Hilal H. Nuha

Associate-Professor
School of Computing
Telkom University
Indonesia

Adil Balghonaim

Assistant-Professor
King Fahd University of Petroleum &
Minerals, Dhahran
Saudi Arabia

Rizka Reza Pahlevi

Assistant-Professor
School of Computing
Telkom University
Indonesia

S. Rehman

Research Engineer
King Fahd University of Petroleum &
Minerals, Dhahran
Saudi Arabia

M. Mohandes

Professor
King Fahd University of Petroleum &
Minerals, Dhahran
Saudi Arabia

Vertical Wind Speed Extrapolation Using Statistical Approaches

The wind power industry has experienced a significant increase and popularity in recent times, and the latest statistics indicate that this sector is still thriving. However, one of the essential steps in developing wind energy projects is finding suitable sites for wind farms, which involves understanding the nature of wind speed, wind direction, terrain, and environmental impacts. To predict the wind energy production over the expected lifespan of a wind farm, vertical wind speed extrapolation to the hub height of the wind turbine is necessary. Therefore, this study presents a comprehensive evaluation of seven statistical approaches for vertical wind speed extrapolation, including Generalized Linear Models (GLM), Linear Regression (LR), Support Vector Machines (SVM), Generalized Additive Models (GAM), Gaussian Process Regression (GPR), Regression Tree (RT), and Ensemble Regression (ER). The accuracy of these methods is assessed using performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Normalized RMSE (NRMSE), Normalized MSE (NMSE), Mean Bias Error (MBE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and R-squared (R2). The study concludes that, on average, GLM performs the best out of all seven statistical methods.

Keywords: Vertical Wind Speed Extrapolation, Statistical Approaches, Regression, GLM, SVM

1. INTRODUCTION

Global wind power deployment has experienced an enormous surge due to its technological maturity, ease of maintenance, minimal human interaction, and commercial acceptance. In 2022, the global cumulative wind power installed capacity reached a record high of 898.24 GW [1]. This significant growth was largely driven by policy deadlines in China and the United States, both pushing for cleaner and sustainable energy source utilization. With more and more countries committing to reducing their carbon footprints, wind power will likely play an incre–asingly important role in meeting global energy demands in the years to come.

Wind speed (WS) and direction are the most fluc–tuating meteorological parameters. These parameters change with time, location, and height of measure–ment. Therefore, thorough knowledge of wind speed variability is important in -identifying suitable site/s for wind farm development. Usually, the wind speed mea–surements are carried out at heights below the wind turbine hub height. So, vertical wind speed extra–polation to the wind turbine hub height is useful for estimating the wind energy production over the lifetime of a wind farm. By analyzing wind speed data over a long period of time, wind energy companies can better understand the wind resources at a given site and optimize the placement of turbines to maximize energy production. Figures 1 and 2 indicate

that WS increases with height.

Numerous methods have been used for the extra–polation of WS to hub heights using measured values taken at more than one height. Newman and Klien [2] discovered that the power law does not perform effectively in unstable atmospheric conditions. To determine the local wind shear exponent (LWSE), WS measurements are needed at multiple heights. The LWSE is utilized to extrapolate the WS at higher heights. Although the LWSE proves to be the reliable option for WS extrapolation, it is a highly site-dependent parameter. So, one requires LWSE for almost every site under consideration which is a cost and time-intensive process. Ayodele et al. [3] utilized WS data from 20 and 60 meters to calculate the LWSE, resulting in a more precise wind power assessment. Tizpar et al. [4] employed WS observations at 10, 30, and 40 meters to predict the LWSE, ensuring a dependable evaluation of wind power potential at the hub height. In a study conducted in Northern Cyprus, Solyali et al. [5] determined the LWSE by using WS at 50, 80, and 90 meters to accurately estimate WS at the required hub height for precise wind power potential assessment.

Boro et al. [6] investigated the characteristics of the vertical wind profile at Burkina Faso using wind data at 10 m above ground level (AGL) and satellite data at 50 m height in the atmosphere boundary. The authors used the standard power and the logarithmic laws to estimate WS data from 20 m to 50 m. Barantiev and Batchvarova [7] analyzed wind speed profile statistics from acoustic soundings at the black sea coastal site. The authors used seven years of remote sensing data covering from 30 m to 600 m heights with vertical resolution of 10 m.
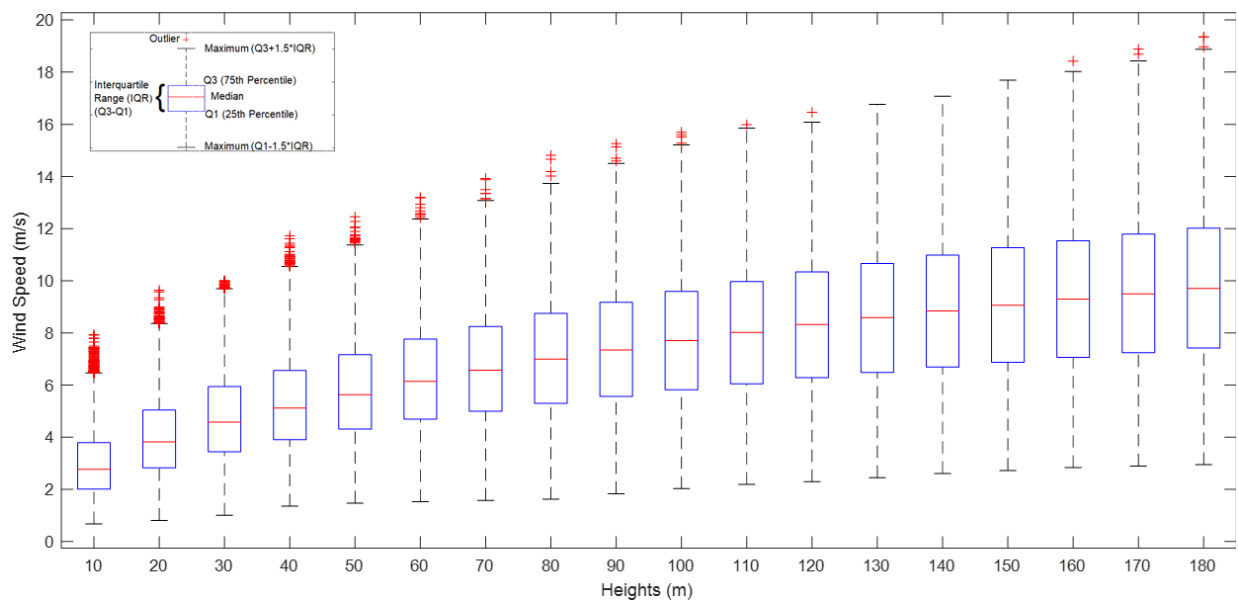
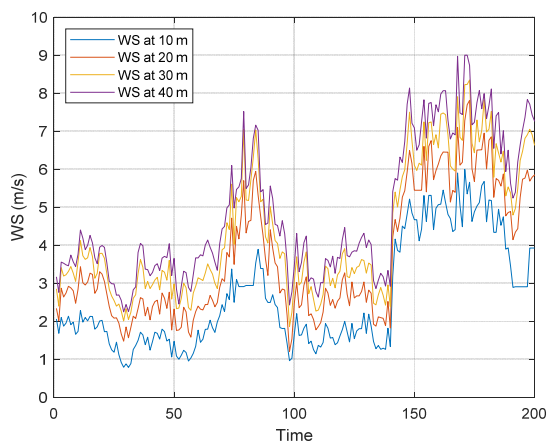**Figure 1. Windspeed statistics using boxplot**

Figure 2. Windspeed measured at 10, 20, 30, and 40 m using LiDAR

Steinheuer and Friederichs [8] used multivariate extreme value theory to derive a conditional distribution for hourly peak WS as a function of height. For training the system, the authors used peak WS observations at 5 vertical levels between 10 m and 250 m from the Hamburg weather mast. Sucevic and Djurisic [9] analyzed the atmospheric stability influence on wind speed profile and showed that the proposed method outperformed the classic logarithmic law. Studies on WS extrapolation with other methods can be found in [10-15].

Existing studies on WS extrapolation have shown a significant reliance on extensive site-dependent parameters, highlighting a critical gap in the field. Traditional models, especially in unstable atmospheric conditions, lack the versatility needed for diverse environmental scenarios. The reliance on the LWSE, while providing reliable results, necessitates detailed, site-specific measurements, leading to processes that are both time-consuming and costly. This issue is further compounded by the limitations in the range of heights at which WS data is collected in many studies, restricting the ability to efficiently interpolate and extrapolate WS across broader vertical spans. Additionally, the focus of

many studies on specific methodological approaches, such as multivariate extreme value theory or atmospheric stability analysis, underscores the need for a more comprehensive model capable of integrating a wider array of factors affecting WS. This gap underscores the necessity for a more adaptable and less site-dependent approach in WS extrapolation research. This gap can be solved utilizing statistical methods that exploit the information provided by the available WS data measured at lower heights. These methods can model the statistical properties of the data that can be used to extrapolate WS at higher heights.

This paper investigates the performance of seven different statistical approaches for vertical WS extra–polation. The statistical approaches used include Gene–ralized Linear Models (GLM), Linear Regression (LR), Support Vector Machines (SVM), Generalized Additive Models (GAM), Gaussian Process Regression (GPR), Regression Tree (RT), and Ensemble Regression (ER). To evaluate the accuracy of these methods, several per–formance metrics are used, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Nor–malized RMSE (NRMSE), Normalized MSE (NMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Mean Absolute Percentage Error (MAPE), Mean Per–centage Error (MPE), Symmetric Mean Absolute Per–centage Error (SMAPE), and coefficient of deter–mination (R^2). By examining the performance of these seven methods, it can be determined which approach provides the most accurate estimation of vertical WS extrapolation.

The main scientific merit and contribution to the practice of engineering of this paper is as follows. This study presents the use of diverse statistical models like GLM, SVM, and GPR indicating a robust and inno–vative approach to data analysis. This contributes to the scientific field by exploring and validating various methodologies for predictive modeling. In addition to that, the detailed assessment of performance metrics like MSE, RMSE, and MAE demonstrates a comprehensive approach to evaluating the accuracy and reliability of the models. This contributes to scientific rigor and helps

in establishing benchmarks for future studies. The application of these statistical methods in engineering can significantly improve predictive modeling, essential for planning, design, and operational decisions. For ins–tance, in energy engineering, accurate WS extrapolation is vital for wind farm installation planning.

The remainder of this paper is organized as follows. The Methodology section presents the approaches and the mathematical equations used in this paper. The experimental data collection section provides a brief description of the vertical wind speed profiler. The results section presents a comprehensive analysis of the performance of each algorithm, including a comparison of their accuracy metrics. Finally, the conclusion section summarizes the findings and highlights the significance of this study for wind energy applications.

2. METHODOLOGY

This section presents all statistical approaches used to perform vertical WS extrapolation and metrics to mea–sure accuracy.

2.1 Generalized Linear Model (GLM)

The GLMs [16] are a broad class of statistical models that can handle a wide range of response variables, including binary, count, and continuous data. This paper uses a GLM with a quadratic model. It means that the GLM is fitted using a quadratic function of the predictor variables. Mathematically, a quadratic function has the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 \qquad (1)$$

where y denotes the response variable, β_0 is the bias, β_1 and β_2 are the coefficients of the predictor variables x_1 and x_2^2, respectively. The predictor variable x_2 is squared to introduce a nonlinear relationship between the response and the predictor variables.

In this model, x_1 is defined as a vector representing WS measurements at p heights ($h_1 \ldots h_p$). Specifically, $x_1 = \{x_{h1}, x_{h2}, \ldots, x_{hp}\}$, where each element x_{hi} corresponds to the wind speed measured at a different height. The term x_2^2 in the model is not a separate set of measu–rements but rather the square of the vector x_1. In the model, β_1 and β_2 are vectors of coefficients that corres–pond to the predictor variables x_1 and x_2, respectively.

This quadratic model is useful when the relationship between the response and predictor variables is nonlinear and can be approximated by a quadratic function. By fitting the quadratic model, the GLM can capture the curvature of the relationship between the response and predictor variables, which cannot be captured by a simple linear model. The coefficients, β_0, β_1 and β_2 can be calculated as follows:

$$(\beta_0, \beta_1, \beta_2) = (X'X)^{-1} X'y \qquad (2)$$

where X is the data input matrix, which is constructed by stacking the predictor variable x and its squared value x^2, and y is the response variable. In this model, the matrix X consists of input vectors from N samples:

$$X = \begin{pmatrix} x_1(1) & x_2^2(1) \\ \vdots & \vdots \\ \vdots & \vdots \\ x_1(N) & x_2^2(N) \end{pmatrix} = \begin{pmatrix} x_{1,h1}(1)\ldots x_{1,hp}(1) & x_{2,h1}^2(1)\ldots x_{2,hp}^2(1) \\ \vdots & \vdots \\ \vdots & \vdots \\ x_{1,h1}(N)\ldots x_{1,hp}(N) & x_{2,h1}^2(N)\ldots x_{2,hp}^2(N) \end{pmatrix} \qquad (3)$$

where $x_j(n)$ denotes the input vector j from sample n.

2.2 Linear Regression (LR)

The mathematical representation of a linear regression [17] model is given by the following equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \qquad (4)$$

where, y is the response variable and $x_1, x_2, \ldots, x_p$ are the predictor variables. $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are the coefficients or parameters to be estimated. The LR estimates the values of the coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ that best fit the data by minimizing the sum of squared residuals. The variables x_1 through x_p represent wind speed measurements at sequential heights, starting from the first measurement height and going up to the p-th height. The value of p, which denotes the number of measurement heights included in the model, varies depending on the specific height at which the wind speed is being extrapolated. For example, consider a scenario where the goal is to extrapolate wind speed at a height of 50 meters. In this case, the model would use wind speed measurements from lower heights as inputs to predict the wind speed at the 50-meter level. If the available measurements are at heights of 10, 20, 30, and 40 meters, then these measurements correspond to x_1, x_2, x_3 and x_4 respectively, and therefore, $p = 4$. Each x_i in this sequence represents the wind speed recorded at the i-th height.

This process entails employing the ordinary least squares (OLS) technique, which requires the coeffi–cients' values that minimize the total of the squared variances between the observed response variable and the model-predicted values. Similar to the GLM, standard LR coefficients can be calculated using equation 2 where the squared values are excluded from the data input matrix.

2.3 Generalized additive model (GAM)

The next statistical method is the generalized additive model (GAM) [18] which is specified as given below:

$$Y = g^{-1}\left(\alpha + f_1(X_1) + f_2(X_2) + \ldots + f_p(X) + \varepsilon\right) \qquad (5)$$

where Y is the response variable, $X_1, X_2, \ldots, X_p$ are the predictor variables, α is the bias, g is the link function that maps the linear predictor to the response space, and $f_1, f_2, \ldots, f_p$ are the smooth functions that capture the nonlinear relationship between the predictors and the response. The smooth function used in this paper is the logistic function as follows:

$$f(x) = \frac{1}{1 + e^{-x}} \qquad (6)$$

2.4 Support vector machine (SVM)

A support vector machine (SVM) [19] regression model predicts the response variable based on the values of

one or more predictor variables. The SVM algorithm finds a hyperplane that best separates the data into two classes. In regression, the SVM algorithm tries to find a hyperplane that best fits the data by minimizing the mean-squared error between the predicted and actual responses. Mathematically, the SVM regression model can be represented as:

$$y = w'x + b + e \qquad (7)$$

where *y*, *x*, *w*, *b* and *e* denote the predicted response, the vector of predictor variables, the weight vector, the bias term, and the error term; respectively. Basically, SVM finds the best linear fit that separates data points of different predicted values, with a focus on maximizing the margin between data points and the decision boundary. Therefore, SVM is less sensitive to outliers compared to LR. However, SVMs can be computa–tionally more intensive than LR. This can be a drawback for very large datasets or real-time processing.

On the other hand, LR seeks to find a linear rela–tionship between predictor variables and a conti–nuous response variable. It does this by minimizing the sum of the squares of the differences between observed and predicted values. LR is straightforward to imp–lement and interpret. It's well-suited for situations where the relationship between variables is expected to be linear. LR models are computationally efficient, making them suitable for large datasets or real-time analysis. However, LR can be significantly influenced by outliers in the data.

2.5 Gaussian process regression (GPR)

A Gaussian process regression (GPR) [20] model uses a Gaussian kernel function with an isotropic (spherical) covariance function. Mathematically, the GPR model can be described as a set of input data $X = [x_1, x_2, \ldots, x_N]$ and corresponding output data $Y = [y_1, y_2, \ldots, y_N]$, where each x_n is a *d*-dimensional vector, and each y_i is a scalar. The GPR model aims to find a function f(x) that maps the inputs to the outputs with uncertainty quantified by a Gaussian distribution. Specifically, for a test input vec–tor x_q, the GPR model estimates the output as follows:

$$f\left(x_q\right) = \mu\left(x_q\right) + \varepsilon\left(x_q\right) \qquad (8)$$

where $\mu(x_q)$ is the mean function and $\varepsilon(x_q)$ is a random variable representing the residual error at x_q. The mean function is defined as:

$$\mu\left(x_q\right) = K\left(x_q, X\right)' \left(K\left(X, X\right) + \sigma_0^2 I\right)^{-1} Y \qquad (9)$$

where $K(x_q, X)$ represents the covariance matrix where each component of the matrix is the value of kernel function $k(x_1, x_2)$ that quantifies the similarity between x_q and each input vector in X, σ_n is the noise standard deviation, I is the identity matrix, and Y is the vector of output data. If there is no strong prior knowledge about the data, $\mu(x_q)$ can use a zero-mean function. The kernel function is defined as:

$$k\left(x_1, x_2\right) = \sigma_f^2 \exp\left(-\frac{\|x_1 - x_2\|^2}{2l^2}\right) \qquad (10)$$

where σ_f^2 is the signal variance, l is the length scale parameter, and $\|x_1 - x_2\|$ is the Euclidean distance between the two input vectors.

2.6 Regression Tree (RT)

The regression tree (RT) [21] model is used for pre–dicting a response variable based on aet of predictor variables. The RT for building the decision tree is based on the recursive partitioning of the predictor variables into smaller subspaces that are as homogeneous as pos–sible with respect to the response variable. The RT algorithm works as explained below:

1. The algorithm starts with the entire dataset and selects a predictor variable and a split point that opti–mizes the splitting criterion. The splitting criterion is typically the reduction in the mean squared error (MSE) between the response variable and the predicted values for each subspace.

2. The algorithm then divides the dataset into two smaller subsets based on the selected split point: one subset where the predictor variable is less than or equal to the split point and another subset where the predictor variable is greater than the split point.

3. For each subset, the algorithm recursively repeats steps 1 and 2 until a stopping criterion is met. The stop–ping criterion can be based on the maximum depth of the tree, the minimum number of observations in a leaf node, or the minimum reduction in MSE achieved by the split.

4. At each leaf node of the tree, the response variable is predicted as the mean value of the response variable for all observations in the leaf node.

Once the decision tree model has been created using the training data, it can be used to predict the response variable for new observations by traversing the tree from the root node to the appropriate leaf node based on the values of the predictor variables. The predicted response variable for a new observation is then the mean value of the response variable for all observations in the leaf node.

2.7 Ensemble Regression (ER)

The ensemble regression (ER) [22] is a training method for ensembling the RT. The method combines the predictions of multiple regression trees to produce a more accurate and robust model. Given a training set of input-output pairs $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N),\}$, where x_n is a vector of input features and y_n is the corres–ponding output value, the ER algorithm proceeds as follows:

a. Draw a bootstrap sample of the training data, i.e., randomly select n samples from the training set with replacement.

b. Grow a regression tree on the bootstrap sample. At each node of the tree, select the feature and split point that minimize the mean squared error of the predictions.

The ensemble's prediction for a new input vector *x* is the average of the predictions of each individual tree. In numerical experimentation, the number of trees in the

ensemble is set to 100. The tree depth is unlimited, which means that each tree is grown until all the leaves are pure or have fewer than 5 observations with mean squared error as the splitting criterion. The ER algorithm uses bagging and random feature subsampling to improve the generalization performance of the model and prevent overfitting.

2.8 Models Summary

All models used in this paper are summarized in Table 1.

Table 1. Summary of ModelsMain Feature

No	Model	Main Feature
1	GLM	Handles a wide range of response variables including binary, count, and continuous data; can incorporate a quadratic function for non-linear relationships.
2	LR	Models the relationship between the response and predictor variables linearly; employs Ordinary Least Squares for parameter estimation.
3	GAM	Captures non-linear relationships between predictors and response through smooth functions; highly flexible.
4	SVM	Finds a hyperplane in a high-dimensional space that best fits the data; minimizes mean-squared error between predicted and actual responses.
5	GPR	Utilizes Gaussian kernels with isotropic covariance; maps inputs to outputs with quantified uncertainty.
6	RT	Builds decision trees based on recursive partitioning; nodes split based on mean squared error reduction criterion.
7	ER	Combines multiple regression trees to create a more accurate and robust model; uses techniques like bagging to improve performance.

2.9 Error Metrics

This paper uses several metrics that are commonly used to evaluate the performance of each method.

1. Mean Squared Error (MSE) [23] is a measure of the average squared difference between the predicted and actual values. It is obtained by calculating the mean of the squared differences between the predicted and actual values:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2 \qquad (11)$$

where N, y_n, and $\hat{y}_n$ denote the number of observations, the actual value of the n-th observation, and the predicted value of the n-th observation, respectively.

2. Root Mean Squared Error (RMSE) [23] is the square root of the MSE and is also a measure of the average difference between the predicted and actual values:

$$RMSE = \sqrt{MSE} \qquad (12)$$

Normalized RMSE (NRMSE) [23] is the RMSE divided by the range of the actual values. It is a measure of the relative error of the predictions:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \qquad (13)$$

where y_{max} and y_{min} are the maximum and minimum values of the actual values, respectively.

4. Normalized MSE (NMSE) [25] is the MSE divided by the energy of the actual values. It is also a measure of the relative error of the predictions:

$$NSME = \frac{MSE}{\sum_{n=1}^{N} y_n^2} \qquad (14)$$

5. Mean Bias Error (MBE) [26] is the average difference between the predicted and actual values. It is a measure of the bias of the predictions:

$$MBE = \frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n) \qquad (15)$$

6. Mean Absolute Error (MAE) [23] is the average absolute difference between the actual and predicted values. It is a measure of the magnitude of the errors:

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|y_n - \hat{y}_n| \qquad (16)$$

7. Mean Percentage Error (MPE) [27] is the average percentage difference between the predicted and actual values. It provides the magnitude and direction of the errors:

$$MPE = \frac{1}{N}\sum_{n=1}^{N}\frac{y_n - \hat{y}_n}{y_n}\times100\% \qquad (17)$$

8. Mean Absolute Percentage Error (MAPE) [23] is the average absolute percentage difference between the actual and predicted values.

$$MAPE = \frac{1}{N}\sum_{n=1}^{N}\frac{|y_n - \hat{y}_n|}{y_n}\times100\% \qquad (18)$$

9. Symmetric Mean Absolute Percentage Error (SMAPE) [23] is a symmetric version of the MAPE, which measures the average absolute percentage difference between the predicted and actual values but uses the average of the predicted and actual values in the denominator. This makes it less sensitive to outliers:

$$SMAPE = \frac{1}{N}\sum_{n=1}^{N}\frac{|y_n - \hat{y}_n|}{\frac{1}{2}(y_n + \hat{y}_n)}\times100\% \qquad (19)$$

10. Coefficient of determination (R^2) [23] is a measure of the goodness of fit of the regression model. It represents the proportion of the variance in the dependent variable that is explained by the independent variables. It is calculated as:

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \bar{y}_n)^2} \qquad (20)$$

where   is the average of the measured values. R^2 ranges between 0 and 1, with higher values indicating a better fit.

3. EXPERIMENTAL MEASUREMENT RESULTS

WS measurements are often made up to 40 meters due to resource limitations and accordingly, estimation techniques are needed to obtain wind speed data at higher heights. Figures 1 and 2 show that WS increases with height.

The dataset used in this research on WS estimation originates from Dhahran, Saudi Arabia, and consists of 34,660 samples collected from June 20, 2015, to Feb–ruary 29, 2016, with measurements taken every 10 mi–nutes. The data includes WS measurements from heig–hts ranging from 10 to 180 meters. It is divided into three subsets for the purpose of model training and evaluation: 70% (about 24,262 samples) for training, 10% (around 3,466 samples) for validation, and 20% (approximately 6,932 samples) for testing. The research aims to estimate wind speeds at higher altitudes using measurements from lower heights, using an iterative estimation process where LiDAR-measured wind spe–eds at lower heights (10-40 meters) are used as refe–rence for training models, and the output is the esti–mated wind speeds at higher altitudes, up to 180 meters. Although the dataset is specific to Dhahran and reflects its local atmospheric conditions, the methodologies and models developed could potentially be adapted for or compared with other locations, subject to similar data collection criteria and environmental conditions.

The estimation process involves training the models using measured WSs at lower heights as input and at higher heights as desired output. Subsequently, the proposed statistical estimation methods are trained using both actual and estimated wind speed values at lower heights to estimate wind speeds at the next level. This process is iterated until wind speeds at 180 meters, utilizing actual data from 10-40 meters and estimated data from 50-170 meters, are obtained.

As shown in Table 2, the GLM model performs well, with low values for the MSE, RMSE, NRMSE, and NMSE. However, as the height increases, the errors tend to increase as well, as indicated by the increasing values for these metrics. All of the above error values keep on increasing, slightly, as the height increases. The MBE values show slight under-prediction of WSs at most of the heights while positive bias is observed at 160 to 180 meters levels. The MAE, MPE, MAPE, and SMAPE all increase as the height increases, indicating larger errors for higher heights due to being away from the measured values of WSs at 10 to 40-meter heights.

Table 2 GLM Performance

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
MSE	0.041	0.149	0.251	0.408	0.585	0.81	1.01	1.25	1.46	1.71	1.93	2.19	2.36	2.58
RMSE	0.202	0.386	0.501	0.639	0.765	0.90	1.00	1.12	1.21	1.31	1.38	1.48	1.53	1.60
NRMSE	0.023	0.042	0.052	0.065	0.070	0.07	0.08	0.08	0.09	0.10	0.10	0.11	0.11	0.12
NMSE	0.001	0.003	0.005	0.007	0.009	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02
MBE	0.003	-0.05	-0.05	-0.08	-0.08	-0.11	-0.09	-0.10	-0.07	-0.05	-0.02	0.00	0.06	0.10
MAE	0.154	0.301	0.393	0.504	0.603	0.71	0.79	0.88	0.94	1.02	1.09	1.17	1.22	1.28
MPE	0.120	0.850	0.748	0.977	0.920	1.04	0.75	0.59	0.15	-0.19	-0.72	-1.18	-1.80	-2.35
MAPE	2.530	4.602	5.667	6.869	7.781	8.73	9.32	9.97	10.41	10.92	11.34	11.81	12.14	12.54
SMAPE	2.530	4.631	5.694	6.917	7.831	8.80	9.37	10.00	10.41	10.88	11.24	11.67	11.92	12.25

Table 3. LR Performance

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
MSE	0.049	0.182	0.312	0.500	0.715	0.994	1.232	1.521	1.77	2.07	2.31	2.61	2.79	3.02
RMSE	0.222	0.426	0.559	0.707	0.846	0.997	1.110	1.233	1.33	1.43	1.52	1.61	1.67	1.74
NRMSE	0.025	0.046	0.058	0.071	0.078	0.083	0.090	0.098	0.10	0.11	0.11	0.12	0.12	0.13
NMSE	0.001	0.004	0.006	0.008	0.011	0.014	0.016	0.018	0.02	0.02	0.02	0.02	0.02	0.02
MBE	0.002	-0.05	-0.05	-0.10	-0.13	-0.18	-0.20	-0.240	-0.25	-0.27	-0.28	-0.29	-0.27	-0.26
MAE	0.165	0.325	0.430	0.550	0.659	0.779	0.866	0.962	1.03	1.11	1.18	1.25	1.30	1.36
MPE	0.155	0.924	0.938	1.370	1.586	2.033	2.094	2.306	2.24	2.26	2.09	1.97	1.67	1.39
MAPE	2.676	4.899	6.108	7.382	8.383	9.413	10.04	10.727	11.16	11.65	11.99	12.38	12.60	12.86
SMAPE	2.676	4.927	6.140	7.450	8.478	9.564	10.21	10.933	11.37	11.88	12.20	12.59	12.77	13.01

Table 4. GAM Performance

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
MSE	0.060	0.167	0.301	0.519	0.751	1.061	1.349	1.671	1.935	2.260	2.507	2.848	3.080	3.342
RMSE	0.244	0.409	0.548	0.721	0.866	1.030	1.161	1.293	1.391	1.503	1.583	1.688	1.755	1.828
NRMSE	0.028	0.044	0.057	0.073	0.079	0.086	0.094	0.103	0.110	0.117	0.122	0.128	0.133	0.137
NMSE	0.001	0.004	0.006	0.009	0.012	0.015	0.017	0.020	0.022	0.024	0.025	0.027	0.028	0.029
MBE	0.056	-0.01	-0.02	-0.04	-0.08	-0.13	-0.14	-0.193	-0.20	-0.23	-0.24	-0.26	-0.26	-0.26
MAE	0.183	0.316	0.425	0.562	0.676	0.804	0.907	1.015	1.090	1.176	1.239	1.320	1.377	1.441
MPE	-0.54	0.607	0.789	1.083	1.477	1.893	1.924	2.235	2.179	2.289	2.210	2.215	2.091	2.030
MAPE	2.968	4.842	6.102	7.596	8.657	9.787	10.59	11.406	11.84	12.36	12.67	13.12	13.40	13.74
SMAPE	2.953	4.868	6.146	7.673	8.773	9.961	10.77	11.642	12.08	12.64	12.95	13.41	13.68	14.03

Table 5. SVM Performance

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
MSE	0.033	0.130	0.222	0.369	0.541	0.776	1.005	1.304	1.635	2.059	2.548	3.156	3.80	4.59
RMSE	0.183	0.360	0.471	0.608	0.735	0.881	1.002	1.142	1.279	1.435	1.596	1.777	1.951	2.144
NRMSE	0.021	0.039	0.049	0.061	0.067	0.073	0.082	0.091	0.102	0.112	0.123	0.135	0.147	0.161
NMSE	0.001	0.003	0.004	0.006	0.008	0.011	0.013	0.016	0.018	0.022	0.026	0.030	0.035	0.040
MBE	0.014	-.018	0.020	0.044	0.104	0.160	0.268	0.375	0.524	0.672	0.851	1.031	1.245	1.460
MAE	0.141	0.283	0.373	0.485	0.588	0.703	0.802	0.917	1.033	1.169	1.318	1.484	1.641	1.814
MPE	-0.25	0.024	-.778	-1.41	-2.44	-3.39	-4.84	-6.222	-7.96	-9.62	-11.5	-13.4	-15.5	-17.5
MAPE	2.364	4.386	5.554	6.911	8.037	9.228	10.26	11.432	12.59	13.91	15.38	16.95	18.48	20.11
SMAPE	2.355	4.377	5.491	6.783	7.803	8.876	9.738	10.709	11.63	12.70	13.86	15.10	16.25	17.47

The results in Table 3 suggest that the accuracy of the LR method of wind speed estimation decreases as the height increases more rapidly than in the case of the GLM method. The lowest RMSE and MAE values are observed at 50 m, with the error increasing as the height increases. The MBE values are close to zero at the lower heights but become increasingly negatively biased higher heights, indicating a systematic underestimation of wind speed. The MPE and MAPE values show that the percentage errors increase with height, with MAPE values ranging from 2.7% at 50 m to 12.9% at 180 m. The SMAPE values also show a similar trend as MAPE, indi–cating that the errors are symmetric for the different heights.

From Table 4, it can be noticed that the GAM accuracy of the model decreases as the height increases, with the highest errors being seen at 180m. The model tends to underestimate the values, as indicated by the negative MBE values. The MAPE and SMAPE values at 180 m are 13.74% and 14.03 %, respectively.

The performance of the SVM technique is sum–marized in Table 5. In this case, the MBE values show that the model tends to slightly overestimate the wind speed at lower heights (50m, 70m) and underestimate it at higher heights (110m and above). The MAE also inc–reases with height, indicating that the model is less accurate at higher heights. The MPE shows that the model tends to underestimate the wind speed at higher heights, with a maximum underestimation of around 17.5% at 180m. The MAPE and SMAPE also show similar trends, with larger errors at higher heights. Overall, this analysis shows that the SVM model is not very accurate at predicting WSs at higher heights, and its performance decreases as the height increases. Therefore, it may not be suitable for extrapolating wind speeds to heights beyond 100-110m.

In Table 6, the GPR performance for vertical wind speed extrapolation appears to be reasonable. The RMSE values range from 0.344 m/s to 1.753 m/s, which indi–cates that the model's predictions are generally within 1-2 m/s of the actual wind speed values. The NRMSE values

are also relatively low, ranging from 0.039 to 0.132, which suggests that the model's predictions are relatively accurate and reliable across different heights. However, it's worth noting that the MPE and MAPE values are relatively high, ranging from 1.045% to 13.922%. This suggests that the model's predictions may be biased to–wards over- or under-estimating the wind speed at certain heights. Additionally, the MBE values are negative for most of the heights, which indicates that the model tends to underestimate the wind speed. In summary, while the GPR model appears to perform reasonably well, further investigation is needed to understand why the model may be biased in certain ways and whether there are ways to improve its accuracy and reliability.

The values of the performance measures in Table 7 provide insights into the accuracy of the RT approach. For instance, the MAE is less than 2 at all the heights. However, the MAPE and SMAPE are greater than 10% at most of the heights, which suggests that the percentage error can be quite high for some heights. While the regression tree appears to have some accuracy in predicting heights, the relatively high values of MAPE and SMAPE suggest that the model may not be suitable for all applications, especially those where high accuracy is critical.

The results of ER method, Table 8, show that the model's performance decreases as the height increases, with higher values observed for all error metrics at the highest height of 180 meters. However, the errors remain relatively small, with an RMSE of less than 2 meters per second at all heights. The model's MAE ranges from 0.157 m/s to 1.442 m/s, indicating that the predicted values are generally within 1-2 m/s of the actual values. The MAPE ranges from 2.606% to 13.795%, indicating that the model's predictions are within 14% of the actual values on average. The results indicate that the ensemble regression model appears to be a viable method for extrapolating vertical wind speed at various heights.

Table 6. GPR Performance

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
MSE	0.118	0.345	0.600	0.981	1.298	1.706	1.969	2.269	2.443	2.654	2.764	2.928	2.965	3.073
RMSE	0.344	0.587	0.774	0.990	1.139	1.306	1.403	1.506	1.563	1.629	1.663	1.711	1.722	1.753
NRMSE	0.039	0.063	0.081	0.100	0.104	0.109	0.114	0.120	0.124	0.127	0.128	0.130	0.130	0.132
NMSE	0.003	0.007	0.011	0.017	0.020	0.024	0.025	0.027	0.028	0.028	0.028	0.028	0.027	0.027
MBE	-.246	-0.43	-0.56	-0.70	-0.78	-0.86	-0.88	-0.882	-0.84	-0.79	-0.69	-0.59	-0.44	-0.29
MAE	0.282	0.481	0.634	0.806	0.923	1.051	1.126	1.208	1.256	1.311	1.348	1.391	1.412	1.444
MPE	3.769	6.072	7.414	8.679	9.121	9.504	9.212	8.798	7.946	6.997	5.717	4.367	2.728	1.045
MAPE	4.696	7.304	9.016	10.74	11.66	12.60	12.97	13.379	13.46	13.61	13.63	13.73	13.76	13.92
SMAPE	4.836	7.662	9.563	11.52	12.57	13.66	14.05	14.469	14.48	14.53	14.40	14.34	14.15	14.08

Table 7. RT Performance

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
MSE	0.046	0.168	0.298	0.505	0.755	1.086	1.399	1.775	2.121	2.525	2.874	3.284	3.590	3.951
RMSE	0.213	0.410	0.546	0.711	0.869	1.042	1.183	1.332	1.456	1.589	1.695	1.812	1.895	1.988
NRMSE	0.024	0.044	0.057	0.072	0.080	0.087	0.096	0.106	0.116	0.124	0.131	0.137	0.143	0.149
NMSE	0.001	0.004	0.006	0.009	0.012	0.015	0.018	0.021	0.024	0.027	0.029	0.031	0.033	0.035
MBE	0.007	-0.04	-0.04	-0.07	-0.09	-0.14	-0.16	-0.193	-0.21	-0.24	-0.25	-0.26	-0.26	-0.25
MAE	0.163	0.320	0.428	0.562	0.687	0.825	0.941	1.063	1.163	1.269	1.354	1.447	1.516	1.590
MPE	0.085	0.878	0.837	1.217	1.380	1.852	1.928	2.205	2.259	2.407	2.383	2.403	2.222	2.044
MAPE	2.708	4.921	6.183	7.640	8.842	10.10	11.05	12.007	12.70	13.41	13.92	14.47	14.84	15.24
SMAPE	2.709	4.960	6.232	7.733	8.970	10.29	11.27	12.287	13.00	13.75	14.27	14.83	15.18	15.56

Table 8. ER Performance

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
MSE	0.042	0.157	0.277	0.457	0.670	0.947	1.203	1.518	1.801	2.145	2.452	2.821	3.091	3.401
RMSE	0.204	0.396	0.527	0.676	0.819	0.973	1.097	1.232	1.342	1.465	1.566	1.680	1.758	1.844
NRMSE	0.023	0.043	0.055	0.068	0.075	0.081	0.089	0.098	0.107	0.114	0.121	0.127	0.133	0.139
NMSE	0.001	0.003	0.005	0.008	0.010	0.013	0.016	0.018	0.020	0.023	0.025	0.027	0.028	0.030
MBE	0.001	-.040	-.033	-0.06	-0.07	-0.10	-0.10	-0.127	-0.13	-0.14	-0.14	-0.16	-0.15	-0.15
MAE	0.157	0.308	0.411	0.531	0.642	0.762	0.857	0.962	1.045	1.139	1.220	1.308	1.374	1.442
MPE	0.158	0.805	0.665	0.887	0.888	1.157	1.085	1.200	1.056	1.063	0.935	0.902	0.701	0.609
MAPE	2.606	4.723	5.934	7.237	8.296	9.361	10.08	10.879	11.42	12.04	12.52	13.05	13.42	13.79
SMAPE	2.608	4.752	5.962	7.287	8.354	9.456	10.18	11.000	11.53	12.16	12.63	13.15	13.49	13.86

A satisfactory model would ideally have low errors in terms of both RMSE and MAE. While keeping RMSE or MAE below 2 m/s for WSs below 16 m/s might seem adequate in some contexts, this threshold can be considered high for power generation purposes. Relative differences in the range of 10-20% might be acceptable in some meteo–rological or navigational applications. However, for wind energy generation, where accurate power predictions are necessary, these levels of relative error could be prob–lematic. As mentioned, due to the cubic relationship bet–ween WS and power output, even a 10% error (equivalent to MAPE) in WS estimation can lead to much larger errors in power estimation. If power estimation errors are limited to 30 % or more than $a_p \geq 70\%$ of accuracy, then the WS estimation error threshold can be modeled as follows.

$$a_{WS} = \left(1 - MAPE\right) \tag{21}$$

where a_{WS} is the WS accuracy. To achieve $a_p \geq 70\%$, a_{WS} is governed by the following:

$$a_p \sim a_{WS}^3 \geq 0.7 \tag{22}$$

$$a_{WS} \geq \sqrt[3]{0.7}\left(1 - MAPE\right) \geq \sqrt[3]{0.7}\left(1 - MAPE\right) \tag{23}$$

$$MAPE < 0.1121 \tag{24}$$

Therefore, to achieve less than 30% of accuracy, the MAPE must be less than 11.21 %.

The correlation coefficient between the measured and estimated WS values is compared through scatter diagrams and the resulting R^2 values are summarized in Table 9 for all methods (GLM, LR, GAM, SVM, GPR, ER, and RT) used to extrapolate WS at different heig–hts. It can be noticed that all the models have high R^2 values up to a height of 70 meters, indicating good estimates of WS. With further height increase, the R^2 value keeps on decreasing progressively. Among the models, at lower heights (50, 60, and 70 m), SVM has the highest R^2 values followed by GLM and ER methods. However, for heights 80 to 180 m, GLM achieved the highest R^2 values. On the other hand, GPR has the lowest R^2 values for most of the heights, indica–ting that it is the least accurate model for estimating the WS. However, even GPR has an R^2 value above 50% for all the heights, which means that it is still able to capture some of the variation in the data.

Figure 3 shows the scatter plots of all methods at different heights, specifically at heights of 50m, 120m, and 180m, representing low, medium, and maximum heights, respectively. Upon examination of the scatter plots, it becomes apparent that as the height of the mea–surement increases, the accuracy of the predictions tends to deteriorate, as evidenced by the widespread between the measured and estimated values and lower R^2 values. Despite this general trend, there are still some notable differences between the different methods. At lower heights, GLM and SVM seem to perform in a similar fashion, with both models producing relatively accurate estimates. However, at higher heights, GLM outperforms SVM, as the latter tends to overestimate the WS in these scenarios. This trend is confirmed in Figure 4 where estimated WS using GLM, SVM, and GPR methods are compared with the actual WS at different heights (50 m and 180 m). It can be noticed that despite having good estimation at 50 m, SVM performs poorly at 180 m.

Table 9. All Methods R^2 Values

Heights	50	60	70	80	90	100	110	120	130	140	150	160	170	180
GLM	98.72	95.66	93.17	89.67	86.10	81.79	78.58	74.91	72.16	69.02	66.86	64.31	62.94	61.19
LR	98.46	94.81	91.86	88.02	84.06	79.47	75.97	72.08	69.04	65.68	63.31	60.61	59.07	57.20
GAM	98.60	95.83	93.09	89.19	85.23	80.49	76.78	72.80	70.09	66.84	64.77	62.00	60.47	58.71
SVM	98.87	95.84	93.26	89.51	85.66	80.98	77.45	73.41	70.44	67.09	64.76	62.05	60.53	58.64
GPR	98.06	94.77	91.14	85.93	81.26	75.72	71.59	67.20	64.22	61.05	59.10	56.92	55.82	54.47

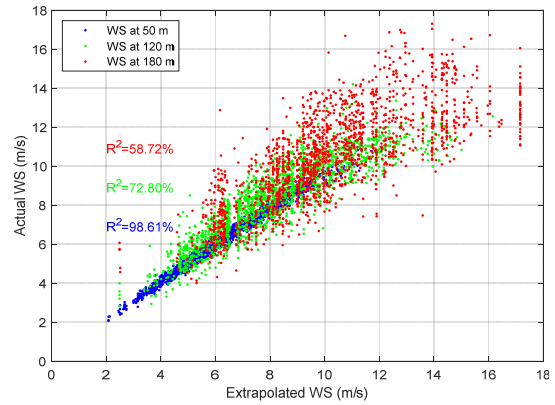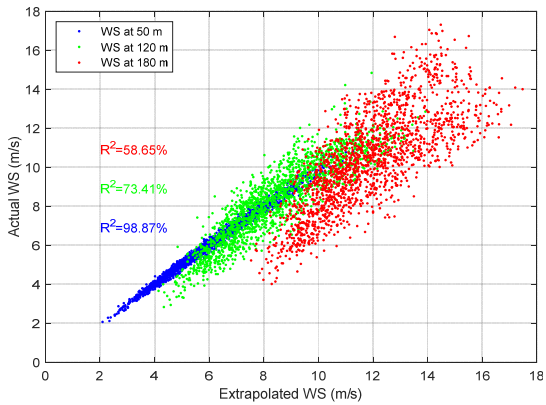| RT | 98.57 | 95.22 | 92.34 | 88.23 | 84.00 | 79.07 | 75.26 | 71.13 | 68.00 | 64.59 | 62.21 | 59.51 | 57.86 | 55.80 |
| ER | 98.67 | 95.47 | 92.70 | 88.95 | 85.01 | 80.42 | 76.76 | 72.67 | 69.50 | 65.93 | 63.30 | 60.32 | 58.38 | 56.16 |
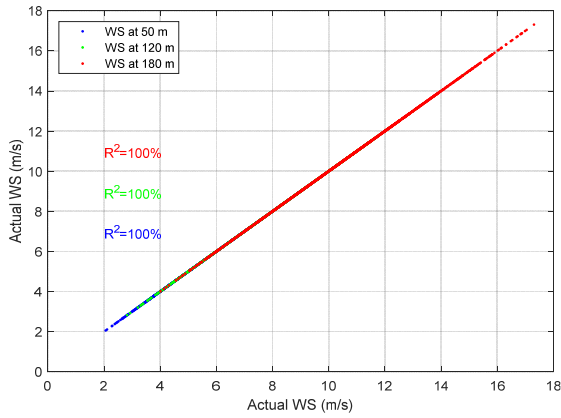
(a) Actual WS Scatter Plot
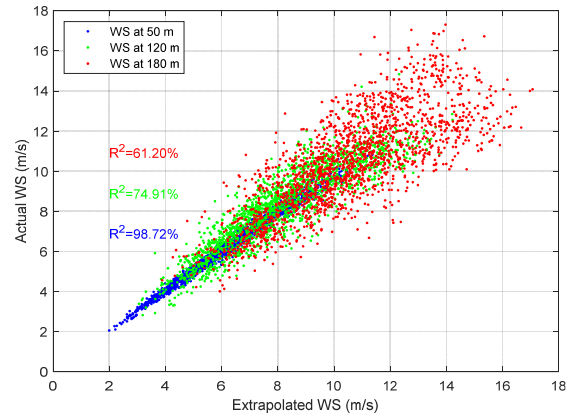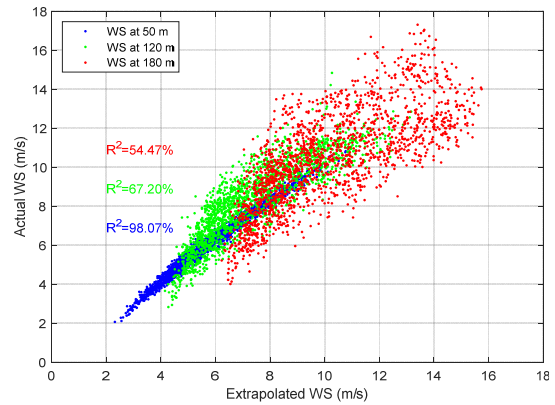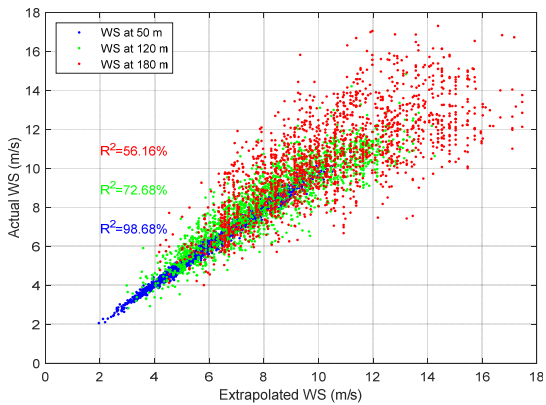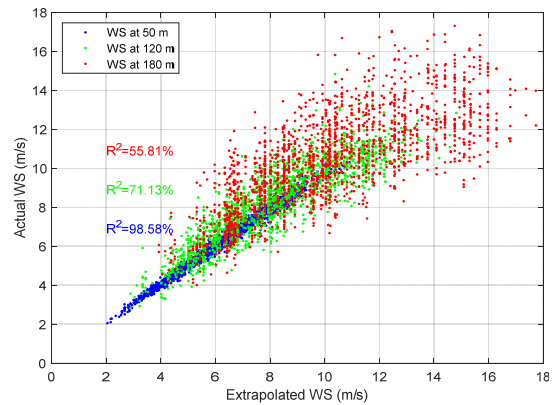
(b) GLM Extrapolated WS Scatter Plot
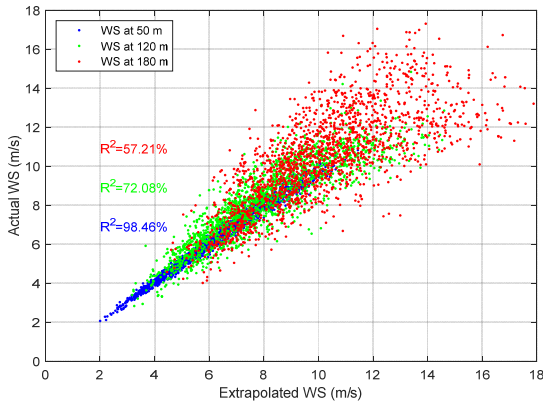
(c) LR Extrapolated WS Scatter Plot

(d) GAM Extrapolated WS Scatter Plot

(e) SVM Extrapolated WS Scatter Plot

(f) GPR Extrapolated WS Scatter Plot

(g) ER Extrapolated WS Scatter Plot

(h) RT Extrapolated WS Scatter Plot

Figure 3. WS extrapolation scatter plots at different heights (50, 120, and 180 m)
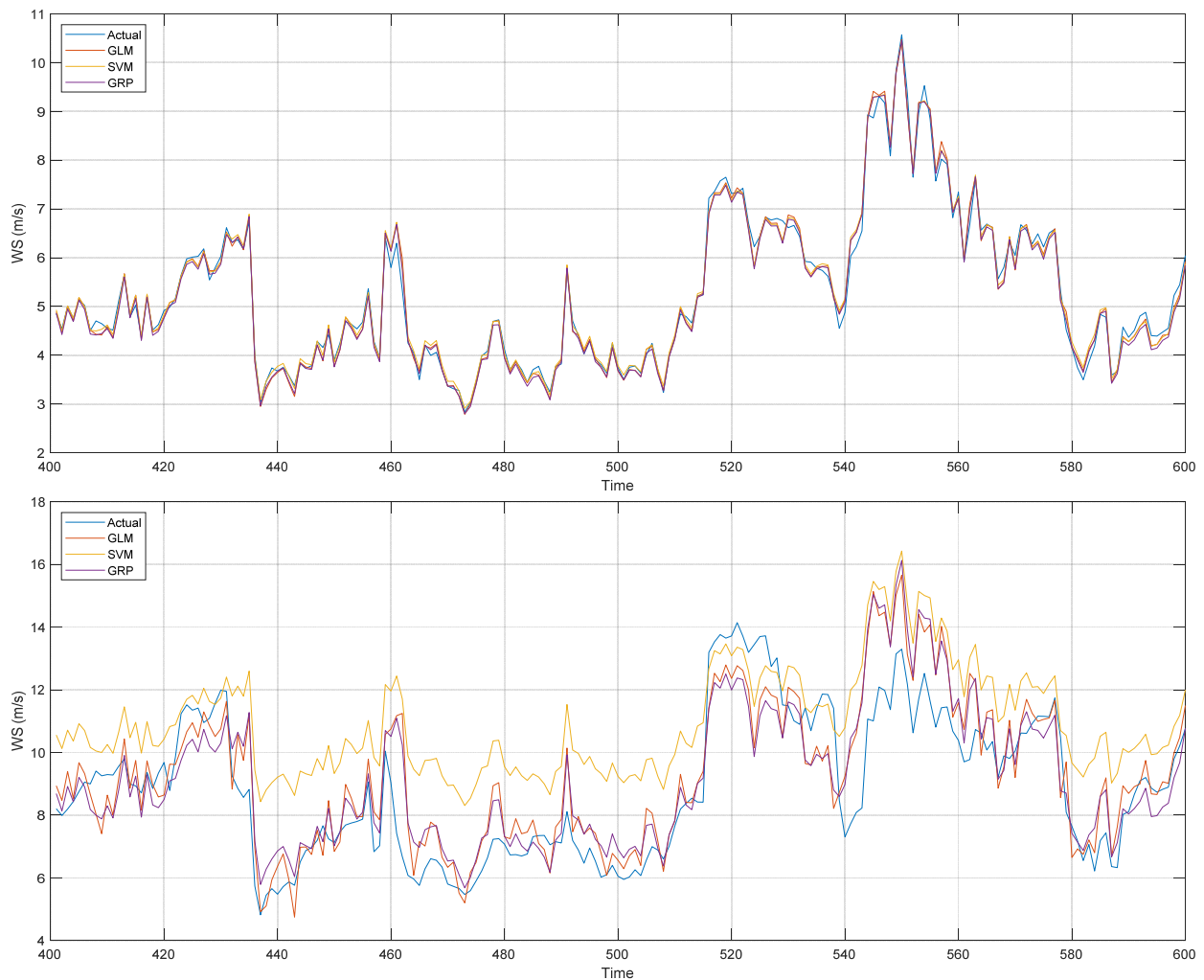
Figure 4. Extrapolated WS compared with the actual WS at (a) 50 m and (b) 180 m

4. CONCLUSION

This study aimed to evaluate the performance of seven statistical approaches for vertical wind speed extra–polation, including GLM, LR, SVM, GAM, GPR, RT, and ER. The results showed that the GLM, ER, and RT models appeared to provide reasonable accuracy in estimating the WSs across different heights. However, the accuracy of other models, such as SVM and GAM, tended to decrease as the height increased. Additionally, some models showed bias towards over- or under-estimating wind speeds at certain heights, as indicated by the MPE and MBE values. GLM arguably performed well at higher heights even as it is the only method that achieved more than 60% R^2 value at 180 m.To enhance the performance of the tested models, the study suggests exploring these statistical models in combination with global optimization methods like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA). Future steps in this research domain could include applying the models to more diverse datasets, enhancing model robustness, and developing capabilities for real-time predictions, expanding the applicability and reliability of the models in various practical scenarios.

Computer program

All scripts are written using Matlab and can be accessed here https://github.com/hilalnuha/WSstatistics.

Acronyms and abbreviations

ER	Ensemble Regression
GAM	Generalized Additive Models
GLM	Generalized Linear Models
GPR	Gaussian Process Regression
LR	Linear Regression
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Bias Error
MPE	Mean Percentage Error
MSE	Mean Squared Error
NMSE	Normalized MSE
NRMSE	Normalized RMSE
R^2	R-squared
RMSE	Root Mean Squared Error
RT	Regression Tree
SMAPE	Symmetric Mean Absolute Percentage Error
SVM	Support Vector Machines

REFERENCES

[1] L. Fernández, "Global Wind Power Market – Sta–tistics & Facts," 2023. [Online]. Available: https://www.statista.com/topics/4564/global-wind-energy/.

[2] J. F. Newman, P. M. Klein, "The impacts of atmo–spheric stability on the accuracy of wind speed extrapolation methods," Resources, vol. 3, no. 1, pp. 81-105, 2014.

[3] T. R. et al. "A statistical analysis of wind distribution and wind power potential in the coastal region of South Africa," Int. J. Green Energy, vol. 10, no. 8, pp. 814-834, 2013.

[4] A. Tizpar, M. Satkin, M. B. Roshan, Y. Armoudli, "Wind resource assessment and wind power poten–tial of Mil-E Nader region in Sistan and Baluc–hestan Province Iran - Part 1: Annual energy esti–mation," Energy Convers. Manage., vol. 79, no. March, pp. 273-280,, 2014.

[5] D. Solyali, M. et al. "Wind resource assessment of Northern Cyprus," Renew. Sustain. Energy Rev., vol. 55, no. March, p. 2016., 180-187.

[6] D. Boro, H. Donnou, I. Kossi, N. Bado, F. Kieno, J. Bathiebo, "Vertical Profile of Wind Speed in the Atmospheric Boundary Layer and Assessment of Wind Resource on the Bobo Dioulasso Site in Burkina Faso," Smart Grid and Renewable Energy, pp. 257-278, 2019.

[7] D. Barantiev, E. Batchvarova, "Wind speed profile statistics from acoustic soundings at a black sea coastal site," Atmosphere, vol. 12, p. 1122, 2021.

[8] J. Steinheuer, P. Friederichs, "Vertical profiles of wind gust statistics from a regional reanalysis using multivariate extreme value theory," Nonlinear Processes in Geophysics, vol. 27, p. 239–252, 2020.

[9] N. Sucevic, Z. Djurisic, "Vertical wind speed pro–files estimation recognizing atmospheric stability," in 2011 10th International Conference on Environ–ment and Electrical Engineering, 2011.

[10] M. Mohandes, S. Rehman, H. Nuha, M. S. Islam, F. H. Schulze, "Accuracy of wind speed predictability with heights using Recurrent Neural networks," FME Transactions, vol. 49, p. 908–918, 2021.

[11] H. Nuha, M. et al. "Vertical wind speed extrapola–tion using regularized extreme learning machine," FME Transactions, vol. 50, p. 412–421, 2022.

[12] T. Kaneko, A. Uehara, T. Senjyu, A. Yona, N. Urasaki, "An integrated control method for a wind farm to reduce frequency deviations in a small power system," Applied Energy, vol. 88, p. 1049–1058, 2011.

[13] B.P. Rašuo, A.Č. Bengin, "Optimization of wind farm layout," FME Transactions, vol. 38, p. 107–114, 2010.

[14] B. Rašuo, A. Bengin, A. Veg, "On aerodynamic optimization of wind farm layout," PAMM, vol. 10, p. 539–540, 2010.

[15] B. Rašuo, M. Dinulović, A. Veg, A. Grbović, A. Bengin, "Harmonization of new wind turbine rotor blades development process: A review," Renewable and Sustainable Energy Reviews, vol. 39, p. 874–882, 2014.

[16] L. Kumle, M. L.-H. Võ, D. Draschkow, "Estima–ting power in (generalized) linear mixed models: An open introduction and tutorial in R," Behavior research methods, vol. 53, no. 6, pp. 2528-2543, 2021.

[17] S. Barhmi, O. Elfatni, I. Belhaj, "Forecasting of wind speed using multiple linear regression and artificial neural networks," Energy Systems, vol. 11, pp. 935-946, 2020.

[18] K. Ravindra, P. Rattan, S. Mor, A. N. Aggarwal, "Generalized additive models: Building evidence of air pollution, climate change and human health," Environment international, vol. 132, p. 104987, 2019.

[19] X. Mi, H. Liu, Y. Li, "Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine," Energy conversion and management, vol. 180, pp. 196-205, 2019.

[20] H. Cai, X. Jia, J. Feng, W. Li, Y.-M. Hsu, J. Lee, "Gaussian process regression for numerical wind speed prediction enhancement," Renewable energy, vol. 146, pp. 2112-2123, 2020.

[21] M. S. Elbisy, A. M. Elbisy, "Prediction of signi–ficant wave height by artificial neural networks and multiple additive regression trees," Ocean Engine–ering, vol. 230, p. 109077, 2021.

[22] M. Karthikeyan, R. Rengaraj, "Short-term wind speed forecasting using ensemble learning," in 2021, 7th International Conference on Electrical Energy Systems (ICEES), 2021.

[23] D. Chicco, M. J. Warrens, G. Jurman, "The coef–ficient of determination R-squared is more infor–mative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, vol. 7, p. e623, 2021.

[24] H. Petetin, D. Bowdalo, A. Soret, M. Guevara, O. Jorba, K. Serradell and C. P. García-Pando, "Meteorology-normalized impact of the COVID-19 lockdown upon NO 2 pollution in Spain," Atmospheric Chemistry and Physics, vol. 20, no. 18, pp. 11119-11141, 2020.

[25] A. A. Poli and M. C. Cirillo, "On the use of the normalized mean square error in evaluating dispersion model performance," tmospheric Environment. Part A. General Topics, vol. 27, no. 15, pp. 2427-2434, 1993.

[26] X. Hu, Z. Xie, F. Liu, "Assessment of speckle pattern quality in digital image correlation from the perspective of mean bias error," Measurement, vol. 173, p. 108618, 2021.

[27] I. Karakoti, B. Pande, K. Pandey, "Evaluation of different diffuse radiation models for Indian stati–ons and predicting the best fit model," Renewable and Sustainable Energy Reviews, vol. 15, no. 5, pp. 2378-2384, 2011.

ЕКСТРАПОЛАЦИЈА ВЕРТИКАЛНЕ БРЗИНЕ ВЕТРА КОРИШЋЕЊЕМ СТАТИСТИЧКИХ ПРИСТУПА

**Х.Х. Нуха, А. Балгонаим, Р.Р. Пахлеви,
С. Рехман, М. Мохандес**

Ветроелектрана је у последње време доживела значајан пораст и популарност, а најновије статистике указују да овај сектор и даље напредује. Међутим, један од суштинских корака у развоју пројеката енергије ветра је проналажење одговарајућих локација за ветроелектране, што укључује разумевање природе брзине ветра, правца ветра, терена и утицаја на животну средину. Да би се предвидела производња енергије ветра током очекиваног животног века ветроелектране, неопходна је вертикална екстраполација брзине ветра на висину чворишта ветротурбине. Стога, ова студија представља свеобухватну процену седам статистичких приступа за вертикалну екстрапола– цију брзине ветра, укључујући генерализоване линеарне моделе (ГЛМ), линеарну регресију (ЛР), машине за векторе подршке (СВМ), генерализоване адитивне моделе (ГАМ), регресију Гаусовог процеса (ГПР), стабло регресије (РТ) и регресија ансамбла (ЕР). Тачност ових метода се процењује коришћењем метрика перформанси као што су средња квадратна грешка (МСЕ), средња средња квадратна грешка (РМСЕ), нормализована РМСЕ (НРМСЕ), нормализована МСЕ (НМСЕ), средња грешка пристрасности (МБЕ), средња апсолутна грешка (МАЕ), средња процентуална грешка (МПЕ), средња апсолутна процентуална грешка (МАПЕ), симетрична средња апсолутна грешка у процентима (СМАПЕ) и Р-квадрат (Р2). Студија закључује да, у просеку, ГЛМ ради најбоље од свих седам статистичких метода.