



Hamburg University of Applied Sciences

Faculty NIW

**Long-Term Turbulence Intensity Characterization and Evaluation of
Statistical, Machine Learning, and Remodelling Calibration Methods**

Bachelor Thesis

B.Sc. Process Engineering

RWE Offshore GmbH
Wind Resource Assessment

submitted by

Heba Irshaid



Hamburg, Germany

23.03.2026

Supervisor: Prof. Dr. Rainer Stank Hamburg University of Applied Sciences
Company Supervisor: Zachary Parker RWE Offshore GmbH

Declaration of Academic Integrity

I, Heba Irshaid, born on [REDACTED], hereby declare in lieu of oath that I have written the following bachelor's thesis independently and without the use of sources or aids other than those indicated:

Long-Term Turbulence Intensity Characterization and Evaluation of Statistical, Machine Learning, and Remodelling Calibration Methods

All passages taken literally or in substance from published or unpublished works have been identified as such. This thesis has not been submitted in whole or in part in the same or similar form in connection with any other academic examination.

I am aware that a false declaration will have legal consequences.

Hamburg, 23.03.2026

Dedicated to my mother and my father, for their unconditional love and support.

Abstract

An accurate long-term characterization of turbulence intensity is crucial for assessing wind conditions at future wind farm sites. Yet, no standardized method exists for extrapolating short-term measurements into reliable long-term predictions.

This thesis investigates how well selected statistical, machine-learning, and remodelling methods can predict long-term turbulence intensity using short-term measurements and long-term model data. Two contrasting sites are analyzed: an offshore site with stable wind conditions and an onshore site with a more irregular wind speed range. Eight calibration methods are tested and evaluated using three metrics - MRBE, RRMSE and DEL - and further refined through a three-tier post-processing correction framework.

Results show that the Vortex Times Remodelling and Extreme Gradient Boosting perform most consistently across MRBE and RRMSE, particularly at the offshore site. The three-tier correction framework can partially improve variance representation and extreme behavior across both sites. Regarding the DEL, some methods produce errors below the sites' natural year-to-year variability, meaning they outperform the current industry standard of relying on a single measurement year.

In conclusion, none of the tested methods consistently stay within industry-defined thresholds across all conditions, meaning full reliability remains an open challenge. Nevertheless, all calibration approaches bring the long-term model data closer to measured turbulence behavior. This study offers a clear comparison of the methods' strengths and limitations, providing practical guidance for method selection and improvement in future turbulence intensity predictions.

Table of Contents

List of Figures.....	I
List of Symbols and Abbreviations.....	II
1 Introduction	1
2 State of the Art	2
3 Theoretical Background.....	3
3.1 Turbulence Intensity.....	3
3.2 Statistical and Machine Learning Principles	4
3.2.1 Quantile Mapping.....	4
3.2.2 Decision Trees and Gradient Boosting	5
3.2.3 Neural Networks	5
3.3 Metrics	6
3.3.1 Mean Relative Bias Error	6
3.3.2 Relative Root Mean Square Error	8
3.3.3 Damage Equivalent Load	9
4 Approach and Methodologies	10
4.1 Input Data and Data Splitting.....	10
4.2 Methodologies.....	10
4.2.1 Quantile Mapping.....	11
4.2.2 Extreme Gradient Boosting	12
4.2.3 Remodel	12
4.2.4 Simple Scaling	12
4.2.5 Linear Least Squares	13
4.2.6 Matrix Scaling.....	14
4.3 Post-Processing: The Multi-Tier Correction Concept.....	15
5 Evaluation Results	17
5.1 MRBE	17
5.2 RRMSE	23
5.3 DEL	29
6 Discussion.....	37
7 Conclusion	41

List of Figures

Figure 1 Offshore Turbulence Intensity as a Function of Wind Speed in m/s, own illustration based on [3, pp. 116-117].....	4
Figure 2 Concept of the Long-Term Characterization Process	11
Figure 3 Concept of Multi-Tier Correction	15
Figure 4 Site 1 : Tier 1 MRBE in % and Best Method by Wind Speed Bin	17
Figure 5 Site 1 : Tier 2 MRBE in % and Best Method by Wind Speed Bin	18
Figure 6 Site 1 : Tier 3 MRBE in % and Best Method by Wind Speed Bin	19
Figure 7 Site 2 : Tier 1 MRBE in % and Best Method by Wind Speed Bin	20
Figure 8 Site 2 : Tier 2 MRBE in % and Best Method by Wind Speed Bin	21
Figure 9 Site 2 : Tier 3 MRBE in % and Best Method by Wind Speed Bin	22
Figure 10 Site 1 : Tier 1 RRMSE in % and Best Method by Wind Speed Bin.....	23
Figure 11 Site 1 : Tier 2 RRMSE in % and Best Method by Wind Speed Bin.....	24
Figure 12 Site 1 : Tier 3 RRMSE in % and Best Method by Wind Speed Bin.....	25
Figure 13 Site 2 : Tier 1 RRMSE in % and Best Method by Wind Speed Bin.....	26
Figure 14 Site 2 : Tier 2 RRMSE in % and Best Method by Wind Speed Bin.....	27
Figure 15 Site 2 : Tier 3 RRMSE in % and Best Method by Wind Speed Bin.....	28
Figure 16 Site 1 : Interannual Variability of DEL in %	29
Figure 17 Site 1 : Tier 1 DEL Error in % and Best Method by Wind Speed Bin.....	30
Figure 18 Site 1 : Tier 2 DEL Error in % and Best Method by Wind Speed Bin.....	31
Figure 19 Site 1 : Tier 3 DEL Error in % and Best Method by Wind Speed Bin.....	32
Figure 20 Site 2 : Interannual Variability of DEL in %	33
Figure 21 Site 2 : Tier 1 DEL Error in % and Best Method by Wind Speed Bin.....	34
Figure 22 Site 2 : Tier 2 DEL Error in % and Best Method by Wind Speed Bin.....	35
Figure 23 Site 2 : Tier 3 DEL Error in % and Best Method by Wind Speed Bin.....	36

List of Symbols and Abbreviations

Symbol/Abbreviation	Full Form
MCP	Measure-Correlate-Predict
TI	Turbulence Intensity
WS	Wind Speed
SD / σ	Standard Deviation
pred	Prediction
meas	Measurement
mod	Model
D	Distribution
MRBE	Mean Relative Bias Error
RRMSE	Relative Root Mean Square Error
DEL	Damage Equivalent Load
QM	Quantile Mapping
XGB	Extreme Gradient Boosting
RM	Remodel
SS	Simple Scaling
MMS	Monthly Matrix Scaling
YMS	Yearly Matrix Scaling
LLS	Linear Least Squares
WLLS	Wind Speed Binned Linear Least Squares
CV	Coefficient of Variation

1 Introduction

Turbulence intensity is an important parameter in wind energy since it influences turbine loading and impacts long-term site assessments. Reliable long-term predictions of turbulence intensity are required for site suitability and fatigue load evaluations, yet these predictions are difficult to obtain when only short-term measurements are available. In current practice, long-term turbulence characteristics are often estimated directly from short-term data, which introduces additional uncertainty. A standardized method for long-term turbulence extrapolation does not exist. Several statistical, machine-learning, and remodelling approaches are used in industry, but their performance has not been compared systematically. It is also unclear how these methods behave under different site conditions, especially when comparing offshore and onshore environments with fundamentally different turbulence characteristics.

The aim of this study is to evaluate calibration methods for predicting long-term turbulence intensity based on short-term measurements. To achieve this aim, the work examines how the statistical, machine-learning, and remodelling approaches behave across different wind speed ranges and site conditions, and how their performance changes when a three-tier correction framework is applied. The thesis investigates whether any method can reproduce long-term turbulence behavior reliably. The evaluation is based on the metrics MRBE, RRMSE, and DEL, which provide measurable criteria to determine whether the defined objective has been met.

The methods evaluated in this study are described in Chapter 4 and build on the theoretical principles introduced in Chapter 3. After the methods are presented, the evaluation results are shown and discussed. The discussion explains the observed results by linking them back to the theoretical background. Finally, the conclusion summarizes the main findings.

2 State of the Art

Wind resource assessment relies heavily on Measure-Correlate-Predict (MCP) methods to convert short-term on-site measurements into long-term predictions. This works by establishing a relationship between the measurement and a long-term reference dataset. The long-term reference dataset is a product developed by Vortex FDC. It is a MCP approach for long-term extrapolation of wind speed data.

Although MCP is widely established for wind speed, there is no standardized methodology for long-term turbulence intensity. In practice, data analysts rely on short-term turbulence intensity of measurements to represent long-term trends. This approach introduces uncertainty into both site suitability assessments and fatigue load evaluations.

Despite these developments, a systematic comparison of statistical, machine learning, and remodelling calibration methods of turbulence intensity has not been conducted, nor has the influence of the site types - onshore and offshore - been examined. This study addresses these gaps through an evaluation of multiple approaches across both site types.

3 Theoretical Background

Turbulence intensity is the key variable of this study. This chapter introduces its definition and connects its meaning to both onshore and offshore sites. The statistical and machine learning algorithms used in this study are introduced, which will be evaluated in addition to other methods, mentioned in Chapter 4, using a set of defined metrics.

3.1 Turbulence Intensity

Turbulence intensity is a crucial parameter in the design and operation of wind turbines, as it directly influences both the structural loads experienced by the turbines and their power output. Mathematically, turbulence intensity is defined as the ratio of the standard deviation of wind speed over a given time interval (Equation 1). With higher wind speeds turbulence decreases while lower wind speeds result in higher turbulence. This influences the turbine and results in increased fatigue loads. [1, p. 757]

$$TI_i = \frac{SD_i}{WS_i}$$

Equation 1

TI: = Turbulence Intensity

SD: = Standard Deviation in m/s

WS: = Wind Speed in m/s

i = Timestamp

Thermal conditions such as the heating and cooling of the atmosphere leads to generally higher values during warmer periods and lower values during colder periods [1, p. 760]. Overall, turbulence intensity is influenced by atmospheric stability, hub height, and terrain roughness, and therefore varies between the site types – offshore and onshore [1, p. 758].

Konagaya et al. [2] demonstrated the difference between offshore and onshore sites clearly, finding that the turbulence reference value at an onshore site was 37% higher than at an offshore site located only 1.5 km away. Additionally, onshore turbulence intensity depends strongly on wind direction, as fixed obstacles such as forests and buildings affect the wind flow in that sector, leading to greater variability between wind directions compared to offshore sites. [2, p. 236]. Furthermore, the land-sea contrast in turbulence is not solely driven by mechanical factors such as surface roughness, but also by thermodynamic factors, specifically seasonal variations in atmospheric stability. Onshore, atmospheric stability fluctuates more strongly across seasons due to the lower heat capacity of the land surface compared to the sea, leading to greater year-to-year variability in turbulence patterns at onshore sites [2, pp. 236-237].

For onshore sites, trends as the wind speed increases are complex and can be inconsistent. However, at offshore sites a decrease is rapid and continues relatively consistently, reaching

a minimum value usually around wind speeds of 12 m/s. Beyond this minimum point, offshore turbulence intensity begins to increase with further increases in wind speed. At low wind speeds, turbulence is mainly caused by thermal effects [3, pp. 116-117]. An overall trend of offshore turbulence intensity over wind speed in m/s is illustrated in Figure 1.

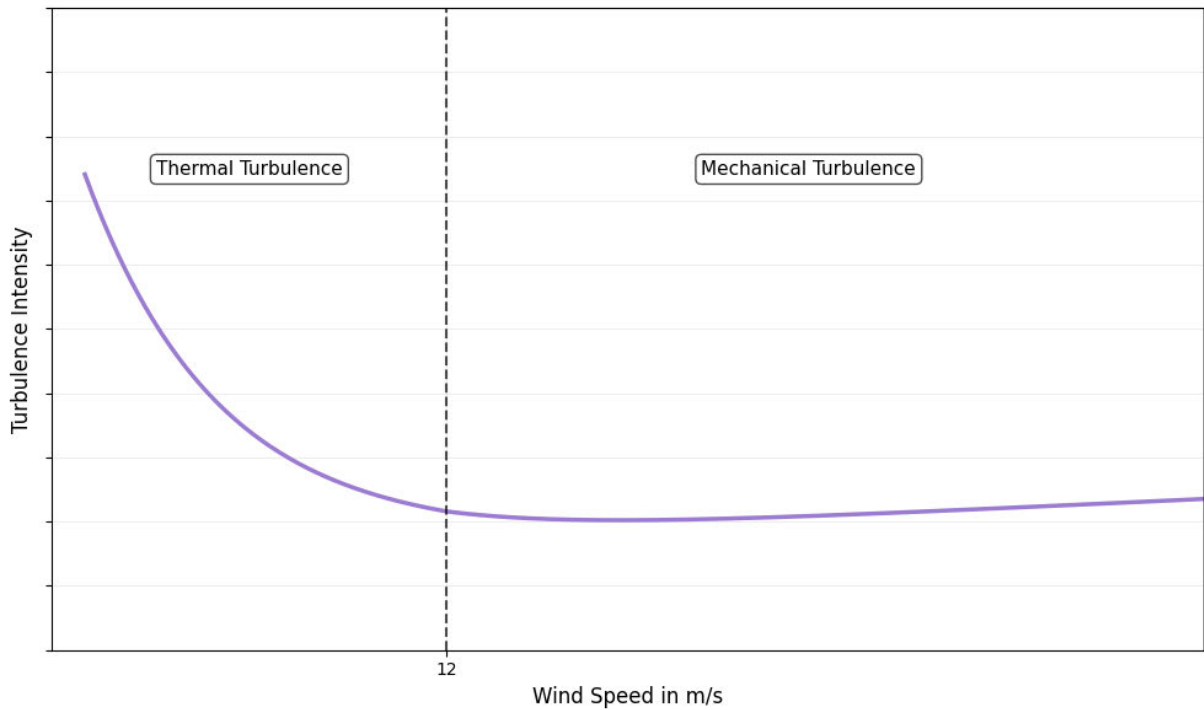


Figure 1 Offshore Turbulence Intensity as a Function of Wind Speed in m/s, own illustration based on [3, pp. 116-117]

The water with higher temperature increases the temperature of the atmosphere, causing the air to rise rapidly. This process leads to higher turbulence. As wind speed increases, thermal turbulence becomes less important. Mechanical turbulence takes over as the dominant source. This type of turbulence is created by wind shear and the roughness of the sea surface. Higher wind speeds produce larger waves, which make the surface rougher and increase friction between the air and water, resulting in higher turbulence. [3, pp. 116-117]

3.2 Statistical and Machine Learning Principles

This subchapter introduces the statistical and machine learning algorithms that are used in addition to self calibrated methodologies defined in Chapter 4 to predict long-term turbulence intensity. Each algorithm represents a different approach to correct a bias between measurement and base model data for a specific site. Understanding the principles is essential for interpreting the algorithms and results presented in Chapter 5.

3.2.1 Quantile Mapping

“Quantile Mapping” is a statistical post-processing method used to adjust the bias of the model data so that the distribution matches the measurement. Equation 2 describes a function-based

mapping, where modeled turbulence TI_{mod} is first located within its distribution D_{mod} and then assigned the value from the measured distribution D_{meas} that corresponds to the same percentile. This means the equation represents a distribution-to-distribution transformation. In other words, the model value is put through a function that looks up the position in the model distribution and then replaces it with the value that sits in the same position in the measurement data. This means the model value is matched to the observed distribution. [4, pp. 3383-3384]

$$TI_{pred} = \frac{D_{mod}(TI_{mod})}{D_{meas}}$$

Equation 2

TI: = Turbulence Intensity

D: = Distribution

pred: = Prediction

mod: = Model

meas = Measurement

3.2.2 Decision Trees and Gradient Boosting

A decision tree is a simple machine-learning model that makes predictions by asking a sequence of small yes-or-no questions about the input data. Each question checks a condition, such as whether a value is above or below a certain threshold, and depending on the answer the data moves down a different branch of the tree. This process continues until the data reaches a final point called a leaf, which contains the model's prediction. Decision trees can be used for classification, where each leaf represents a category, or for regression, where each leaf holds a numerical value. They are also easy to interpret because the entire prediction process is a transparent series of decisions. [5]

Gradient boosting builds on the idea of decision trees by combining many small, simple trees into a single, strong model. It starts with a very weak model that makes poor predictions. Then it calculates the errors this model makes and trains a small new decision tree specifically to correct those errors. This new tree is added to the model, improving it a little. The process repeats: each new tree focuses on whatever mistakes remain after the previous trees. Over time, many of these small correction-trees are added together, and the result is a model that performs far better than any of the individual trees alone. Gradient boosting therefore improves step by step, with each tree correcting the weaknesses of the model built so far. [6]

3.2.3 Neural Networks

A neural network is a type of machine-learning model designed to learn complex, nonlinear patterns in data. Unlike simple linear models, neural networks can discover curved or irregular relationships automatically during training. Each neural network is built from layers of small computing units called nodes. A node takes input values of model and measurement data,

performs a simple calculation on them, and passes the result forward. These nodes are arranged in layers: an input layer that receives the data, one or more hidden layers that gradually transform the information, and an output layer that produces the final prediction. The hidden layers play the key role as they allow the network to perform multiple transformations in sequence, enabling it to model nonlinear relationships that would be impossible for a linear model [7].

3.3 Metrics

After developing the results for each algorithm, their performance must be evaluated by comparing the predictions to the measurement data. The resulting differences, or biases, provide a basis for assessing how well each method captures the true turbulence characteristics. The following subchapters introduce the bias metrics used for this evaluation.

3.3.1 Mean Relative Bias Error

The Mean Relative Bias Error (MRBE) is used to quantify the difference from the modeled compared to the measurement. In Equation 3 it is defined as the percentage difference between the modeled and the measured turbulence intensity, normalized by the measured value and expressed. A positive MRBE indicates that the model systematically overestimates the true turbulence intensity, while a negative MRBE reflects a systematic underestimation. A value of zero would represent a perfectly unbiased model with no systematic deviation from the measured reference. The MRBE is therefore computed as a frequency-weighted metric across wind speed bins, in order to assess bias across the full wind speed range while maintaining statistical robustness through binning. [8, p. 345]

$$MRBE = \frac{(TI_{mod} - TI_{meas})}{TI_{meas}} \cdot 100\%$$

Equation 3

MRBE = Mean Relative Bias Error in %

TI = Turbulence Intensity

mod = Model

meas = Measurement

While the MRBE provides a straightforward measure of relative bias for a given training period, RWE's technical analysts extend Equation 3 to Equation 4. The MRBE is computed by averaging the relative differences between the modeled values $TI_{mod,n,i}$ and the measured values $TI_{meas,n,i}$ over a number of datapoints N_i . Then an overall frequency-weighted MRBE in percent is computed across wind speed bins for each time period i . The frequency-weighted MRBE is defined as:

$$MRBE_i = \sum_{wsb=1}^{WSB} \cdot f_{wsb} \left(\frac{1}{N_i} \sum_{n=1}^{N_i} \frac{(TI_{mod,n,i} - TI_{meas,n,i})}{TI_{meas,n,i}} \right) \cdot 100\%$$

Equation 4

MRBE = Mean Relative Bias Error in %

N = Number of Datapoints

TI = Turbulence Intensity

mod = Model

meas = Measurement

i = Time Period

f = Frequency

wsb = wind speed bin

WSB = Total Number of Wind Speed Bins

The acceptable thresholds for the frequency weighted MRBE applied to the categories of site suitability and load validation are defined by DNV, an international assurance and standards organization active in the wind-energy sector, as follows:

Site Suitability

$$-6\% \leq MRBE \leq +10\% \quad \text{for wind speeds below 7 m/s}$$

$$-3\% \leq MRBE \leq +10\% \quad \text{for wind speeds above 7 m/s}$$

Load Validation

$$MRBE \leq \pm 5\%$$

These thresholds ensure that the model's results remain within an acceptable range of the measurements, which is important for confirming site suitability and reliability.

3.3.2 Relative Root Mean Square Error

The Relative Root Mean Squared Error (RRMSE) provides a broader measure of the overall magnitude of the prediction error relative to the measured turbulence intensity and is defined in Equation 5 [8, p. 345].

$$RRMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{TI_{mod} - TI_{meas}}{TI_{meas}} \right)^2} \cdot 100\%$$

Equation 5

RRMSE = Relative Root Mean Squared Error in %

N = Number of Datapoints

TI = Turbulence Intensity

mod = Model

meas = Measurement

For each training period i , the RRMSE is computed by taking the square root of the mean of the squared relative differences between the modeled values TI_{mod} and the measured values TI_{meas} over all data points N , expressed as a percentage. Unlike the MRBE, which captures systematic bias and indicates the direction of the deviation, the RRMSE is sensitive to both systematic and random errors, as squaring the residuals gives more influence on larger deviations. However, since the residuals are squared before averaging, the RRMSE does not provide any indication of under- or overestimation of the error. [8, p. 345]

Consistent with the MRBE, RWE's analysts extend Equation 5 to a frequency-weighted metric in Equation 6 across wind speed bins, following the same weighting structure as Equation 4.

$$RRMSE_i = \sum_{wsb=1}^{WSB} \left(\sqrt{\frac{1}{N_i} \sum_{n=1}^{N_i} \left(\frac{TI_{mod,n,i} - TI_{meas,n,i}}{TI_{meas,n,i}} \right)^2} \right)$$

Equation 6

RRMSE = Relative Root Mean Squared Error in %

N = Number of Datapoints

TI = Turbulence Intensity

mod = Model

meas = Measurement

i = Training Period

Just as MRBE, RRMSE thresholds ensure that the model stays within an acceptable deviation from the measurements and meets the required accuracy for site suitability and load validation. The thresholds are defined by DNV as follows:

Site Suitability

$$\begin{aligned} RRMSE &\leq 30\% && \text{for wind speeds below 7 m/s} \\ RRMSE &\leq 15\% && \text{for wind speeds above 7 m/s} \end{aligned}$$

Load Validation

$$RRMSE \leq 15\%$$

3.3.3 Damage Equivalent Load

The Damage Equivalent Load (DEL) is a single value that summarizes how much fatigue a cyclic load would cause. Instead of looking at every small fluctuation in the load time series, the DEL turns all that variability into one constant load that would create the same amount of fatigue. This makes the DEL a metric for comparing fatigue loading across different operating conditions, time periods, or turbines, and for assessing how well a model reproduces the fatigue-relevant behavior observed in measurements. [9, pp. 1171-1172]

One mathematical definition of DEL used within RWE is expressed in Equation 7 as the standard deviation of the wind speed at each time step is taken and raised to the Wöhler exponent m , which reflects the material's fatigue sensitivity [9, p. 1171]. These values are then averaged over the number of data points, and finally the $1/m$ power is applied.

$$DEL = \left[\frac{1}{N} \sum_{i=1}^N SD_i^m \right]^{\frac{1}{m}}$$

Equation 7

DEL = Damage Equivalent Load

N = Number of Datapoints

SD = Standard Deviation

i = Time Step

m = Wöhler Exponent

4 Approach and Methodologies

This chapter outlines the overall approach used to predict long-term turbulence intensity. It introduces the input data, presents the tested methodologies and describes the multi-tier correction concept.

4.1 Input Data and Data Splitting

The input data consist of two long-term measurement datasets and the corresponding model datasets for the same periods.

The two sites analyzed in this study are an offshore site - Site 1 - and an onshore site - Site 2. Although both sites provide long-term observations, the available measurement periods differ: Site 1 contains four years of data collected from 2004 to 2007, whereas Site 2 provides nine years of measurements spanning 2010 to 2018. All measurement datasets are available at a 10-minute resolution and include mean wind speed in m/s, the standard deviation of wind speed in m/s, and wind direction in degrees. Turbulence intensity is calculated for each 10-minute time step using Equation 1. For each site, the corresponding model data – which is a Vortex TIMES product - follow the same 10-minute resolution and contain the same set of variables, allowing turbulence intensity to be calculated in the same manner.

For each measurement dataset, the time series is divided into yearly segments, referred to in this study as training periods. Each training period therefore represents a short-term dataset, with Site 1 providing four training periods and Site 2 providing nine. The algorithms are trained separately on each individual year, and only the data within that respective training period are used for prediction. This setup reflects a realistic situation in which only limited measurement data are available for calibration. Using all available training periods individually for each method allows the study to compare how the calibrated results vary from year to year.

4.2 Methodologies

The following subchapters introduce the methods used to predict long-term turbulence intensity, while several of these methods build on the concepts described in Chapter 3.2.

All methods follow the concept illustrated in Figure 2, where wind speed and its standard deviation are predicted.

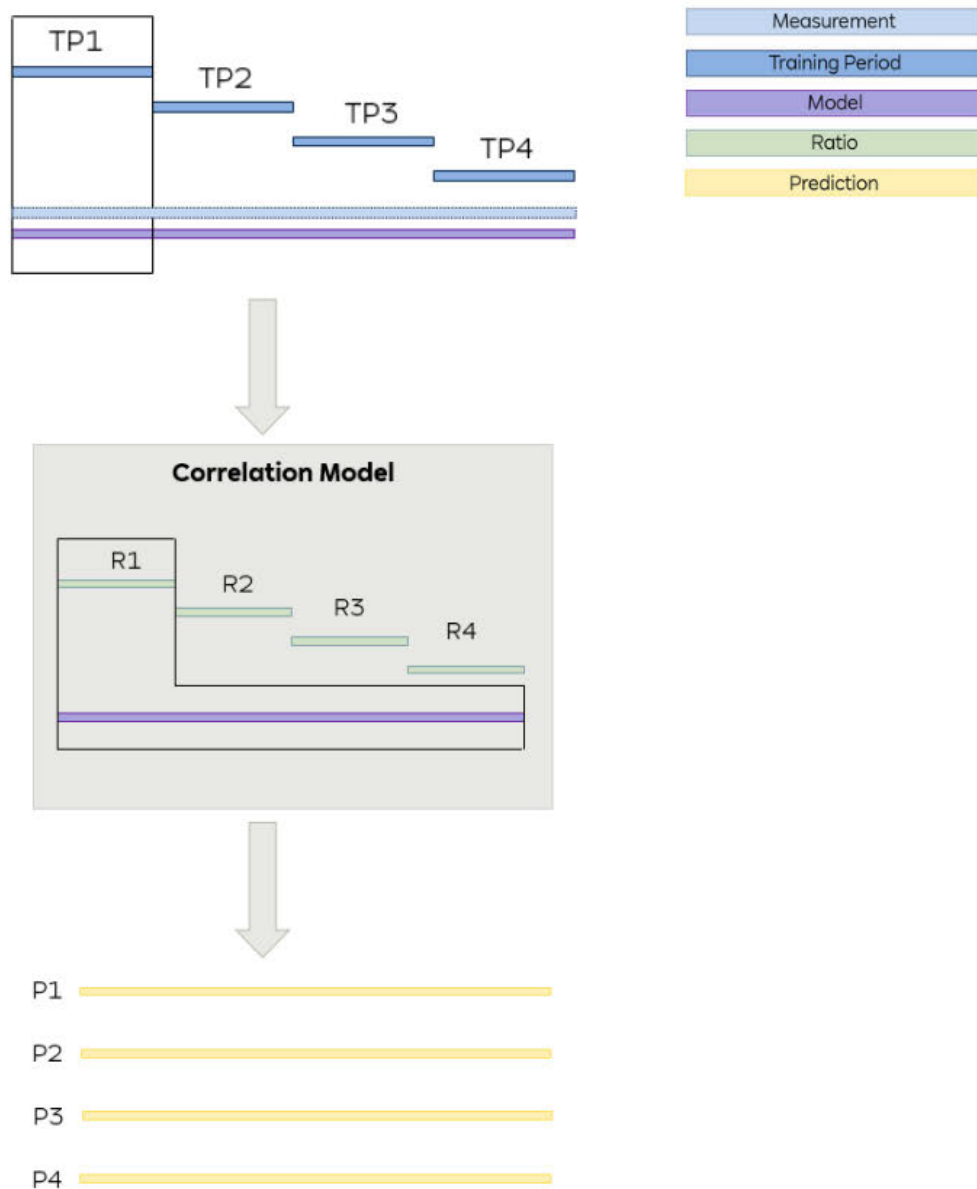


Figure 2 Concept of the Long-Term Characterization Process

The measurement is divided into multiple training periods. In that training period, a relationship between the measurement and the corresponding model is determined through the selected method – the correlation model. Once established, this relationship is applied to the entire model long-term dataset, allowing each training period of each algorithm to produce its own predicted long-term time series.

Since all methods generate predictions for wind speed and its standard deviation only, turbulence intensity is calculated afterwards using Equation 1.

4.2.1 Quantile Mapping

“Quantile Mapping” (QM) applies the idea from Chapter 3.2.1 by adjusting the modelled wind speed and standard deviation. For every month of the year, the method compares how the model and the measurements are distributed and then replaces each model value with the

value that sits at the same percentile in the measured distribution. This allows the corrected values to align with the observed distribution while still preserving month-to-month differences in variability. When a month does not contain enough information to construct a reliable distribution, the original model values are retained, and values outside the learned percentile range are limited to the lower or upper end of the measured distribution. In this way, QM provides a month-specific adjustment.

4.2.2 Extreme Gradient Boosting

“Extreme Gradient Boosting” (XGB) is a tree-boosting method that builds decision trees one after another. The concept is described in Chapter 3.2.2 where each new tree learns from the mistakes of the trees built before it and corrects their residual errors. The process starts with the original datasets, and several trees are trained in sequence. The predictions from all trees are then combined to form the final prediction output. The method improves accuracy by integrating what each previous tree has already learned. It also uses an objective function that balances how well the model fits the data and how complex the trees are. A regularization term in this objective function helps prevent overfitting by penalizing overly complicated trees. The number of trees and their depth play an important role in how well the model performs. [10, pp. 126-127]

4.2.3 Remodel

The Vortex “TIMES Remodeling” (RM) is an extension of the Vortex TIMES product that enables long-term extrapolation of wind data from measurement campaigns [11, p. 4].

The first step improves the temporal correlation by training a neural-network ensemble, described in Chapter 3.2.3, on the wind and turbulence data together with the model, enabling the method to learn consistent relationships over time. Next, daily and seasonal cycles are adjusted using quantile mapping, introduced in Chapter 3.2.1. This step refines long-term patterns while maintaining the temporal structure learned by the neural networks. Finally, sub-hourly variability is refined so that the calibrated dataset reproduces the short-term fluctuations seen in the measurements, based on a comparison with the non-calibrated simulation. [11, p. 5].

4.2.4 Simple Scaling

“Simple Scaling” (SS) is a statistical method based on the linear relationship expressed in Equation 8.

$$f(x) = mx + b$$

Equation 8

f(x):= Predicted Wind Data

m:= Relationship between Measurement and Model Data

x:= Model Wind Data

b:= intercept

This method uses a single relationship between the measurement and model, making the intercept equal to zero.

For each training period, the monthly mean values are calculated for both measured and modeled wind speed and standard deviation by averaging all 10-minute time steps within each month. The ratio between measured and modeled values is then determined using Equation 8 for each month. Since training periods are defined as one year, each training period contains 12 monthly wind speed ratios and 12 monthly standard deviation ratios. These ratios serve as scaling factors to correct the long-term modeled data. For each month, the scaling factors for wind speed and standard deviation are calculated as presented in the Equation 9.

$$R_{WS,i} = \frac{\overline{WS_{meas,i}}}{\overline{WS_{mod,i}}}$$

$$R_{SD,i} = \frac{\overline{SD_{meas,i}}}{\overline{SD_{mod,i}}}$$

Equation 9

R:= Relationship/ Ratio

WS:= Monthly Mean Wind Speed

SD:= Monthly Mean Standard Deviation

i:= MMY, M:=Month Y:=Year

4.2.5 Linear Least Squares

The general concept of “Linear Least Squares” (LLS) is that it fits a straight-line relationship - expressed in Equation 8 - between two the measurements and the models. In this study, two versions of the LLS approach are tested: the standard LLS method and a wind speed binned LLS variant referred to as WLLS.

LLS determines the slope and intercept in Equation 8 so that the fitted straight line matches the measurement data as closely as possible. It does this by minimizing the differences between the measured values and the line, effectively finding the linear relationship that best connects the model output to the observations. Once this relationship is established, the model values can be adjusted using the fitted line: if the model systematically underestimates the measurements, the correction shifts the values upward, while consistent overestimation leads to a downward adjustment. In this way, the linear transformation produces a corrected time series that aligns more closely with the measurements.

Wind-Speed-Binned Linear Least Squares (WLLS) extends the basic LLS approach by not fitting one single linear relationship for the entire dataset, but instead fitting separate linear relationships for different wind speed bins. The modeled data are first grouped into wind speed bins, and for each bin a separate pair of coefficients - m and b - is determined. This means that the correction applied to the model does not follow one global line; instead, it adapts to the behavior of the data within each wind-speed interval.

The idea behind WLLS is that the relationship between modeled and measured values may change with wind speed. As a result, each wind speed bin receives its own fitted correction, which can adjust the model more precisely in ranges where the model consistently over- or underestimates the measurements. The corrected values are then obtained by applying the bin-specific version of Equation 8 to all wind speed and standard deviation data points within that wind speed bin. After predicting the wind speed and its standard deviation, the turbulence intensity is calculated using Equation 1.

4.2.6 Matrix Scaling

Two matrix-based scaling techniques are used to correct modeled wind data: “Monthly Matrix Scaling” (MMS) and “Yearly Matrix Scaling” (YMS). Both methods follow the same core approach but differ in temporal resolution. MMS produces separate scaling matrices for each of the 12 calendar months, while YMS produces a single set of matrices for the entire training year.

Binning Structure

Both methods begin by organizing wind data into a two-dimensional bin grid. Wind speeds are rounded to the nearest whole number to define wind speed bins (wsb), and wind directions are divided into 12 bins of 30° each to define wind direction bins (wdb). Every 10-minute data point in the training period is then assigned to its corresponding wsb - wdb combination.

Ratio Calculation

For each bin, the algorithm calculates the average measured and average modeled values for both wind speed and standard deviation. The ratio of measured to modeled values yields a bin-specific scaling factor. This process produces two scaling matrices per training period: one for wind speed and one for standard deviation.

If a bin contains no data - due to some wind-speed and direction combinations not occurring during the training period - the missing scaling factor is replaced with the average value from all neighboring non-empty bins. This ensures that every possible wind condition is still covered.

Application to Long-Term Data

Once established, the scaling matrices are applied to the full long-term modeled dataset. For each 10-minute time step, the corresponding wind speed bin and wind direction bin are identified. In the case of MMS, the calendar month is used to select the appropriate monthly matrix, while for YMS the appropriate yearly matrix is identified. The retrieved scaling factors are then multiplied with the modeled wind speed and standard deviation to produce the corrected values.

4.3 Post-Processing: The Multi-Tier Correction Concept

The predictions of all algorithms undergo a series of post-processing steps, referred to as tiers, where each tier represents an additional level of refinement. The overall concept is illustrated in Figure 3.

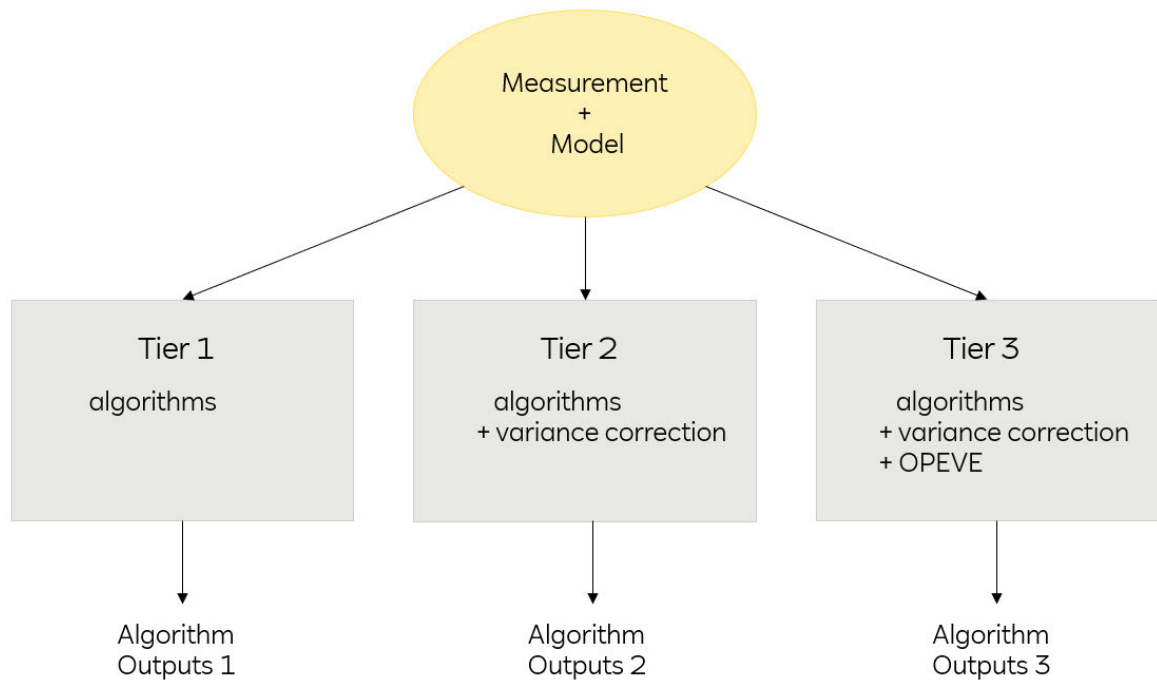


Figure 3 Concept of Multi-Tier Correction

Tier 1 represents the baseline output, where the raw algorithm prediction of turbulence intensity, the concept of which is illustrated in Figure 2, are used without any further adjustment or correction.

Tier 2 builds on Tier 1 by applying variance correction to the predictions. The variance correction addresses a limitation of the base calibration methods and modeled datasets. The algorithms predict the correct average turbulence intensity while producing values that are too clustered, underestimating the variability seen in actual measurements. The correction is applied separately for each month by rescaling the predicted values around the monthly mean to match the variance of the measurements. Monthly resolution is used because turbulence variance is

seasonal, and a single annual correction would fail to capture this behavior. Importantly, the variance correction does not significantly alter the mean and does not consider extreme values.

Tier 3 combines the baseline predictions of Tier 1 with the variance correction of Tier 2 and additionally applying an “Optimised Pareto Extreme Value Extrapolation” (OPEVE), which adjusts the extreme values in the dataset to more accurately reflect the behavior observed in real wind measurement data. OPEVE matches the quantiles of the measured distribution. This means a predicted value at the 95th percentile of the method's tail is replaced with the corresponding 95th percentile value of the measured tail. This correction is particularly important for highly non-linear scaling.

5 Evaluation Results

This chapter presents the evaluation results of all correction methods. The analysis focuses on the metric outputs within the wind speed range from 3 m/s to 25 m/s, as this range contains the largest amount of data. The results are shown separately for each site and each correction tier, making it possible to visualize how the post-processing steps in Tier 1, Tier 2, and Tier 3 influence the final prediction accuracy.

The results are presented in a figure consisting of two diagrams: the upper diagram shows the error of all methods in percent, while the lower diagram highlights which method's error is closest to zero in each wind-speed bin.

5.1 MRBE

Following figures in this subchapter show the MRBE in % over a wind speed bin of each method in each tier for both Site 1 and Site 2.

Site 1

Figure 4 presents the Tier 1 MRBE results for Site 1.

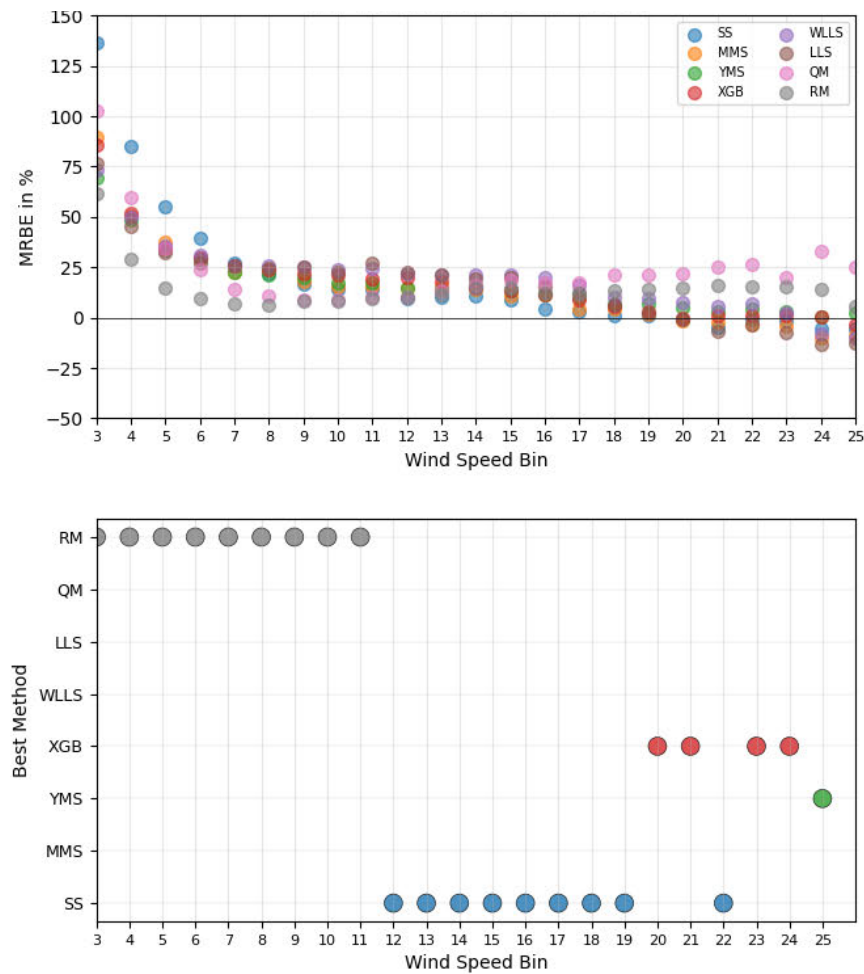


Figure 4 **Site 1**: Tier 1 MRBE in % and Best Method by Wind Speed Bin

In the upper diagram, the MRBE values begin with a very wide positive spread at the lowest wind speed bin. As the bins move from 4 to 6, the MRBE overall decreases rapidly while the spread remains large but starts to narrow, with the data still positive. From around bin 7 onward, the cloud of values becomes increasingly compact. Between the bins 11 and 16, the MRBE values form a tight band with only a few isolated points positioned above or below the main group. At wind speed bin 18, the methods start to spread. In this higher wind speed region, the outliers QM and RM appear above zero, while the cluster remains mainly below zero.

The lower diagram shows which method lies closest to zero for each wind speed bin. RM appears consistently as the method closest to zero across all bins from 3 up to about 11. From 12 to 19, SS becomes dominant, again forming a continuous run. At higher wind-speed bins, the sequence shifts again, with XGB appearing as the best method across several bins. The pattern changes once more, with YMS appearing in the final bin.

Figure 5 presents the Tier 2 MRBE results for Site 1.

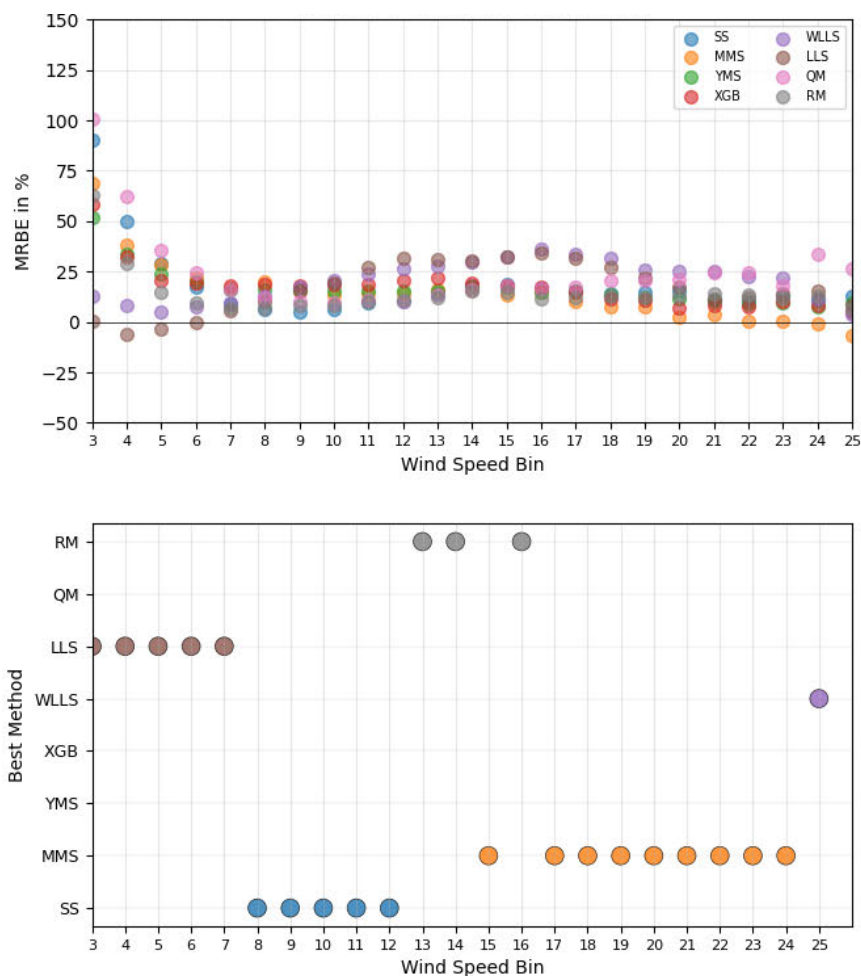


Figure 5 **Site 1**: Tier 2 MRBE in % and Best Method by Wind Speed Bin

In the upper diagram, the MRBE values again begin with a wide positive spread at the lowest wind speed bins. WLLS and LLS lie clearly below the main cluster, with LLS reaching negative

MRBE values. Between bins 4 and 6, the MRBE decreases noticeably while the spread narrows. From wind speed bin 7 onward, the values begin to cluster more tightly. Between bin 11 and 16, the MRBE values form a compact cloud again LLS and WLLS as the positive outlier. Beyond bin 17, the spread widens again, with QM slightly outlying in bins 24 and 25. The error of MMS start to slightly decrease with reaching negative MRBE's from bins 22 onward.

In the lower part of the figure, LLS appears as the best method in the bins from 3 up to 7, forming a continuous sequence. From wind speed bins 8 to 12, SS becomes the best-performing method, again forming an uninterrupted run. At higher wind speed bins, beginning around bin 14, RM emerges as the best method over several consecutive bins. A final shift appears at the highest bins, where MMS becomes the best method for wind speed bins 17 to 24. WLLS appears to be closest to zero at wind speed bin 25.

Figure 6 presents the Tier 3 MRBE results for Site 1.

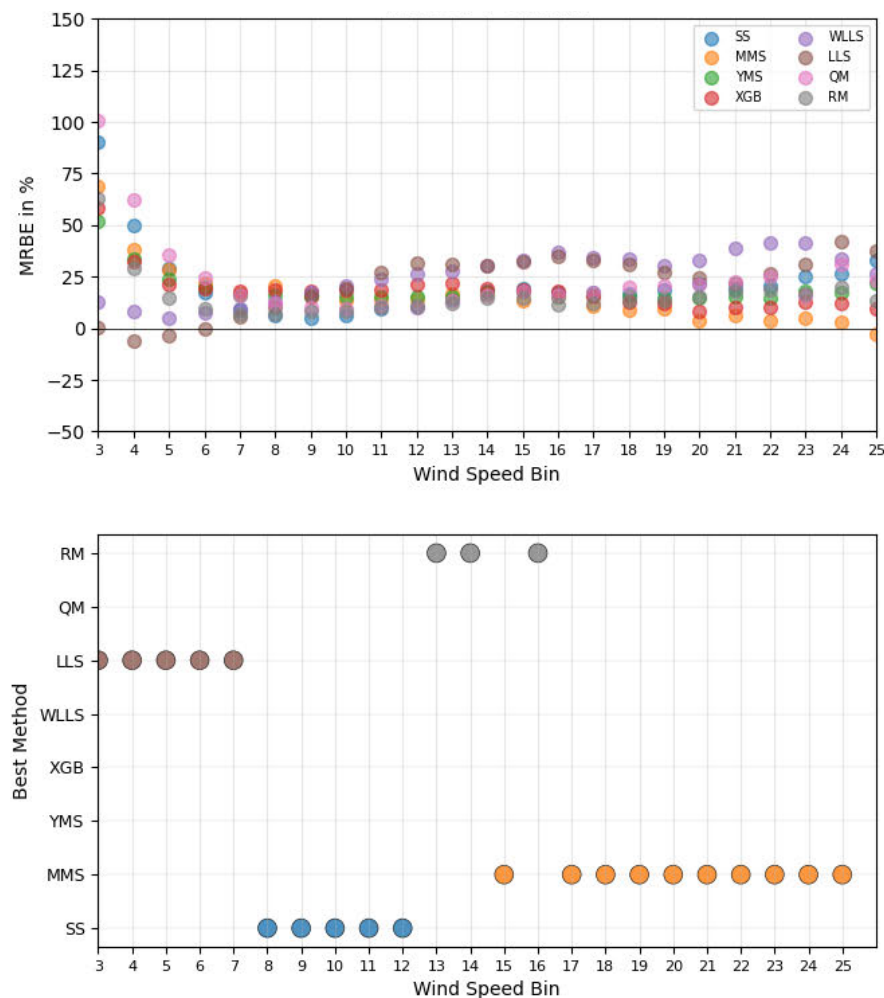


Figure 6 **Site 1:** Tier 3 MRBE in % and Best Method by Wind Speed Bin

The results of Tier 3 are similar to the ones in Tier 2. The MRBE values begin with a wide positive spread at the lowest wind-speed bins with the same outliers as Tier 2. Between bins 4 and 6, the MRBE decreases noticeably, while the spread narrows. From bin 7 onward, the points form a more compact cluster. Starting at bin 11 up until 16, the MRBE values begin to

split, with LLS and WLLS emerging as the positive outliers while the remaining methods stay together as a compact cluster, a pattern that continues up to bin 17. Beyond 17 m/s, the spread increases again and stays mainly positive.

The methods closest to zero follow the same pattern as the ones in Tier 2. LLS appears consistently as the best-performing method across bins 3 to 7. Beginning at bin 8 and continuing through 12, SS becomes the method closest to zero, forming an uninterrupted sequence. RM appears as the best method over several consecutive bins. In the higher wind-speed range from roughly 17 m/s onward, MMS becomes dominant across the bins.

Site 2

Notably, for all MRBE plots at Site 2, certain correction methods are consistently absent across some wind speed bins, particularly at higher wind speeds.

Figure 7 presents the Tier 1 MRBE results for Site 2.

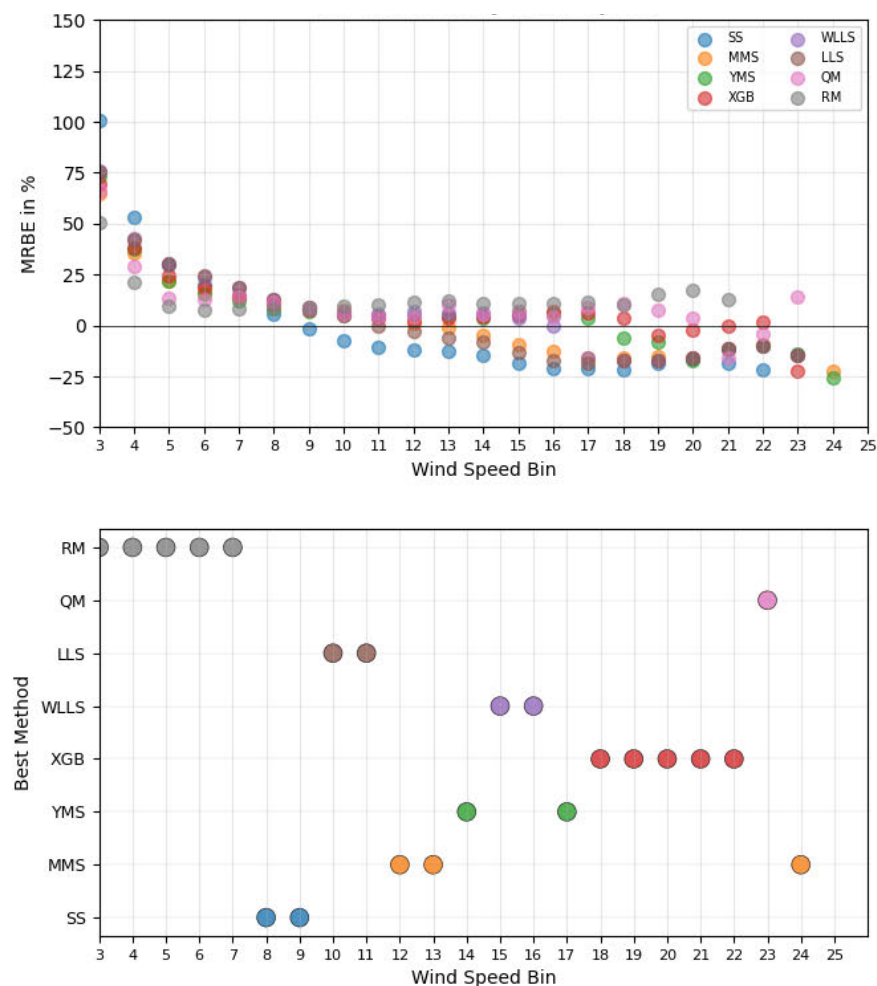


Figure 7 **Site 2:** Tier 1 MRBE in % and Best Method by Wind Speed Bin

In the upper diagram, the MRBE values appear more clustered than in Tier 1 for Site 1, decreasing steadily from wind speed bin 3 until reaching bin 8. SS noticeably underpredicts from bin 9 onwards. At this point, the pattern begins to split: The methods RM, QM and WLLS move

clearly above zero while others drop below, forming two distinct bands. This split persists up to bin 18, after which the separation disappears but the overall spread becomes larger. Toward the highest bins, several methods stop appearing, and at bin 25 no method is listed.

In the lower bins from bin 3 to bin 7, RM consistently appears as the best method. From bin 8 to bin 17, multiple methods take turns being closest to zero, including SS, MMS, YMS, WLLS, and LLS. From bin 18 to bin 22, XGB becomes consistently dominant across several consecutive bins. In the final bins, bin 23 and bin 24, QM and MMS appear once each as the closest-to-zero methods. Compared to Site 1, there are more methods listed as the best method in a wind speed bin.

Figure 8 presents the Tier 2 MRBE results for Site 2.

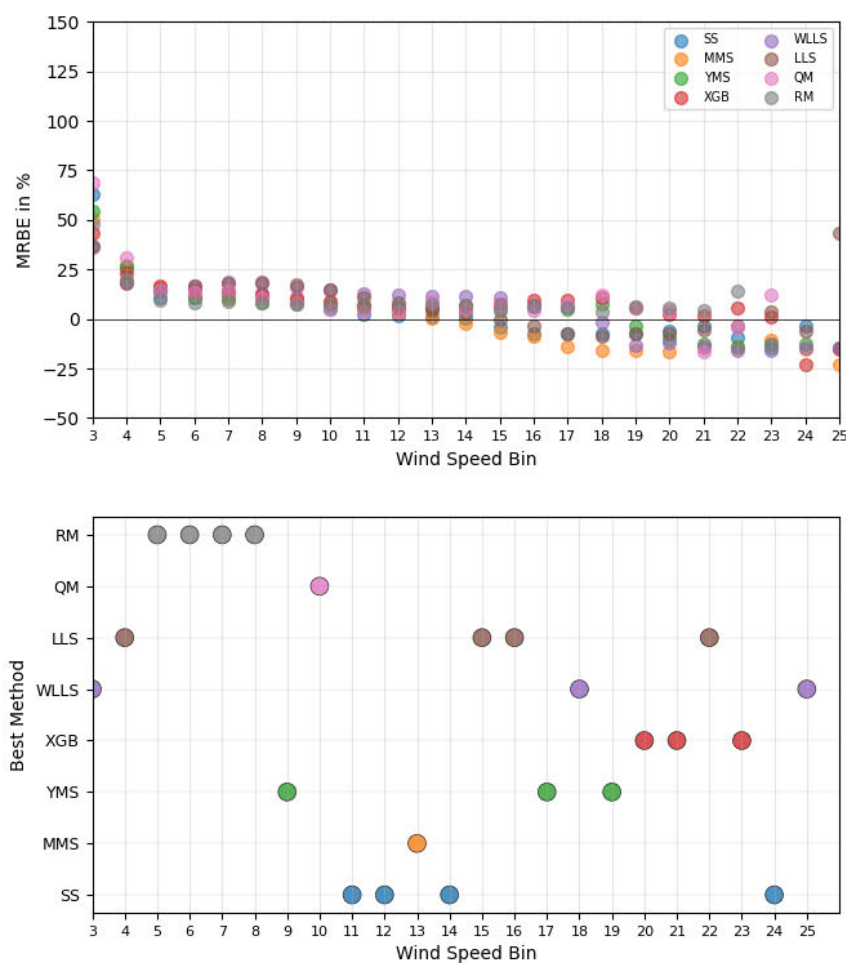


Figure 8 **Site 2:** Tier 2 MRBE in % and Best Method by Wind Speed Bin

The overall MRBE curve for Tier 2 appears more compressed compared to Tier 1, with the high errors still occurring at bin 3 but the overall curve looking considerably more constant across the wind speed range. The MRBE of the methods remain above 0% starting to reach negative values as the spread increases at bin 12. At the highest wind speed bin only QM, WLLS, LLS, and MMS remain listed.

Compared to Tier 1, the number of the best methods over the wind speed bins is larger. As in Tier 1, there is no consistently best method over a larger wind-speed range, and the best-performing method changes from bin to bin. Consistent with Tier 1, RM again achieves the lowest MRBE at some of the lower wind speed bins.

Figure 9 presents the Tier 3 MRBE results for Site 2.

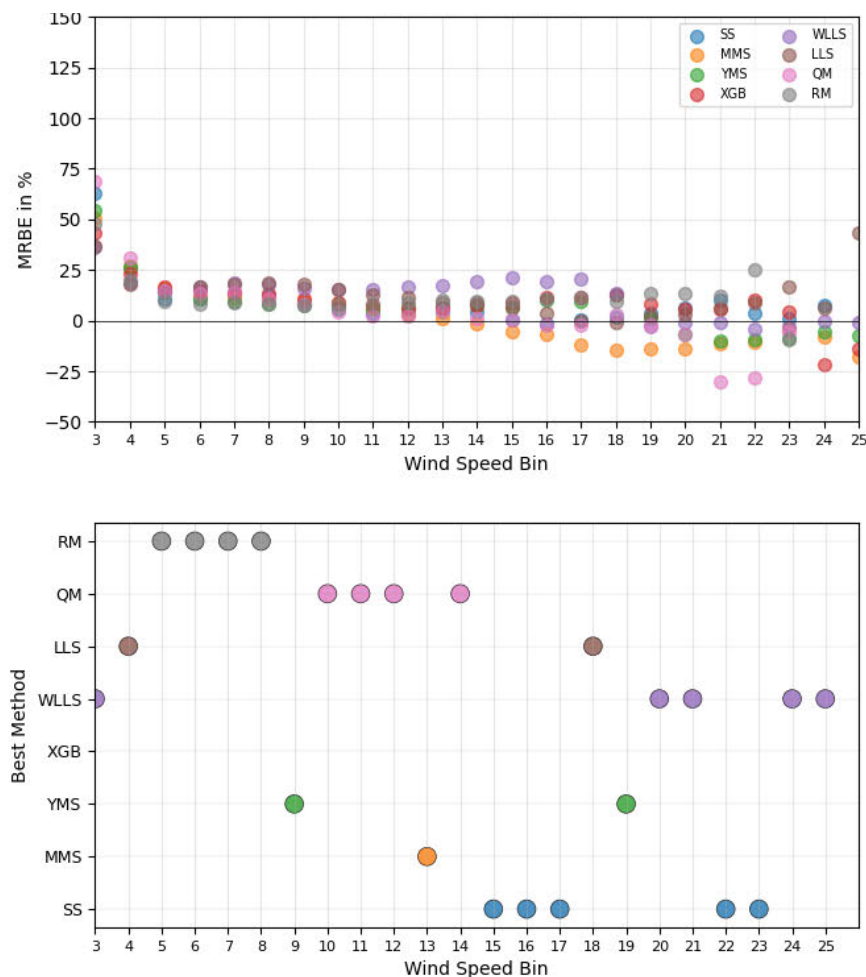


Figure 9 **Site 2**: Tier 3 MRBE in % and Best Method by Wind Speed Bin

The Tier 3 MRBE results for Site 2 show that the largest errors occur at wind speed bin 3, after which the values decrease rapidly and remain relatively stable between bins 5 and 10. Starting at bin 11, the values begin to scatter more noticeably, and from bin 13 onward the first negative MRBE values appear. Between bins 13 and 17, and again at bin 25, WLLS appears as an outlier above the main cluster, while MMS shows outlying behavior between bins 17 and 19.

Tier 3 does not reveal one consistently best-performing method across all bins. Overall, RM tends to perform best at lower wind speeds, followed by QM, which remains close to zero through the lower-to-mid range. SS becomes the method closest to zero from mid to higher wind speeds, while at the highest wind speeds WLLS emerges as the closest-to-zero method.

5.2 RRMSE

Following figures in this subchapter show the RRMSE in % over a wind speed bin of each method in each tier for both Site 1 and Site 2.

Site 1

Figure 10 presents the Tier 1 RRMSE results for Site 1.

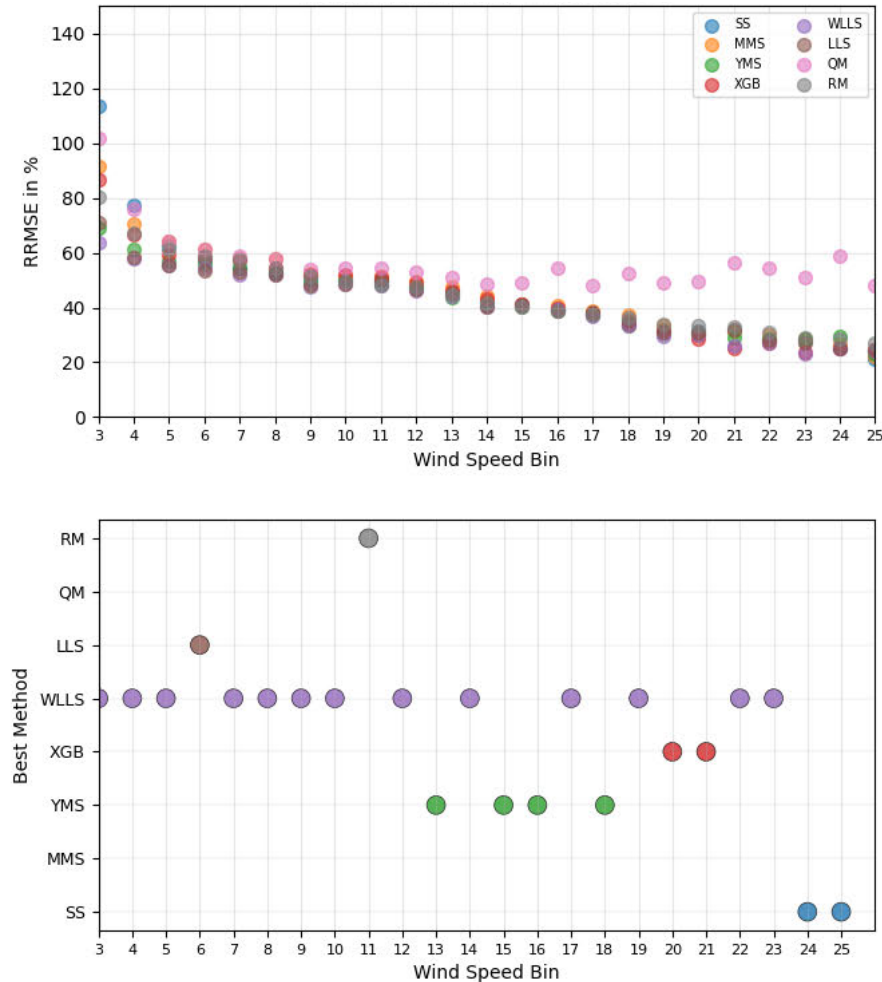


Figure 10 **Site 1**: Tier 1 RRMSE in % and Best Method by Wind Speed Bin

The RRMSE in Tier 1 follows the same general trend as the MRBE, with average errors being highest at low wind speeds and decreasing as wind speed increases. The spread in wind speed bin 3 is the highest ranging from 60% to almost 120%, making the overall spread between methods in increasing bins is relatively small. From bin 15 onwards, QM stands out by remaining relatively constant while all other methods continue to decrease.

Regarding best method performance, WLLS achieves the lowest RRMSE across the majority of wind speed bins, although there are some individual bins where other methods perform better. A notable exception is the mid wind speed range between bins 13 to 18, where YMS occasionally dominates. At the highest wind speeds, between 24 and 25, SS achieves the lowest RRMSE.

Figure 11 presents the Tier 2 RRMSE results for Site 1.

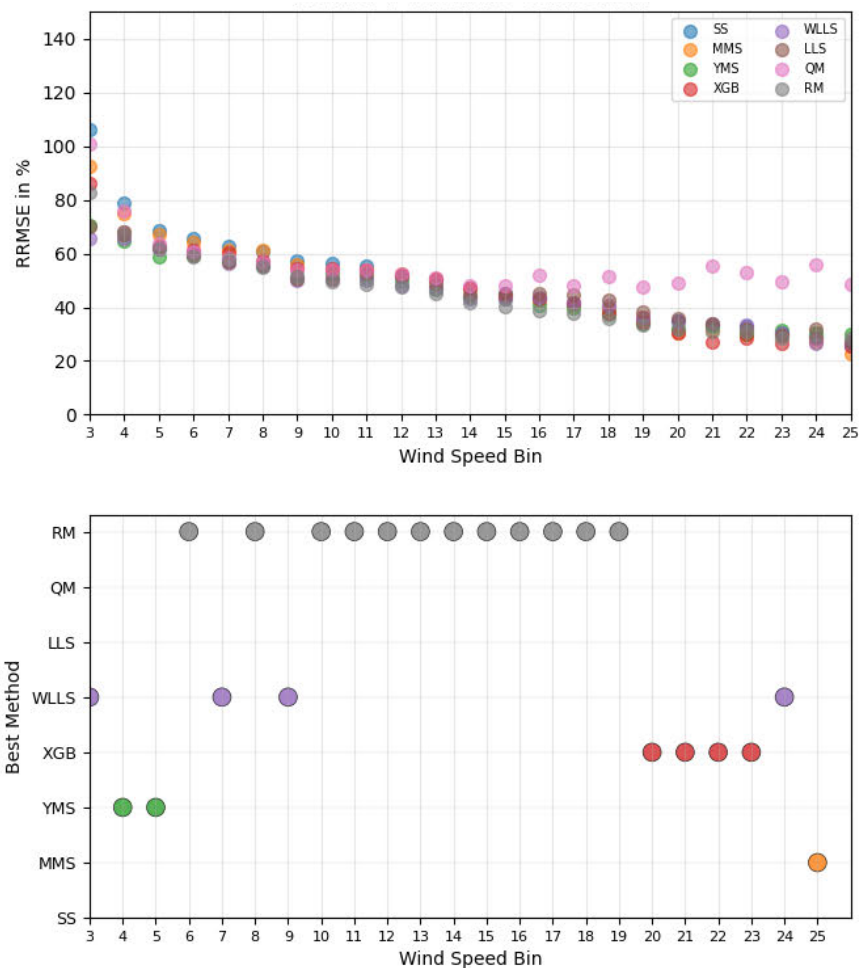


Figure 11 **Site 1**: Tier 2 RRMSE in % and Best Method by Wind Speed Bin

Moving to Tier 2, the RRMSE diagram remains similarly clustered compared to Tier 1, with the decrease across wind speeds being gradual rather than abrupt. QM again shows the same behavior as in Tier 1, remaining relatively constant from 15 onwards while all other methods continue to decrease.

At lower wind speeds, YMS and WLLS stand out with the lowest RRMSE values, suggesting these methods handle the high variability in this range most effectively. From 10 to 19, RM takes over as the dominant method, achieving the lowest RRMSE across this range. At higher wind speeds between 20 and 23, XGB delivers the best performance.

Figure 12 presents the Tier 3 RRMSE results for Site 1.

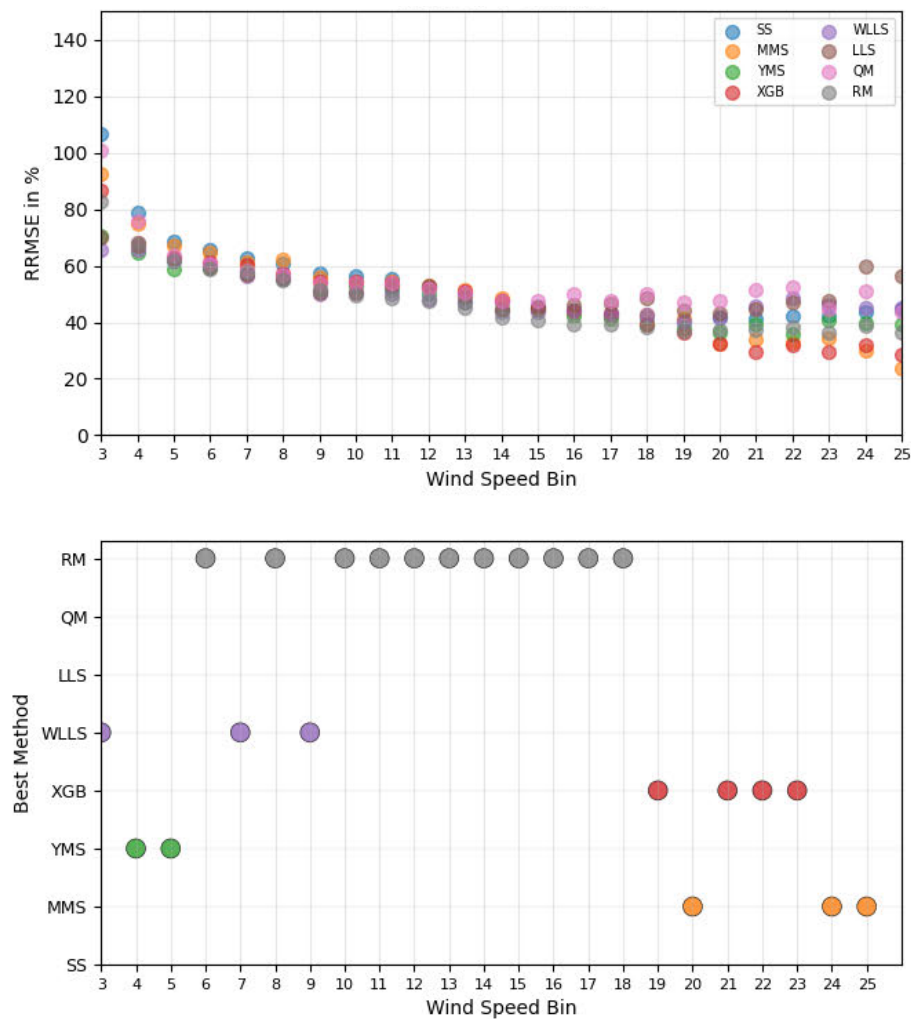


Figure 12 **Site 1**: Tier 3 RRMSE in % and Best Method by Wind Speed Bin

The methods spread in the first bin ranging from approximately 60% to 110% and become more clustered onward at the other lower wind speeds, while the spread between methods increases at wind speed bin 19 onward.

At low wind speeds, YMS and WLLS again achieve the lowest RRMSE, mirroring the pattern observed in Tier 2. From bins 10 to 18, RM strengthens its performance and achieves the lowest RRMSE across this range, building on the dominance already seen in Tier 2. Beyond bin 18, XGB and MMS take over as the best performing methods at the higher wind speeds.

Site 2

Figure 13 presents the Tier 1 RRMSE results for Site 2.

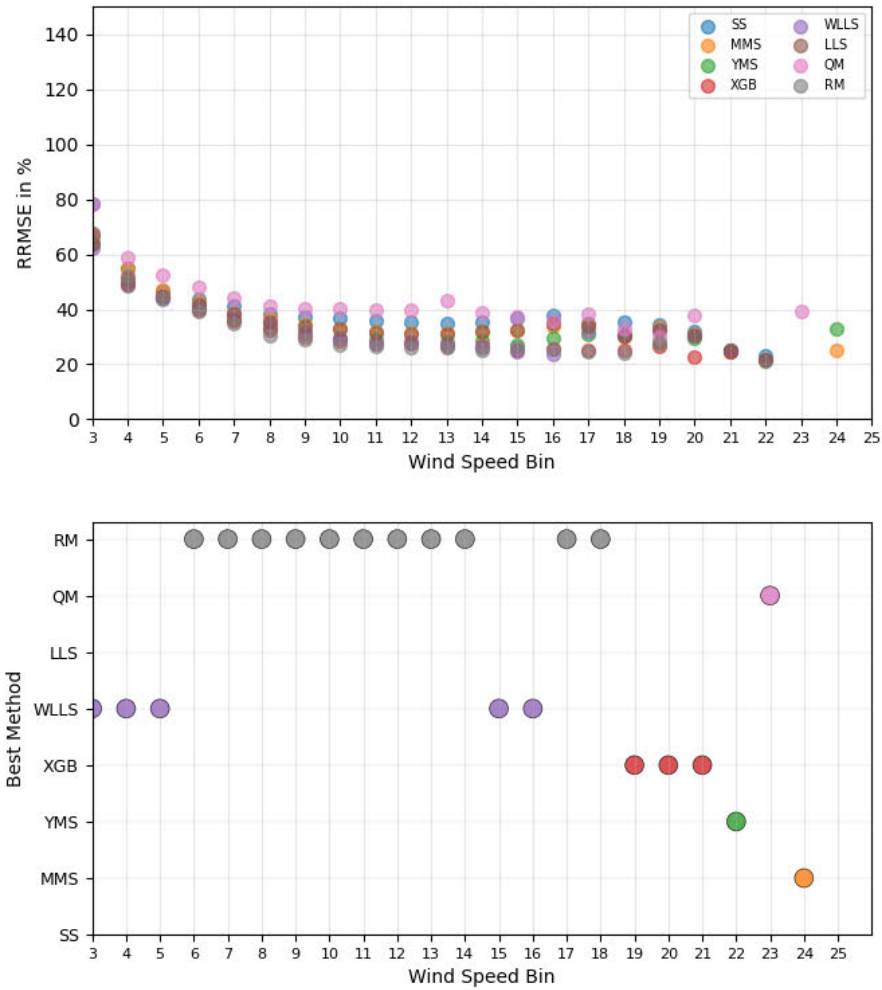


Figure 13 **Site 2**: Tier 1 RRMSE in % and Best Method by Wind Speed Bin

Turning to Site 2, the RRMSE results for Tier 1 appear more constant compared to Site 1 Tier 1, with only a relatively slight decrease visible across the wind speed range rather than the more pronounced decrease. Noticeably, QM stays somewhat above the cluster throughout the majority of the wind speed bins. In higher wind speeds, methods are missing, just like for MRBE. Only this time, the methods missing differ. QM remains in bin 23, while in bis 24 and 25 only MMS and YMS are listed. There is no method listed in bin 25.

At the lowest wind speeds between bins 3 and 5, WLLS achieves the lowest RRMSE, however unlike Site 1 where WLLS performed well across a broad wind speed range, its advantage here is limited to this low wind speeds and bins 15 and 16 only. From bins 6 to 14 and 17 to 18, RM takes over and maintains the lowest RRMSE across this extended range, making it the most dominant method in Tier 1 at Site 2. At higher wind speeds, XGB occasionally achieves the lowest RRMSE in some bins. At higher bins 23, 24, and 25 the methods that were remaining in the upper diagram are listed as best methods: YMS, QM, and MMS.

Figure 14 presents the Tier 2 RRMSE results for Site 2.

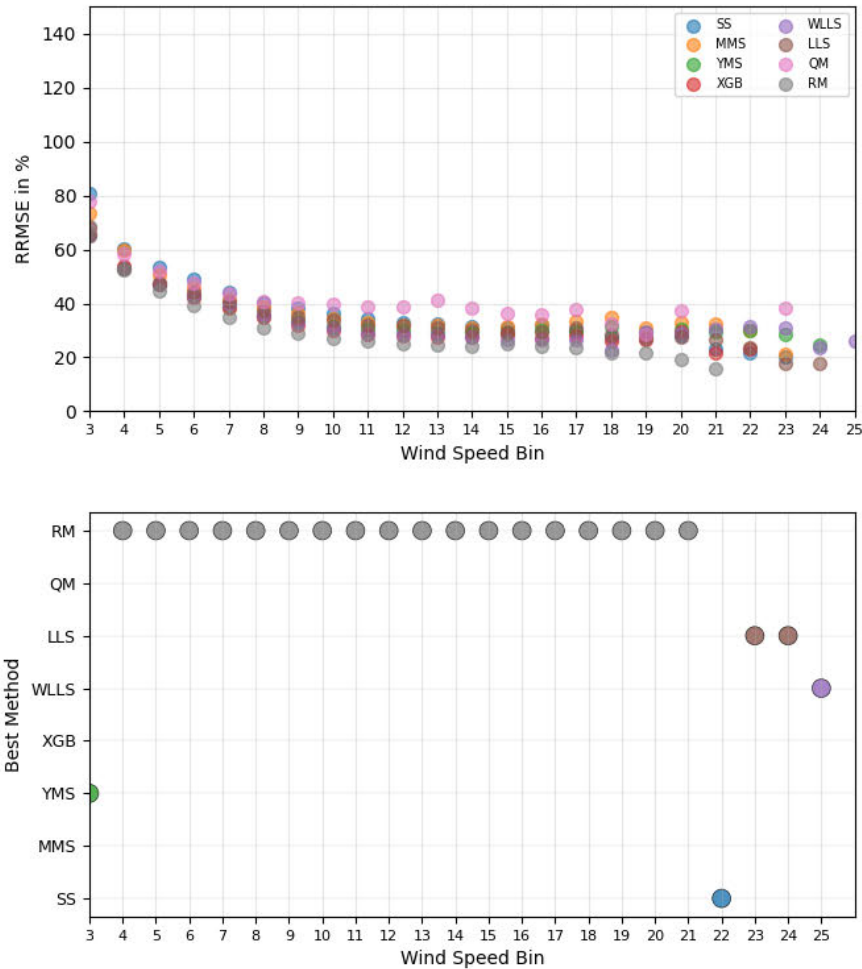


Figure 14 **Site 2:** Tier 2 RRMSE in % and Best Method by Wind Speed Bin

In Figure 14, the methods are closely clustered at lower wind speeds and begin to flare outward as wind speed increases, with the spread between methods becoming more visible at higher wind speeds. From bin 12 onwards, the RRMSE values stabilize, and the trend remains relatively constant across the remaining wind speed range. QM again stays somewhat in the upper part of the cluster in most wind speed ranges. Again, the number of methods decrease with higher wind speeds with YMS, WLLS, and LLS listed in bin 24, and WLLS as the only method remaining in bin 25.

Similar to Tier 2 at Site 1, RM emerges as the most dominant method, however its dominance at Site 2 is considerably stronger and more consistent. Except for bins 3 and from 22 onwards where other methods occasionally perform better. SS appears in bin 22, LLS achieves the lowest RRMSE across bins 23 and 24, and as WLLS is the only method remaining in 25, it is the best performing method listed in that bin.

Figure 15 presents the Tier 3 RRMSE results for Site 2.

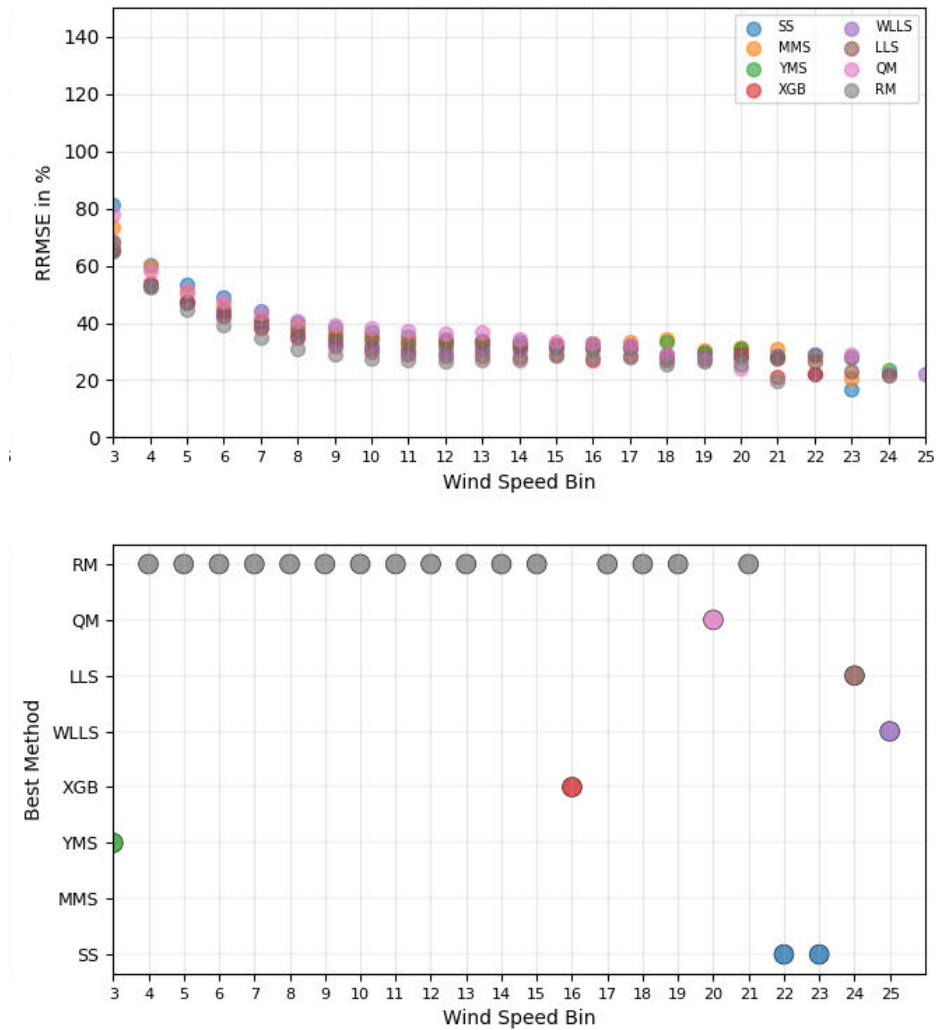


Figure 15 **Site 2:** Tier 3 RRMSE in % and Best Method by Wind Speed Bin

The Tier 3 RRMSE results for Site 2 follow the same pattern as the previous tiers, with errors decreasing across the wind speed range before stabilizing at a relatively constant level. Towards higher wind speeds, the number of methods listed is reduced to only LLS and YMS listed in bins 23 and 24, and WLLS as the only remaining method in bin 25.

RM remains the dominant method throughout the majority of wind speed bins, consistent with its strong performance in Tier 2. YMS takes over at the lowest wind speeds, and the dominance of RM gradually weakens towards the higher end of the wind speed range where other methods begin to occasionally outperform. Overall, the consistent dominance of RM across all three tiers at Site 2 for RRMSE stands out as one of the clearest findings at this site.

5.3 DEL

Following figures in this subchapter show the DEL in % over the wind speed bins of each method in each tier for both Site 1 and Site 2. In addition, a diagram of the DEL interannual variability is shown for each site.

Site 1

Figure 16 shows the interannual variability of DEL for Site 1, spanning the years 2004 to 2007.

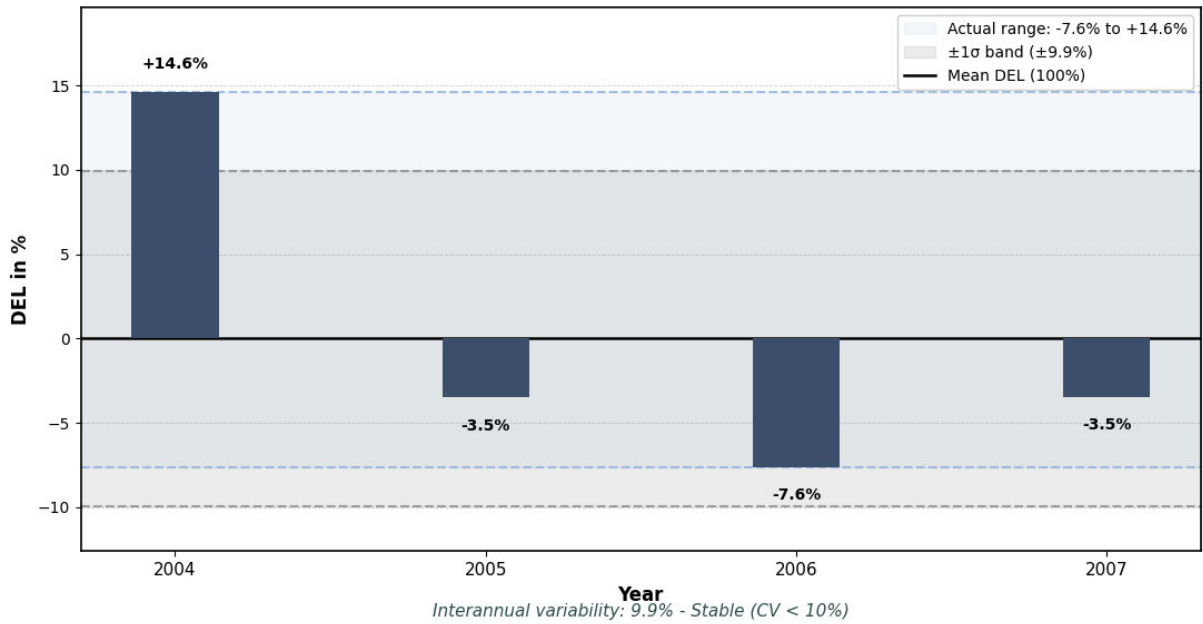


Figure 16 **Site 1**: Interannual Variability of DEL in %

The annual DEL values deviate from the long-term mean by +14.6% in 2004, -3.5% in 2005, -7.6% in 2006, and -3.5% in 2007. The actual range band spans from -7.6% to +14.6%, and the standard deviation ($\pm 1\sigma$) band covers $\pm 9.9\%$ around the mean. The overall coefficient of variation is 9.9%, classifying the site as having stable interannual variability (CV < 10%).

Figure 17 presents the Tier 1 DEL results for Site 1.

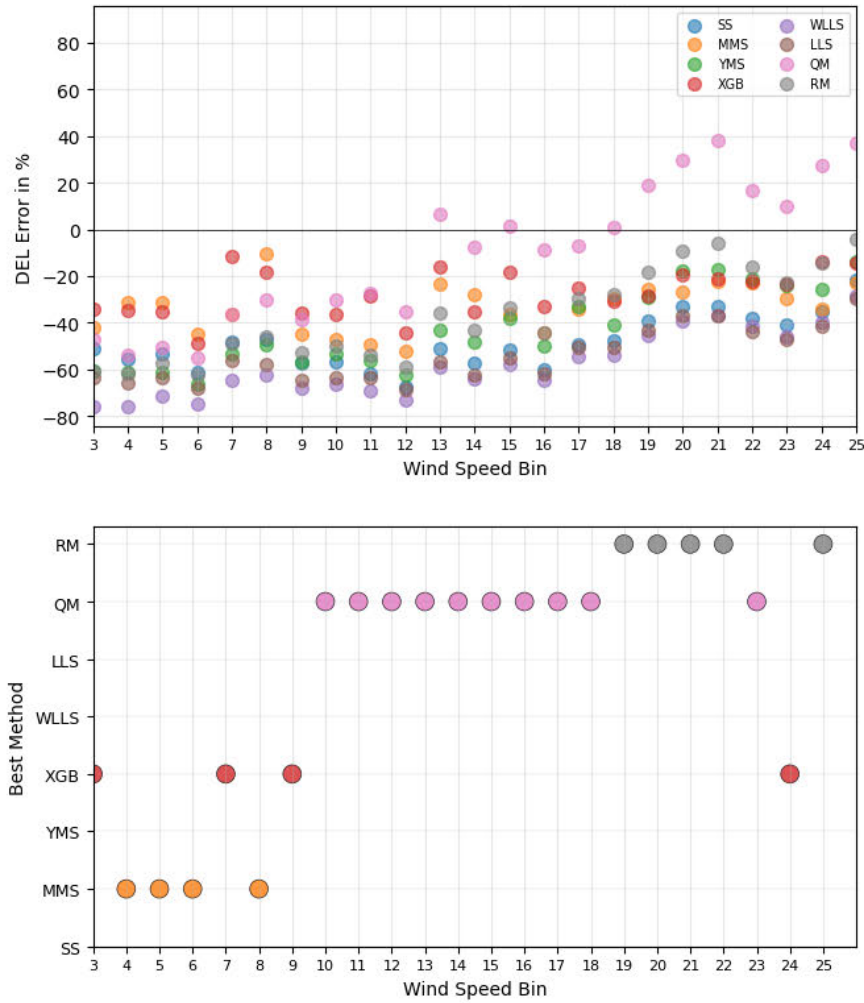


Figure 17 **Site 1:** Tier 1 DEL Error in % and Best Method by Wind Speed Bin

The DEL results for Site 1 Tier 1 are predominantly negative across all wind speed bins, indicating that methods systematically underestimate the fatigue-relevant turbulence content, consistent with what was observed at Site 1. As wind speed increases, the DEL values for most methods tend to converge closer to 0%, suggesting improved accuracy at higher wind speeds. The clear outlier is QM, whose DEL increases rapidly towards positive values, reaching nearly +40% at higher wind speeds, making it the only method that consistently overestimates fatigue loads in this range.

Regarding best method performance, XGB and MMS achieve the lowest DEL errors at lower wind speeds. From bins 10 to 18, QM takes over as the best performing method, after which RM becomes dominant at the higher wind speeds.

Figure 18 presents the Tier 2 DEL results for Site 1.

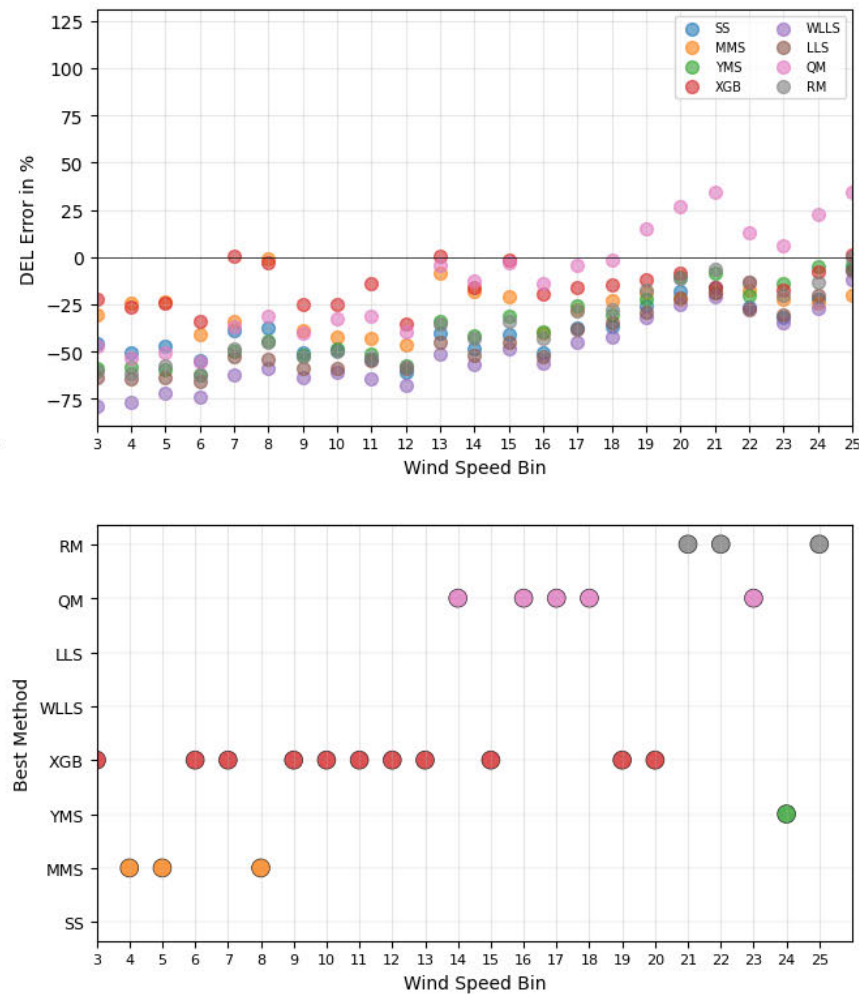


Figure 18 **Site 1:** Tier 2 DEL Error in % and Best Method by Wind Speed Bin

In Figure 18, results show that all methods follow an increasing trend towards 0% as wind speed increases, this time with a slightly higher magnitude compared to Tier 1. In lower wind speed bins XGB is a positive outlier tending to be closest to 0%. QM remains a persistent outlier, reaching noticeably positive DEL values at higher wind speeds, consistent with its behavior in Tier 1.

At lower wind speeds, MMS and XGB overlap closely, both achieving the lowest DEL errors in this range. Moving into the mid wind speed range, XGB becomes the dominant method and briefly overlaps with QM before separating again. At higher wind speeds, RM begins to appear as the best performing method across three bins, while YMS occasionally takes over at bin 24.

Figure 19 presents the Tier 3 DEL results for Site 1.

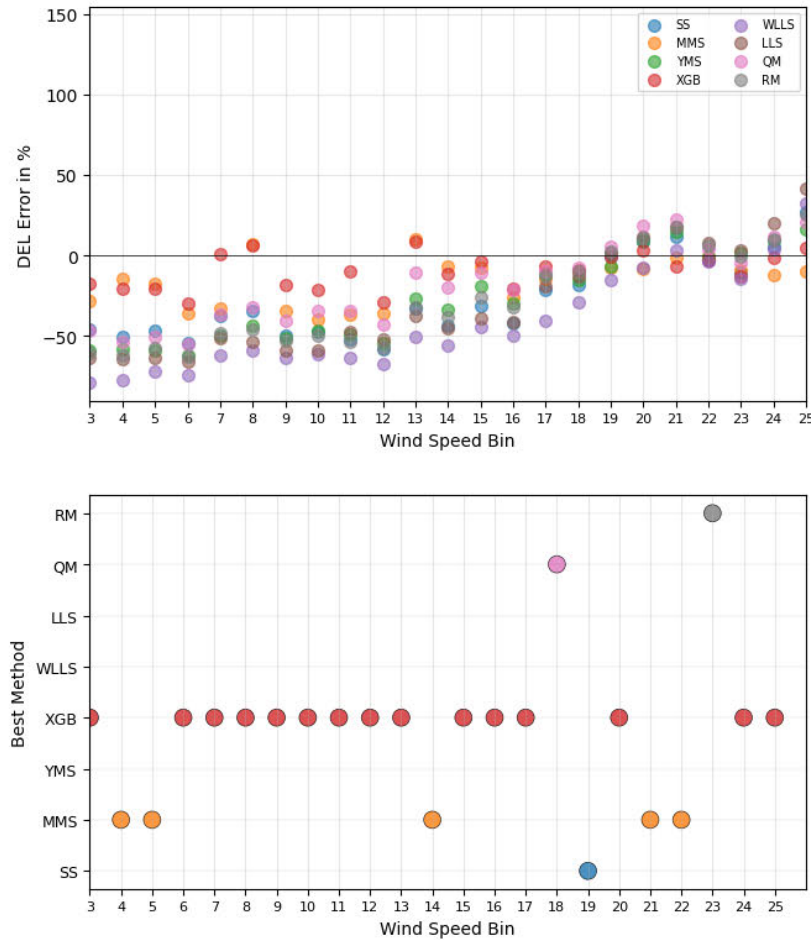


Figure 19 **Site 1:** Tier 3 DEL Error in % and Best Method by Wind Speed Bin

The Tier 3 DEL results show the same upward trend towards 0% as wind speed increases, consistent with Tier 1 and Tier 2. The overall magnitude increases further compared to the previous tiers, with more methods reaching positive DEL values at higher wind speeds. The spread of the methods appears to be highest in lower wind speeds while at higher wind speeds starting from bin 16 the methods form a cluster around a DEL value. This cluster remains consistent until the highest bin where the methods begin to spread slightly from one another.

Compared to Tier 1 and Tier 2, the arrangement of best performing methods across the wind speed range is more stable, with less variation depending on the wind speed bin. XGB stands out as the consistently dominant method throughout the entire wind speed range, achieving the lowest DEL errors across the majority of bins. Notably, MMS appears to have a DEL closest to zero in bins 4 to 5, 14, and 21 to 22.

Site 2

Figure 20 shows the interannual variability of DEL for Site 2, spanning the years 2010 to 2018.

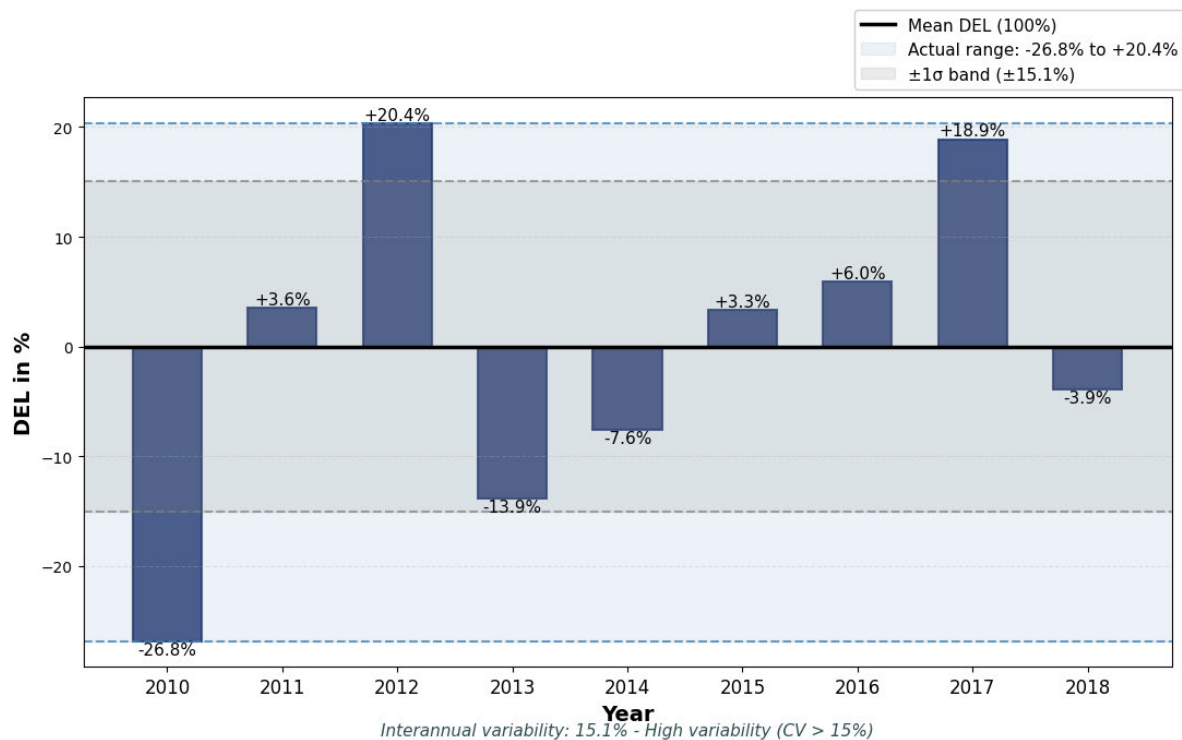


Figure 20 **Site 2:** Interannual Variability of DEL in %

For Site 2, the DEL varies across the years 2010 to 2018, with the largest positive deviations being +20.4% in 2012 and +18.9% in 2017, and the largest negative deviations being -26.8% in 2010 and -13.9% in 2013. The remaining years show more moderate deviations, ranging between -7.6% and +6.0%. The standard deviation ($\pm 1\sigma$) band covers $\pm 15.1\%$ around the mean, and the actual range spans from -26.8% to +20.4%. The overall coefficient of variation is 15.1%, classifying the site as having high interannual variability (CV > 15%).

Figure 21 presents the Tier 1 DEL results for Site 2.

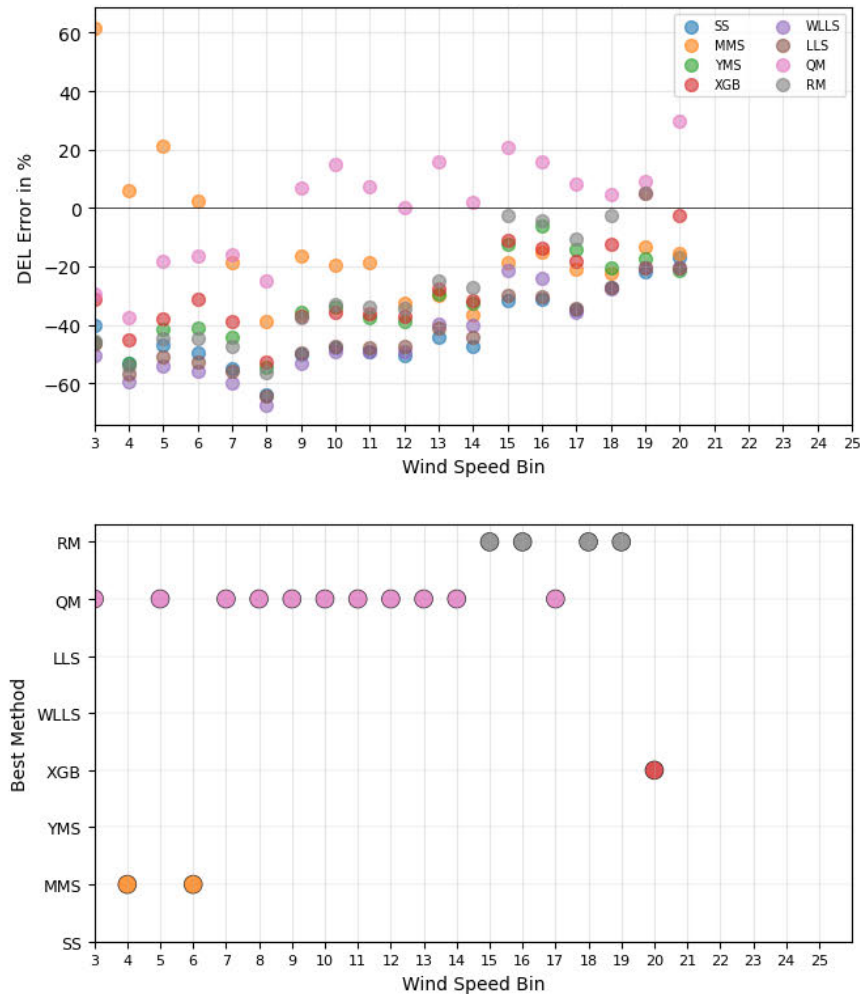


Figure 21 **Site 2:** Tier 1 DEL Error in % and Best Method by Wind Speed Bin

The DEL values for all methods are mostly negative throughout the wind speed range. At the lowest wind speeds, the methods are more widely spread but are still negative. Notable is that MMS is high in lower wind speeds, reaching +60% in bin 3, and presenting an outlier but decreasing towards the other method starting from bin 6 resulting in negative DEL values. As wind speed increases, the curves generally move closer to 0%, although the spread between the methods remains visible across the entire range. At higher wind speeds, several methods show a slight upward movement toward 0%, but the negative DEL error persists for nearly all methods. One exception is visible at the mid wind speed range, where the QM curve increases more steeply than the others and crosses into positive DEL error values. Bins 21 to 25 are empty with no method listed.

Different methods appear as the best performer at isolated bins. Looking at the lower diagram, QM dominates across most wind speed bins, except for bins 4 and 6 where MMS appears and at higher wind speeds, RM appears at bins 15, 16, 18, 19 and XGB appears once at bin 20. Naturally, no method performs in bin 21 to 25.

Figure 22 presents the Tier 2 DEL results for Site 2.

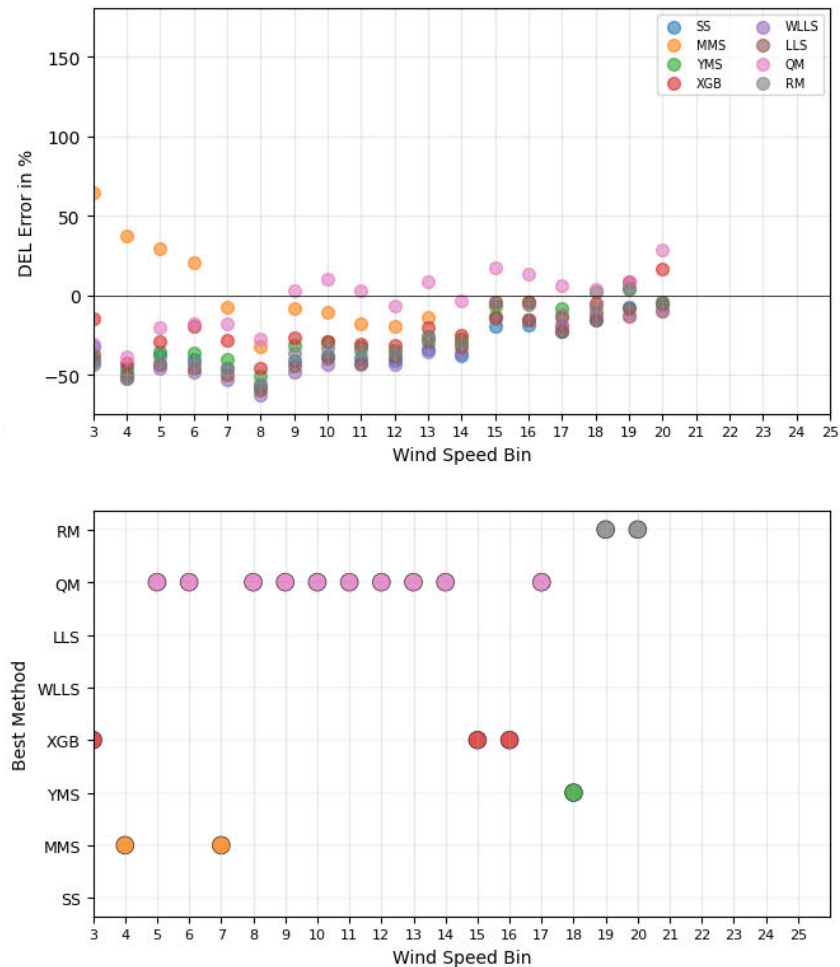


Figure 22 **Site 2:** Tier 2 DEL Error in % and Best Method by Wind Speed Bin

Looking at the upper diagram in Figure 22, the DEL error values for all methods are mostly negative throughout the wind speed range. At the lowest wind speeds, the methods are relatively narrowed, with MMS being an exception, performing a positive outlier in bins 3 to 6 before reaching negative DEL values. As wind speed increases, the values generally move closer to 0%. Starting at bin 9, QM becomes a slight outlier with DEL increasing and reaching positive values. At higher wind speeds, several methods show a slight upward movement toward 0% and beyond, but negative DEL error persists for nearly all methods. Again, no methods appear in bins 21 to 25.

In lower wind speeds, XGB and MMS show a DEL value closest to zero. However, QM dominates across most wind speed bins especially in the mid-range. Some algorithms such as XGB, RM, and YMS have a DEL closer to 0 % in bins 15 to 21. There is no best performing method in bins 21 to 25.

Figure 23 presents the Tier 3 DEL results for Site 2.

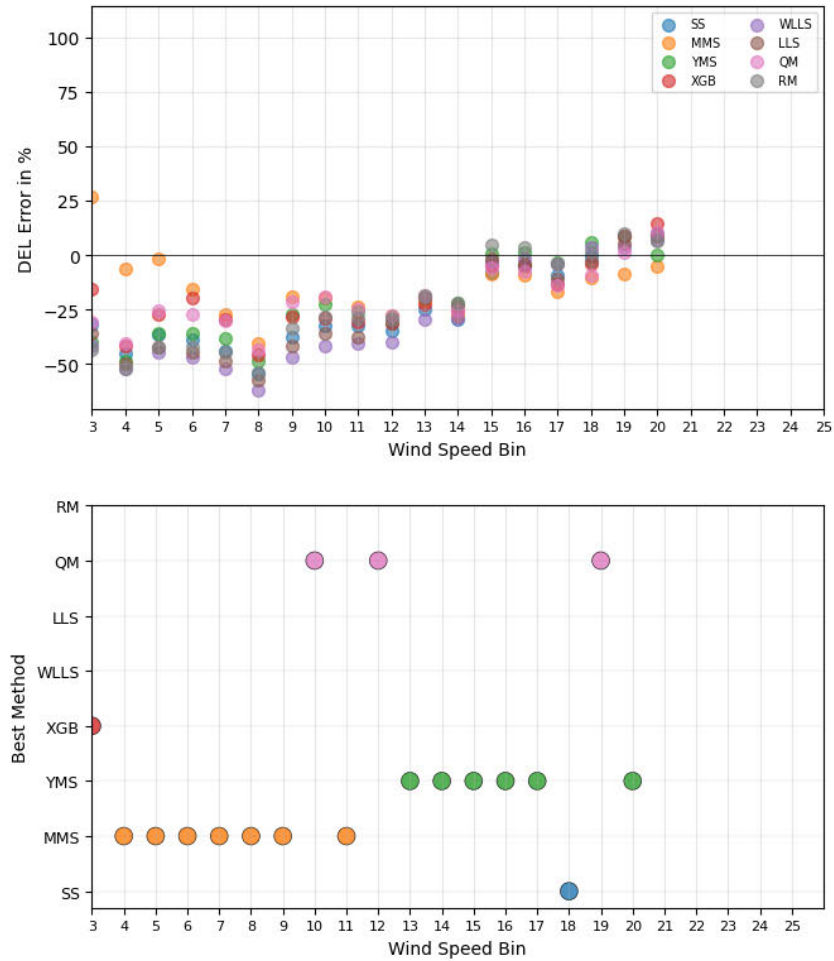


Figure 23 **Site 2:** Tier 3 DEL Error in % and Best Method by Wind Speed Bin

In Tier 3 in Figure 23, MMS starts at a lower DEL compared to Tier 1 and Tier 2 but still presents a positive outlier, then decreases until bin 8. While the spread of the methods is relatively higher than in the rest of the wind speed bins, the DEL values are more compressed across methods compared to Tier 1 and Tier 2. From bins 15 to 20, values are relatively close to 0%. The overall trend is similar to the other tiers, with an overall increase from negative values to values closer to zero. There is no performing method in bins 21 to 25.

In the lower diagram, XGB is the best performing method in bin 3, followed by MMS dominating from bins 4 to 7. From mid to high wind speeds, YMS dominates, with exceptions at bins 10, 12 and 19 where QM performs best. Only in bin 18 SS has the DEL closest to 0%. Bins 21 to 25 remain empty.

6 Discussion

The following discussion reflects on the main findings of this study and connects the results back to the overall goal of predicting long-term turbulence intensity from short-term measurements. After evaluating a range of statistical, machine-learning, and remodelling methods at two fundamentally different sites (one offshore and one onshore), patterns and tendencies become visible across metrics, and tiers. This chapter discusses the most relevant behaviors across methods and sites, and places the results into a broader context to show what they mean for long-term TI prediction.

Site-Specific Differences

The two sites differ strongly in their wind speed distribution. Site 1, the offshore site, covers a wider range of wind speeds and includes high wind speed conditions, as described earlier in Chapter 3.1. Site 2, the onshore site, reaches lower maximum wind speeds, which leads to fewer data points in the high-wind bins. Because of this, several methods simply do not appear in the highest bins of the metrics at Site 2. This is not a method issue but a direct result of the more limited wind speed range in onshore locations.

The sites also differ in their natural year-to-year variability, shown through DEL in Figure 16 and Figure 20. At Site 1, the annual DEL values vary by about $\pm 9.9\%$, meaning that choosing one year or another as a training period - the current industry practice - already introduces an uncertainty of around 10%. Any method with DEL errors consistently below this $\pm 9.9\%$ range can therefore be considered an improvement over relying on a single year of data. At Site 2, the natural variability is much higher, around $\pm 15.1\%$. As explained in Chapter 3.1, onshore turbulence is generally harder to predict because of more irregular and complex wind conditions. This helps explain why Site 2 shows larger year-to-year differences. As a result, the reference threshold for meaningful improvements is higher here: only methods with DEL errors below $\pm 15.1\%$ truly outperform the natural variability of the site.

Site 1 offers only four years of data, while Site 2 offers nine years. Having shorter or longer measurements directly influences the interannual variability. Each year holds a significant weight. For Site 1, each individual year carries a larger weight on the overall results, making the interannual variability less statistically robust. When longer measurements are available, as in Site 2, results are more reliable since the influence of any single year on the overall interannual variability is reduced.

Comparison of the Tiers

Across all metrics, the behavior of the tiers can be explained directly by the structure of the multi-tier correction concept introduced in Chapter 4.3. Tier 1 represents the raw model-meas-

urement calibration of each method; therefore, the results primarily reflect the inherent tendencies of the individual algorithms, including wide spreads, method-specific biases, and strong deviations at the lower wind-speed bins. When Tier 2 applies the variance correction, the predictions are rescaled around the monthly means to match the measured variability. This step specifically targets the issue that most base calibration methods produce values that are too clustered compared to real measurements. As a result, Tier 2 reduces the overly narrow variance in MRBE and RRMSE at low and mid wind speeds and shifts DEL values closer to zero, where the true fatigue-relevant variability lies. Because the correction only adjusts the spread but not the mean, methods in MRBE can still appear more positive or negative in the mid-range, depending on how strongly their original predictions deviated, which explains the observed slightly upward shifts in MRBE.

Tier 3 introduces the OPEVE correction, which directly adjusts the extreme values of the prediction to better match the measured distribution. This explains why Tier 3 affects mainly the upper wind-speed bins: at higher wind speeds, the influence of extreme turbulence becomes stronger, and OPEVE replaces the uncorrected tail behavior from Tier 2 with measurement-based extremes. In MRBE, this leads to stronger spreads because small differences in the extremes can visibly shift the bias. In RRMSE, the effect remains smaller because the metric focuses on overall error magnitude rather than direction - but the tail corrections still cause slight increases in MRBE error at the highest bins. For DEL, Tier 3 has the clearest and most consistent impact. Because DEL amplifies the effect of extremes through the Wöhler exponent, improving the alignment of the predicted tails directly reduces the underestimation that dominates Tier 1 and Tier 2. As a result, DEL error values move steadily toward zero with each tier, demonstrating exactly the improvement Tier 3 is designed to produce.

Overall, the way the tiers influence the results follows naturally from their purpose: Tier 1 reflects the base methods, Tier 2 corrects the spread to match the natural monthly variability of turbulence, and Tier 3 corrects the extremes to match the observed tail behavior. The differences seen across MRBE, RRMSE, and DEL therefore arise not from inconsistencies in the methods, but from the specific aspect of the signal each tier is designed to correct.

The influence of the tier corrections differs between the two sites because of their underlying wind speed characteristics. At the offshore site, the wind speed range is broad and well populated across all bins, which gives both the variance correction in Tier 2 and the extreme-value adjustment in Tier 3 enough information to work as intended. With sufficient data in each bin, Tier 2 can correct the overly narrow spread of the baseline predictions, and Tier 3 can meaningfully adjust the distribution tails where extreme turbulence occurs. This is why the offshore site shows clearer and stronger improvements from one tier to the next, particularly in the higher wind-speed bins where turbulence extremes become more important.

At the onshore site, the wind-speed range is much narrower, and several higher bins contain few or no data points. Because Tier 2 relies on capturing monthly variance and Tier 3 relies on matching the measured tail behavior, both corrections depend on bins being sufficiently populated. When this is not the case, the impact of the corrections becomes weaker, and the differences between tiers remain much smaller. The more irregular and variable onshore wind conditions further limit how much the tiers can adjust the predictions. As a result, the tier corrections have a noticeably stronger effect offshore, while at the onshore site the influence is weak simply because the data provide less information for the corrections to act on.

Method Behavior

Across all methods, errors are highest at low wind speeds. As described in Chapter 3.1, thermal turbulence dominates at low wind speeds and therefore is more difficult to predict regardless of the method applied.

RM demonstrates one of the most stable behaviors across the evaluation, which aligns with its design described in Chapter 4.2.3. By combining a neural-network ensemble with distribution- and variability-based adjustments, the method is less dependent on individual wind-speed bins and can capture the overall model–measurement relationship more consistently. This is reflected clearly in the results: RM frequently achieves the lowest MRBE and RRMSE errors across long sequences of wind-speed bins, particularly in the lower and mid ranges at both sites. However, this strength does not transfer to DEL. As outlined in Chapter 3.3.3, DEL is highly sensitive to extreme turbulence fluctuations due to the Wöhler exponent, meaning that even small mismatches in the tails of the turbulence distribution have a large impact on the final value. The remodel focuses on correcting mean behavior and typical variability rather than explicitly adjusting the extremes, so it tends to underestimate the highest-turbulence events. This explains why the remodel performs strongly in MRBE and RRMSE but is less effective for DEL, where accurate representation of extreme variability is essential.

Simple Scaling, although based only on monthly mean ratios, performs surprisingly well in MRBE. At Site 1 it often outperforms MMS, even though MMS includes wind direction and creates a 2-dimensional binning structure. Simple Scaling also performs well at Site 2 in several wind speed bins. However, its limitations become clear in RRMSE: it tends to produce the highest errors at low wind speeds, where wind variability is greatest. MMS and YMS show only small differences in performance. At Site 1, MMS performs slightly better in MRBE while YMS performs better in RRMSE, which reflects the difference between using monthly versus yearly matrices. At Site 2 the two methods behave almost identically because the narrower and more irregular wind speed distribution reduces the advantage of using monthly time resolution. In such cases, the monthly matrices of MMS can produce fluctuating or poorly supported scaling factors, while YMS benefits from a larger, more uniform sample. This effect is especially visible

when the wind-speed distribution is narrow or irregular, because the yearly aggregation in YMS avoids gaps and reduces sensitivity to monthly differences. As a result, YMS can outperform MMS even though it is less complex, simply because the yearly ratios are more robust when wind speeds vary strongly or when certain bins are sparsely populated.

Among the machine-learning-based and distribution-based methods, XGB shows a very consistent pattern that fits its theoretical structure. In MRBE and RRMSE it performs especially well at higher wind speeds, where the residual-correction behavior of boosted trees becomes effective. In DEL at Site 1, XGB frequently appears as the best method, particularly in Tiers 2 and 3, where the additional tail corrections support its strong performance. At Site 2 it is less dominant but still performs well in selected wind speed ranges. Quantile Mapping stands out the most across all metrics. Its distribution-matching nature often causes it to behave as an outlier - sometimes becoming the best method, sometimes producing unusually high errors. This behavior is especially visible in the DEL results for both sites in Tier 1 and Tier 2, where QM occasionally becomes the best method simply because its adjustments move the values closer to the measured distribution tails. In RRMSE, QM consistently appears at the upper end of the cluster, particularly at Site 1, showing that its distribution correction does not always translate into lower overall error magnitudes.

7 Conclusion

The results demonstrate that the performance of each method depends strongly on the wind speed distribution, the site characteristics, and the metric - and corresponding application - being considered. Offshore conditions, with their broader and more stable wind speed range, allow most methods to operate consistently across bins. Onshore conditions, in contrast, limit method performance in the higher wind speed bins due to data scarcity and stronger natural variability.

All methods tested in this study show distinct strengths. Remodel and XGB perform most reliably across MRBE and RRMSE, especially at the offshore site, where stable wind conditions support their underlying modelling principles. SS performs better than expected in MRBE despite its simplicity, while MMS and YMS show small but meaningful differences depending on the site and metric. QM often stands out as an outlier, occasionally reaching the best performance but often producing larger errors due to its strong sensitivity to distribution shapes. For DEL, which amplifies the effect of extremes through the Wöhler exponent, XGB and MMS perform well at a broad range of TI and wind speed condition, while Remodel - despite its strengths in MRBE and RRMSE - is less suited to capturing the highest-turbulence events. Given how strongly results differ between sites, method selection should always be selected based on the specific site conditions, rather than applying a single universal approach.

The multi-tier correction concept has a clear influence on the results of this study. Tier 2 reduces the overly clustered behavior of the baseline predictions by matching the monthly variability of the measurements, which leads to a more realistic spread, especially at lower and mid wind speeds. Tier 3 further refines the predictions by adjusting the extreme values so that the upper tail of each method more closely reflects the measured distribution. These corrections work particularly well at the offshore site, where the wind speed bins are densely populated and provide sufficient information for both the variance and extreme-value adjustments to take effect across the full wind-speed range. At the onshore site, however, the narrower and more irregular wind-speed distribution limits the impact of the corrections - especially in the higher bins where data are sparse. Overall, the tier effects follow their intended purpose: Tier 2 improves the representation of variability, and Tier 3 strengthens the alignment of extremes.

A key aspect of the evaluation is how the method performances relate to the natural interannual variabilities of the two sites. Both sites have characteristic DEL variability ranges - approximately $\pm 9.9\%$ offshore and $\pm 15.1\%$ onshore - which represent the uncertainty introduced by the current standard industry practice of choosing a single measurement year. In the DEL error diagrams, no method consistently stays within these thresholds across all wind speed bins at either site. Every method exceeds the natural variability limits at some point, regardless of the correction tier. This shows that while the methods can improve certain wind speed ranges,

none of them fully overcome the uncertainty introduced by the year-to-year variability of the sites. The continuing threshold violations can be explained with two main lines of reasoning. The first point being wind conditions varying strongly from year to year, especially onshore, while natural variability is too high to be fully captured by training on only one year of data or by applying simple bias or variance corrections. Second, the correction tiers adjust only certain statistical features, such as the mean, the variance, or the shape of the distribution. However, the errors that matter for load estimation are influenced by more complex effects, including irregular turbulence behavior, non-linear turbine responses, and atmospheric factors that are not directly measured. Because of this, even the best correction methods still leave some errors that fall outside the required thresholds.

Although no single method consistently outperforms the others across all metrics, sites, or wind speed bins, all calibration approaches still move the original long-term model dataset closer to the measured turbulence behavior. In this sense, the methods do improve the long-term prediction, but the improvements are not yet sufficient to reliably mitigate the uncertainty resulting from the natural variability of the sites across the full range of wind speed conditions.

References

- [1] G. Ren, J. Liu, J. Wan, F. Li, Y. Guo und D. Yu, „The analysis of turbulence intensity based on wind speed data in onshore wind farms,“ *Renewable Energy*, pp. 756-766, 3 January 2018.
- [2] M. Konagaya, T. Ohsawa, T. Inoue, T. Mito, H. Kato und K. Kawamoto, „Land–Sea Contrast of Nearshore Wind Conditions: Case Study in Mutsu-Ogawara,“ *SOLA (Scientific Online Letters on the Atmosphere)*, Bd. 17, p. 234–238, 2021.
- [3] S. Emeis, *Wind Energy Meteorology*, Garmisch-Partenkirchen: Springer, 2013.
- [4] L. Gudmundsson, J. B. Bremnes, J. E. Haugen und T. Engen-Skaugen, „Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods,“ *Hydrology and Earth System Sciences*, pp. 3383-3390, 21 September 2012.
- [5] „Google Developers,“ 18 February 2026. [Online]. Available: <https://developers.google.com/machine-learning/decision-forests/decision-trees?hl=de>. [Zugriff am 16 March 2026].
- [6] „Google Developers,“ 18 February 2026. [Online]. Available: <https://developers.google.com/machine-learning/decision-forests/intro-to-gbdt?hl=de>. [Zugriff am 16 March 2026].
- [7] „Google Developers,“ 27 July 2025. [Online]. Available: <https://developers.google.com/machine-learning/crash-course/neural-networks?hl=de>. [Zugriff am 16 March 2026].
- [8] B. Jamil und N. Akhtar, „Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: A case study,“ *Renewable and Sustainable Energy Reviews*, pp. 329-355, 4 May 2017.
- [9] A. Natarajan, „Damage equivalent load synthesis and stochastic extrapolation for fatigue life validation,“ *Wind Energy Science*, pp. 1171-1181, 2 June 2022.
- [10] Z. Wengang, L. Hanlong, W. Lin, Z. Xing und Z. Yanmei, *Application of Machine Learning in Slope Stability Assessment*, China: Science Press Beijing, Springer, 2023.

[11] Vortex, „Times Remodeling - Technical Details and Validation,“ Vortex S.L., Barcelona, 2024.