a structured quality-control pipeline to ensure that the model training is based on valid and physically meaningful operating data. Samples flagged as invalid by the turbine-specific indicator "is_valid"

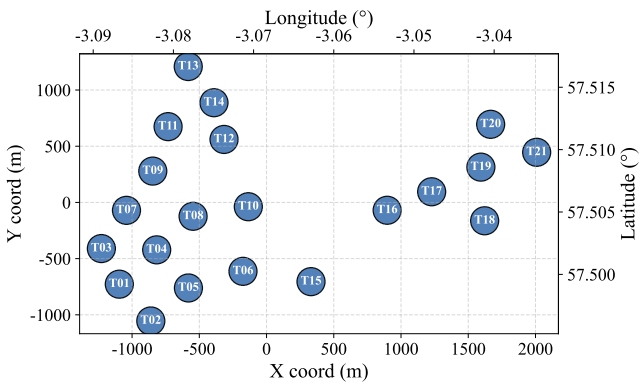


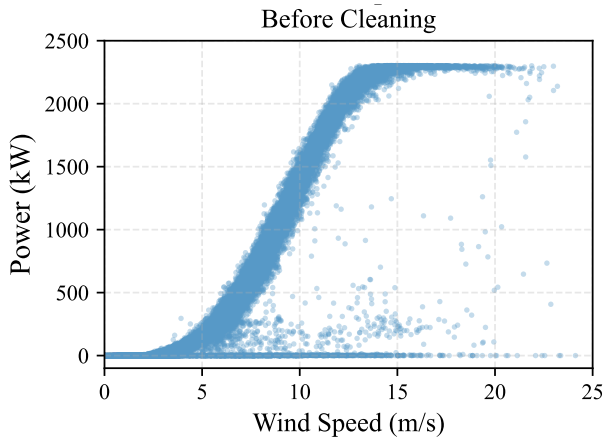

Figure 1. Spatial layout of the 21 wind turbines in the Hill of Towie wind farm.

are removed first, followed by observations with missing target power. To avoid contamination from non-steady operational states, shutdown-related records are then excluded using the shutdown-duration signal, together with a short temporal padding window before and after each shutdown event.

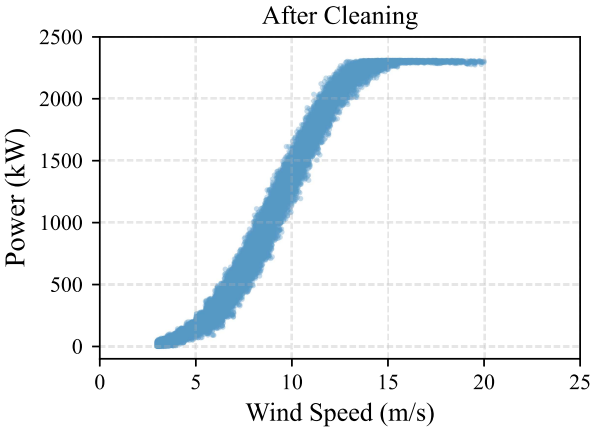
After this initial screening, physically implausible observations are further filtered in the wind speed–power domain. Only samples within the effective operating wind speed range are retained, and records with nonpositive power output under normal generating wind speeds are removed. To further enhance data reliability, a robust median absolute deviation filtering procedure is applied within wind-speed bins rather than over the full dataset. When a blade-pitch variable is available, the pitch is screened first to eliminate locally inconsistent control responses under comparable inflow conditions. The same bin-wise MAD criterion is then applied to turbine power so that abnormal local deviations can be suppressed while preserving the dominant operational distribution. After outlier removal, a small number of remaining missing feature values are smoothed using rolling-window averaging, and directional variables are transformed into sine and cosine components for subsequent modeling. The effectiveness of the overall cleaning procedure is illustrated by the comparison before and after cleaning in Fig. 2.

After data cleaning, feature selection is conducted using a two-stage screening strategy. An XGBoost pretraining step is first employed to rank candidate SCADA variables according to their contribution to turbine power prediction. The most informative variables are then further examined using Pearson correlation analysis, through which highly correlated and therefore redundant descriptors are removed. This process yields a compact but information-rich feature subset that captures inflow intensity, short-term variability, environmental condition, and turbine orientation.

The final variables retained include mean wind speed, standard deviation of wind speed, ambient temperature, nacelle




(a) Before Data Cleaning




(b) After Data Cleaning

Figure 2. Comparison Before and After Data Cleaning

position, and ERA5 wind direction. Since directional measurements are periodic and discontinuous at 0° and 360°, nacelle position and wind direction are transformed into sine and cosine components prior to model training. This treatment preserves directional continuity and improves numerical stability in nonlinear regression. Consequently, the baseline model is trained using mean wind speed, wind speed standard deviation, ambient temperature, and the sine-cosine encodings of nacelle position and wind direction.

To ensure numerical completeness after the cleaning stage, residual missing values in the retained predictors are smoothed using rolling-window averaging and then filled when necessary. Compared with conventional wind-speed-only inputs used in bin-based power-curve methods, this selected feature set provides a more comprehensive description of the turbine power response under varying inflow, thermal, and directional conditions. Such a representation is better suited to capturing nonlinear operational behavior and therefore provides a more reliable baseline for subsequent degradation quantification and uncertainty-aware performance assessment.

### 2.4. Uncertainty-aware XGBoost baseline model

To construct a reliable baseline for turbine power prediction together with predictive uncertainty, a two-stage XGBoost-based mean–variance learning framework is adopted. Instead of using a single deterministic regressor, the model separately characterizes the conditional mean response and the input-dependent residual variance, thereby providing both point prediction and uncertainty estimation.

In the first stage, an XGBoost regressor is trained to learn the nonlinear mapping between the selected SCADA features and turbine power output. Given an input vector $\mathbf{x}_i$, the predicted power is expressed as

$$\hat{\mu}_i = f_\mu(\mathbf{x}_i), \tag{1}$$

where $f_\mu(\cdot)$ denotes the mean model and $\hat{\mu}_i$ represents the conditional mean prediction of turbine power. This component serves as the baseline estimator of the normal operational response.

In the second stage, another XGBoost regressor is introduced to model the heteroscedastic uncertainty associated with the first-stage prediction. Based on the estimated mean response, the squared residual of each sample is defined as

$$r_i^2 = (y_i - \hat{\mu}_i)^2, \tag{2}$$

where $y_i$ is the observed turbine power. To stabilize the variance learning process and avoid numerical issues caused by extremely small residuals, the logarithm of the squared residual is used as the learning target:

$$z_i = \log\left(r_i^2 + \varepsilon\right), \tag{3}$$

where $\varepsilon$ is a small positive constant for numerical stability. The second-stage model, denoted by $f_\sigma(\cdot)$, is therefore trained to approximate this log-variance-related quantity. The output of the second-stage model represents the logarithm of the conditional residual variance. Therefore, the sample-specific predictive standard deviation, denoted by $\hat{\sigma}_i$, can be recovered by exponentiating the predicted log-variance and then taking the square root:

$$\hat{\sigma}_i = \sqrt{\exp\left(f_\sigma(\mathbf{x}_i)\right)}. \tag{4}$$

Here, $\hat{\sigma}_i$ quantifies the input-dependent spread of the prediction residual around the mean estimate $\hat{\mu}_i$, and is used as a measure of local predictive uncertainty under the operating condition $\mathbf{x}_i$.

With both the conditional mean and the predictive standard deviation available, the turbine power output is approximated by a Gaussian predictive distribution,

$$y_i \sim \mathcal{N}\left(\hat{\mu}_i, \hat{\sigma}_i^2\right), \tag{5}$$

which allows the model to represent not only the expected power response but also its local uncertainty level under different operating conditions.

To improve the reliability of interval estimation, the predicted standard deviation is further calibrated using a scaling factor $\alpha$:

$$\hat{\sigma}_i^{\text{cal}} = \alpha \hat{\sigma}_i. \tag{6}$$

The factor $\alpha$ is determined on a calibration set by matching the empirical coverage of the prediction intervals to the target confidence level. Specifically, $\alpha$ is adjusted such that the proportion of calibration samples falling within the nominal interval is as close as possible to the desired coverage. To account for differences in operating characteristics and uncertainty patterns, $\alpha$ is calibrated separately for each turbine or experimental case rather than being fixed globally. This calibration step compensates for possible under- or over-estimation of the raw predictive uncertainty and improves the reliability of interval estimation.

Based on the calibrated uncertainty, a two-sided $(1-\beta)$ prediction interval can be constructed as

$$\left[\hat{\mu}_i - z_{1-\beta/2}\hat{\sigma}_i^{\text{cal}},\ \hat{\mu}_i + z_{1-\beta/2}\hat{\sigma}_i^{\text{cal}}\right], \tag{7}$$

where $\beta$ denotes the significance level and $z_{1-\beta/2}$ is the corresponding standard normal quantile. In this study, $\beta = 0.05$ is used, corresponding to a 95% prediction interval with $z_{0.975} = 1.96$.

This dual-regressor structure enables the model to capture both the central nonlinear power-conversion relationship and the input-dependent uncertainty of the residuals. Compared with a conventional single-output regression model, it provides a more informative baseline for subsequent degradation diagnosis and uncertainty-aware performance assessment.

### 2.5. Experimental setup and model configuration

The dataset is partitioned chronologically to ensure consistent baseline construction and unbiased model evaluation. Although records are available from 2016, 2017 was selected as the healthy-reference period based on data completeness, cleaned sample availability, intra-annual power-curve stability, and inspection of the publicly available alarm and upgrade records. According to the dataset documentation, the documented performance-upgrade interventions, including AeroUp and TuneUp, occurred after 2021 rather than during 2017. The 2017 alarm-log records were also inspected together with the available alarm-code dictionary. For alarm codes with available descriptions, the dominant interpretable records were mainly associated with environmental or operational states, such as generator cut-in/cut-out, low-wind availability, high-wind shutdown, icing-related stops, and pitch-lubrication events. Because not all internal SCADA event codes are mapped to textual descriptions in the public alarm-code dictionary, unmapped records were further examined in

terms of duration, recurrence, and turbine-level concentration rather than being assigned unsupported physical meanings. After invalid and shutdown-related records were excluded during data cleaning, the cleaned 2017 SCADA samples were considered suitable for learning an operationally representative early healthy baseline power response. Data from the first quarter of 2018 are used as the validation set, the second quarter of 2018 as the uncertainty-calibration set, and the second half of 2018 as the test set.

Table 1. Hyperparameter settings of the XGBoost mean and variance models

| Branch | Hyperparameter | Value |
|---|---|---|
| Mean prediction | n_estimators | 40000 |
| | learning_rate | 0.0005 |
| | max_depth | 6 |
| | subsample | 0.8 |
| | colsample_bytree | 0.8 |
| Variance prediction | n_estimators | 3000 |
| | learning_rate | 0.01 |
| | max_depth | 4 |
| | subsample | 0.8 |
| | colsample_bytree | 0.8 |

The hyperparameter values used in the final experiments are selected through preliminary grid-search trials prior to formal evaluation. The detailed configurations of the XGBoost mean and variance models are listed in Table 1.

Both branches are implemented using XGBoost regressors. To avoid excessive complexity in residual modeling, the variance branch is configured more conservatively than the mean branch, enabling it to capture the dominant heteroscedastic pattern while maintaining stability.

A 95% confidence level is adopted for uncertainty quantification, corresponding to $z = 1.96$. Moreover, a small constant $\varepsilon = 10^{-6}$ is used to ensure numerical stability in logarithmic and variance-related operations. The random seed is fixed at 42 in all experiments for reproducibility.

### 2.6. Evaluation metrics

The predictive performance of the baseline model is evaluated using the coefficient of determination ($R^2$), mean absolute error (MAE), and normalized mean bias error (NMBE). $R^2$ reflects the goodness of fit between predicted and observed power outputs, MAE measures the average absolute prediction error, and NMBE quantifies the average signed bias relative to the mean observed power. Together, these metrics provide complementary information on fitting accuracy, error magnitude, and systematic bias.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{8}$$

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{9}$$

$$\mathrm{NMBE} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n\bar{y}} \times 100\% \tag{10}$$

Here, $y_i$ and $\hat{y}_i$ denote the observed and predicted power outputs, respectively, $\bar{y}$ is the mean observed power, and $n$ is the sample size. The reliability of the 95% confidence interval is additionally assessed by examining whether the observed values are generally covered by the calibrated predictive bounds.

## 3. RESULTS AND DISCUSSION

### 3.1. Model performance

Table 2. Prediction performance of the baseline model for turbines T1, T10, and T18.

| | | Training set | Validation set | Test set |
|---|---|---|---|---|
| T1 | R2 | 0.999 | 0.995 | 0.997 |
| | MAE | 12.09 | 24.59 | 19.36 |
| | NMBE(%) | 0.00 | 0.64 | 0.84 |
| T10 | R2 | 0.999 | 0.997 | 0.997 |
| | MAE | 10.13 | 26.38 | 15.11 |
| | NMBE(%) | 0.00 | 0.79 | 0.21 |
| T18 | R2 | 0.999 | 0.996 | 0.997 |
| | MAE | 11.10 | 30.37 | 14.73 |
| | NMBE(%) | 0.00 | 1.10 | 1.30 |

Table 2 summarizes the predictive performance of the probabilistic baseline model for three representative turbines, T1, T10, and T18, which are located in different spatial regions of the wind farm and therefore reflect different inflow and wake-interaction conditions. The model achieves high accuracy across the training, validation, and test sets, with $R^2$ values above 0.995 for all three turbines during validation and testing. The test MAE remains within approximately 15–19 kW, which is small relative to the rated turbine power, while the NMBE values remain close to zero or within a limited positive range. These results indicate that the learned baseline can reproduce the normal power response with limited systematic bias.

It should be noted that the validation MAE is higher than the training MAE, indicating a moderate generalization gap between the 2017 baseline-training period and the subsequent 2018 validation period. This gap may be partly attributed to temporal distribution shifts, including seasonal variability, inflow-condition differences, and changes in operating regimes, rather than model overfitting alone. Importantly, the validation and test $R^2$ values remain consistently high, and the test MAE is comparable to or lower than the validation MAE for the selected turbines. This suggests that the model retains reasonable out-of-sample generalization capability and

does not exhibit strong systematic bias in the healthy-baseline prediction.

To reduce the risk of overfitting, the XGBoost models were configured with moderate tree depth, subsampling, and column subsampling, and the hyperparameters were selected based on validation performance. Future implementations may further incorporate early stopping or stronger regularization to reduce this generalization gap, although the present validation and test results indicate that the current baseline is sufficiently accurate for degradation analysis.

Fig. 3 presents the temporal comparison between the measured power output and the model-predicted baseline power during the evaluation period. The predicted power closely follows the temporal fluctuations of the actual turbine output, indicating that the proposed XGBoost-based baseline model effectively captures the nonlinear relationship between environmental variables and turbine power generation.

The 95% confidence interval further demonstrates the robustness of the uncertainty estimation. Most observed power values fall within the predicted confidence bounds, indicating that the model provides reliable probabilistic predictions of the expected power production under healthy operating conditions. Minor deviations are observed during periods of rapid power variation, which may be associated with transient atmospheric turbulence or short-term control adjustments. Overall, the strong agreement between the predicted and measured power confirms that the model can serve as a credible reference baseline for subsequent degradation analysis.

Fig. 4 presents the semiannual decline rates of T1, T10, and T18 together with their corresponding uncertainty bounds. These three turbines are selected as representative cases because they occupy different spatial regions of the wind farm (Fig. 1), including the farm boundary (T1), the central turbine cluster (T10), and the opposite downstream side (T18). This selection captures different wake interaction environments and reduces spatial bias, providing a more representative view of long-term turbine degradation across the farm.

The results reveal a persistent deterioration in turbine performance over time, as reflected by the progressively more negative decline rates of all three units. This trend indicates that the power response of the turbines gradually departs from the earlier healthy-reference state during long-term operation.

From 2019-H1 to 2022-H1, the degradation trajectories of the three turbines remain relatively smooth and gradual, suggesting a progressive aging process rather than abrupt performance loss. During this period, T10 exhibits the mildest deterioration, with decline rates remaining approximately within $-0.5\%$ to $-2.5\%$. T18 shows a slightly stronger but still moderate decline, generally varying from about $-1.3\%$ to $-2.6\%$. By contrast, T1 experiences the most evident degradation before 2022-H2, with the decline rate decreasing from roughly $-1.4\%$ to about $-3.7\%$. Despite these inter-turbine differences, all three units show a consistent long-term downward tendency before 2022-H2, indicating that the baseline model captures the gradual performance loss reasonably well during this stage.

After 2022-H2, the decline rates become markedly more negative, especially for T1 and T18. T1 drops rapidly to nearly $-9\%$ in 2022-H2 and reaches about $-10\%$ around 2023-H1, before partially recovering, but still remaining at approximately $-8\%$ to $-9\%$ thereafter. T18 also shows a pronounced deterioration after 2022-H2, with the decline rate falling to around $-5\%$ to $-6\%$ and fluctuating within this range in the later period. T10 remains comparatively stable until 2022-H2, but subsequently decreases to roughly $-7\%$ around 2023-H2 and then recovers slightly to approximately $-5.5\%$ by 2024-H2. The more pronounced deviation observed after 2022-H2 may indicate a stronger shift in operating conditions relative to the baseline period. In addition to cumulative aging, it should be noted that the turbine underwent an upgrade in September, which may have further altered its operating characteristics and contributed to the observed mismatch.

The uncertainty bounds are generally narrow in the earlier periods, but become wider in several later half-year intervals, implying increased variability once turbine operation departs further from the original healthy baseline. Overall, these results suggest that long-term degradation is cumulative but strongly turbine-dependent, and that later-stage performance evolution may become increasingly nonlinear and unstable.

It should be noted that Fig. 4 is intended to provide a full-period overview of the semiannual performance evolution up to 2024, including both pre- and post-upgrade operating periods. By contrast, the subsequent wind-speed-dependent, directional, and joint wind-speed–direction analyses are restricted to the pre-upgrade period ending in 2022-H1. This restriction is adopted because the post-2022-H2 records may be affected by upgrade-related or control-induced changes in turbine operation, which could confound gradual aging-related degradation with retrofit-induced changes in the turbine power response.

### 3.2. Wind-speed-dependent degradation characteristics

Fig. 5 compares the annual wind speed–power curves of turbine T1 relative to the 2017 healthy baseline for the pre-upgrade period from 2018 to 2022-H1. This period is selected to isolate wind-speed-dependent aging signatures before possible upgrade-related changes in turbine operation occurred after 2022-H2. Including the post-upgrade records in the same wind-speed-binned comparison could confound gradual aging-related degradation with retrofit- or control-induced changes in the turbine power response. Overall, the curves shift downward progressively in later years, indicating gradual performance deterioration over time.

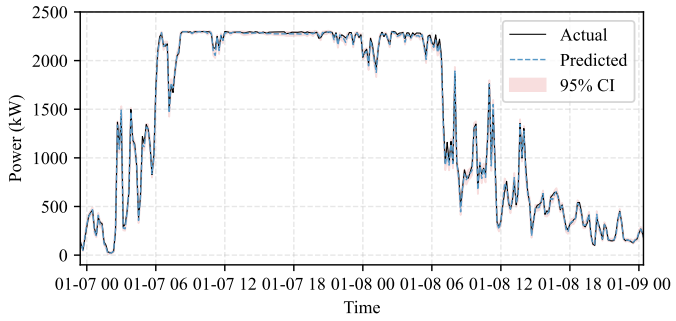



Figure 3. Comparison of daily actual and predicted power with a 95% CI.

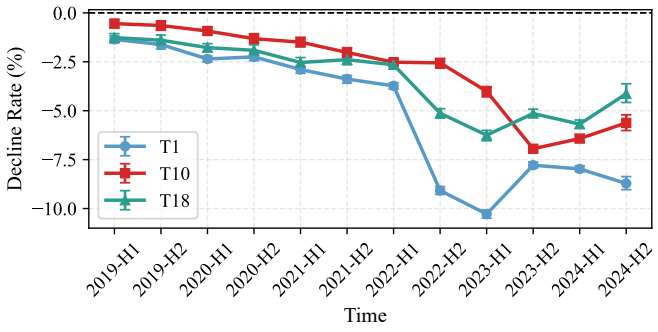



Figure 4. Semiannual performance decline rate with 95% confidence intervals from 2019 to 2024.

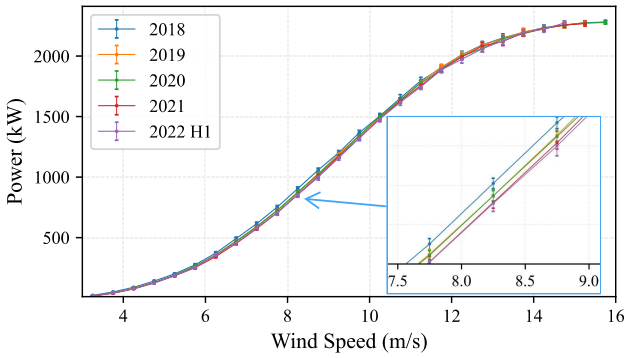



Figure 5. Comparison of annual power curves from 2018 to 2022 H1.

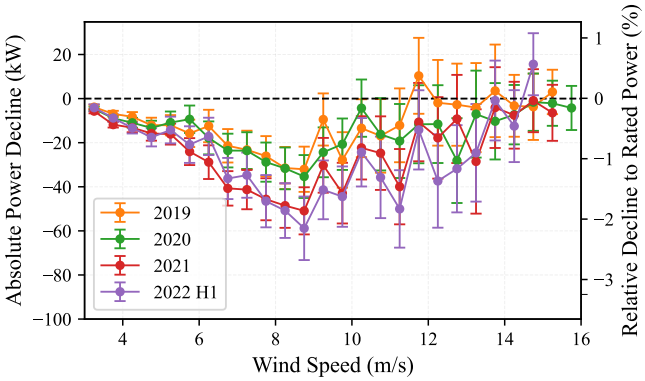



Figure 6. Absolute and relative power decline across different wind speed bins from 2019 to 2022 H1.

Fig. 6 further quantifies this pattern by showing the absolute and relative power decline with respect to the 2017 baseline. The largest losses occur in the partial-load regime, especially between 6 and 10 m/s, where the decline in 2021 and 2022-H1 reaches approximately 40–60 kW in several wind-speed bins, corresponding to about 1.5%–2.5% of rated power. This confirms that the medium wind-speed region is the most sensitive operating range for identifying long-term performance deterioration. At higher wind speeds, particularly above about 12 m/s, the decline becomes smaller and more variable, while the uncertainty bars become more pronounced. This behavior is consistent with the transition to rated operation, where pitch control increasingly constrains the output and reduces the visibility of degradation in absolute power terms.

Taken together, these results show that turbine degradation is strongly wind-speed dependent rather than uniform across the operating range. Relative to the 2017 baseline, the medium wind-speed region provides the clearest and most robust degradation signature, while the later years exhibit systematically larger power deficits, confirming the cumulative nature of long-term performance loss.

### 3.3. Wind turbine degradation performance analysis

Following the wind-speed-dependent analysis, the directional and joint wind-speed–direction degradation analyses are also restricted to the pre-upgrade period ending in 2022-H1. This restriction is intended to isolate direction-dependent aging signatures before possible upgrade-related operational changes occurred and to avoid attributing post-upgrade deviations solely to natural degradation. The directional characteristics of T1 performance degradation relative to the 2017 baseline are illustrated in Fig. 7. For each operating year, the polar curves quantify the direction-dependent performance change with respect to the 2017 healthy reference.

Overall, the degradation is clearly direction-dependent rather than uniformly distributed over all wind directions. The curves for 2019 and 2020 remain relatively close to the baseline over most sectors, indicating that the directional performance loss is still limited during the earlier stage of operation. By contrast, the 2021 and especially the 2022-H1 curves show much stronger negative deviations in several directional sectors, demonstrating that the departure from the 2017 reference condition becomes increasingly pronounced over time.

The most substantial reductions are concentrated in specific inflow sectors rather than being evenly spread around the full directional range. In particular, the largest negative deviations appear in several sectors between approximately $0°$ and $135°$, where the 2022-H1 curve drops well below those of the earlier years. This result suggests that the turbine becomes increasingly sensitive to certain directional inflow conditions as degradation accumulates. Such sector-dependent deterioration may be associated with persistent wake exposure, local terrain-induced flow distortion, or direction-specific aerodynamic and control-response deficiencies.

In contrast, the magnitude of the degradation remains comparatively moderate in some other sectors, particularly over parts of the western and southwestern directional ranges, where the annual curves remain closer to each other. This indicates that the long-term performance loss is strongly modulated by the local inflow environment and cannot be interpreted as a purely time-driven uniform decline. Instead, the figure shows that the directional operating conditions play an important role in shaping how turbine degradation manifests itself.

Together, these results demonstrate that long term turbine degradation has a clear directional signature when measured against the 2017 baseline. The increasing spread and deepening negative deviation of the later-year curves, especially in 2022-H1, provide direct evidence that the turbine response under certain inflow directions deteriorates more rapidly than under others. This directional heterogeneity highlights the need to incorporate wind-direction information into degradation assessment rather than relying solely on direction-aggregated performance indicators.

To investigate the coupled influence of wind speed and wind direction on deterioration of turbine performance, a two dimensional joint analysis is conducted, as shown in Fig. 8. The figure presents the distribution of absolute power loss in the wind speed–direction domain, thereby providing a localized view of how degradation varies under different inflow conditions.

The results reveal that the degradation pattern is highly non-uniform in the joint wind speed–direction space. Rather than being evenly distributed, the absolute power loss is concentrated in several specific directional sectors and wind speed intervals, indicating that turbine deterioration is strongly condition dependent. In particular, the most pronounced losses are observed in the medium-to-high wind speed range, especially around approximately 9–13 m/s, and are mainly clustered in directional sectors near $255°$–$300°$. Within these regions, the loss magnitude becomes substantially larger than that in the surrounding bins, suggesting that the turbine experiences more severe performance deterioration under these combined inflow conditions.

By contrast, at low wind speeds, the absolute loss generally

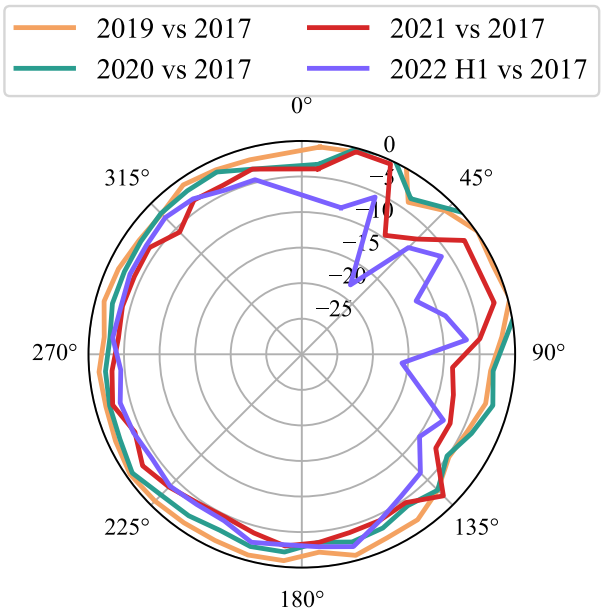



Figure 7. Directional relative power change compared to the 2017 baseline.

remains relatively small and spatially smooth across most directional sectors. This is expected because the available aerodynamic power is limited in the low-wind regime and therefore, even if degradation exists, its manifestation in absolute power terms remains less pronounced. However, in the medium wind speed region, the increased wind energy input and the higher sensitivity of the turbine power response make the degradation signature much more visible. The emergence of localized loss clusters in this range suggests that long-term performance decline is not governed solely by operational age, but is also shaped by the interaction between inflow direction and wind-speed-dependent aerodynamic response.

Another notable feature is the presence of several isolated bins with exceptionally strong negative loss values in the higher wind speed region. These localized anomalies imply that, under certain directional inflow conditions, the turbine departs much more severely from the healthy baseline than under neighboring conditions. This behavior may reflect the combined influence of wake exposure, directional turbulence, terrain-induced flow distortion, or control-response deviations that become more evident in specific operating regimes.

Figures 8 and 9 further extend the directional analysis by resolving degradation simultaneously in the wind speed–wind direction space.

Figure 8 shows the absolute loss change under different inflow combinations. The most pronounced absolute losses are concentrated in a limited number of localized operating regimes, particularly in the medium-to-high wind speed range and within several directional sectors around $240°$–$300°$. This

suggests that the additional energy loss observed in 2022 H1 is not uniformly distributed, but instead emerges under specific wind speed–direction combinations where the turbine is more susceptible to performance deterioration.

Fig. 9 presents the corresponding degradation in relative terms. Compared with the absolute-loss map, the percentage-based decline is more broadly distributed and is especially evident in the low-wind regime, with stronger negative deviations extending across several sectors from approximately 0° to 135°. This indicates that the relative degradation pattern becomes more spatially extensive once the loss is normalized by the expected baseline output, even if the largest absolute losses remain localized.

The apparent differences between Figs. 7, 8, and 9 arise from the different aggregation scales and response metrics used in these visualizations. Fig. 7 aggregates degradation over all wind-speed regimes within each directional sector and therefore reflects the overall directional tendency. Fig. 8, in contrast, isolates localized hotspots in terms of absolute energy loss under specific wind speed–direction combinations, which naturally emphasizes medium-to-high wind speed regions where the power magnitude is larger. Figure 9 further normalizes the loss by the expected baseline output, making low-power operating states more sensitive in relative terms. Therefore, these three figures should be viewed as complementary rather than contradictory: Fig. 7 captures sector-level integrated degradation, Fig. 8 highlights localized absolute-loss hotspots, and Fig. 9 reveals condition-dependent relative deterioration that can be obscured after aggregation.

To provide a more quantitative geometric interpretation of the directional hotspots, the wind-farm layout and turbine coordinates shown in Fig. 1 were used to calculate the bearing angles from T1 to its neighboring turbines. The nearest turbines around T1 are located at bearings of 337.2° for T03, 144.4° for T02, 42.3° for T04, 93.7° for T05, and 4.6° for T07, with corresponding distances of approximately 344.5–660.3 m. Several additional neighboring turbines also fall within the 0°–135° directional range, including T08 at 42.3°, T06 at 82.9°, T09 at 13.9°, T10 at 54.3°, and T15 at 89.1°. This comparison with Figs. 7–9 indicates that the relative-loss hotspot around 0°–135° is geometrically consistent with the bearings of multiple potential upstream turbines, suggesting a possible wake-related contribution to this part of the directional degradation pattern.

In contrast, the absolute-loss hotspot around 240°–300° is not directly aligned with the nearest-neighbor bearings of T1 in Fig. 1. This sector should therefore not be attributed solely to neighboring-turbine wake effects, but may also reflect terrain-induced inflow distortion, directional turbulence, wind-speed distribution differences, or control-response variability. Overall, the geometric comparison provides indicative evidence that wake interaction may contribute to part of the observed directional heterogeneity, while avoiding a definitive causal interpretation.

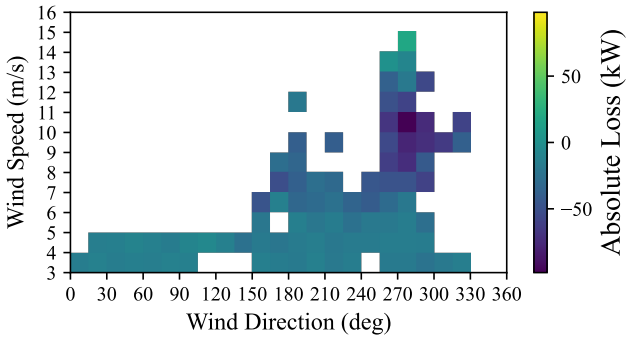



Figure 8. Wind Speed × Wind Direction Absolute Loss Change (2022 H1 vs 2017).

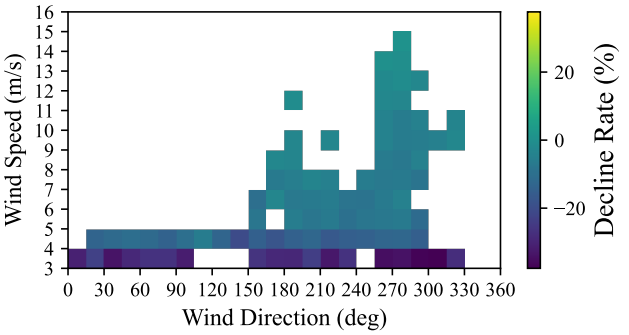



Figure 9. Wind Speed × Wind Direction Relative Loss Change (2022 H1 vs 2017).

## 4. CONCLUSION

This study proposes a probabilistic baseline learning framework for SCADA-based wind turbine aging diagnosis and multi-scale degradation analysis. By combining a two-stage mean–variance XGBoost model with uncertainty calibration, the framework provides both accurate healthy-baseline prediction and confidence-aware degradation quantification, thus improving the reliability of long-term aging assessment under varying operating conditions.

The results show that turbine degradation is cumulative but strongly condition-dependent. In addition to the long-term temporal decline, the degradation patterns exhibit a clear dependence on wind speed, wind direction, and their joint combinations, with the most severe loss concentrated in directional hotspots rather than uniformly distributed across all inflow states. These findings demonstrate that multidimensional analysis is necessary to identify the specific conditions under which turbine aging becomes more pronounced.

More importantly, the proposed framework does not simply improve the accuracy of the aging diagnosis. By pinpointing

the wind-direction sectors and wind-speed–direction regimes associated with the most severe degradation for specific turbines, it can provide wind farm operators with interpretable information for targeted inspection, maintenance planning, and condition-aware operational decision-making. In this sense, the framework offers not only diagnostic value, but also practical insight that may help to prolong turbine service life.

Furthermore, when interpreted together with the wind farm layout, some degradation hot-spots appear to correspond to potential upstream turbine directions, suggesting a possible wake-related contribution. Although this observation does not constitute definitive causal evidence, it indicates that the diagnosed high-risk sectors may provide a useful basis for future evaluation of wake-aware mitigation strategies, such as active wake steering, to reduce long-term downstream performance loss.

## ACKNOWLEDGMENT

This work is partially funded by the grant PID2021-122132OB-C21 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", by the "European Union"; Chongqing Jiaotong University Graduate Student Science and Technology Innovation Program (grant number 2024B0002); and Program of China Scholarship Council (grant number 202508500031).

## REFERENCES

Astolfi, D., Castellani, F., Lombardi, A., & Terzi, L. (2021). Data-driven wind turbine aging models. *Electric Power Systems Research*, *201*, 107495. doi: 10.1016/j.epsr.2021.107495

Astolfi, D., Pandit, R., Celesti, L., Lombardi, A., & Terzi, L. (2022). Scada data analysis for long-term wind turbine performance assessment: A case study. *Sustainable Energy Technologies and Assessments*, *52*, 102357. doi: 10.1016/j.seta.2022.102357

Clerc, A., & Lingkan, E. (2025). *Hill of towie wind farm open dataset (1.0.0) [data set].* Zenodo. doi: 10.5281/zenodo.14870023

Kim, D., & Kim, B. (2025). Quantification of performance degradation due to wind turbine aging: Estimating the reduction in annual energy production using the annual degradation rate. *Energy*, *324*, 136142. doi: 10.1016/j.energy.2025.136142

Mathew, M. S., Kandukuri, S. T., & Omlin, C. W. (2022). Estimation of wind turbine performance degradation with deep neural networks. In *Phme conference* (Vol. 7, pp. 351–359). doi: 10.36001/phme.2022.v7i1.3328

Singh, M. (2024). A hybrid – machine learning and possibilistic – methdology for predicting produced power using wind turbine scada data. In *Phme conference* (Vol. 8, p. 15). doi: 10.36001/phme.2024.v8i1.4006

Staffell, I., & Green, R. (2014). How does wind farm performance decline with age? *Renewable Energy*, *66*, 775–786. doi: 10.1016/j.renene.2013.10.041

Ulmer, M., Jarlskog, E., Pizza, G., Manninen, J., & Goren Huber, L. (2020). Early fault detection based on wind turbine scada data using convolutional neural networks. In *Phme conference* (Vol. 5, p. 9). doi: 10.36001/phme.2020.v5i1.1217

Wang, Q., Yao, G., & Hou, M. (2025). Fatigue life prediction and uncertainty quantification of aerospace metals: A bayesian physics-informed neural network model. *Reliability Engineering & System Safety*, 111724. doi: 10.1016/j.ress.2025.111724

Wang, Q., Yao, G., Kong, G., Wei, L., Yu, X., & Zhang, J. (2024). A data-driven model for predicting fatigue performance of high-strength steel wires based on optimized xgboost. *Engineering Failure Analysis*, *164*, 108710. doi: 10.1016/j.engfailanal.2024.108710

Wang, S., Vidal, Y., & Pozo, F. (2026). Recent advances in wind turbine condition monitoring using SCADA data: A state-of-the-art review. *Reliability Engineering & System Safety*, *267*, 111838. doi: 10.1016/j.ress.2025.111838