https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

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

Influence of atmospheric stability on the *wind speed–energy–revenue* forecasting chain

Linda Canché-Cab[1], Michel Rivero[2], Ali Bassam[3], Liliana San-Pedro[3], and Mauricio Alberto Escalante Soberanis[3]

[1]Posgrado en Energías Renovables, Facultad de Ingeniería, Universidad Autónoma de Yucatán, Mérida, Yucatán, 97203, México
[2]Instituto de Investigaciones en Materiales, Unidad Morelia, Universidad Nacional Autónoma de México, Morelia, Michoacán, 58190, México
[3]Facultad de Ingeniería, Universidad Autónoma de Yucatán, Mérida, Yucatán, 97203, México

Correspondence: Michel Rivero (mrivero@materiales.unam.mx) and Mauricio Alberto Escalante Soberanis (mauricio.escalante@correo.uady.mx)

Abstract. Ultra-short-term wind forecasting ($<$ 30 min) is essential for the reliable operation and economic planning of wind energy systems, particularly in regions with high atmospheric variability. Atmospheric stability influences wind dynamics and predictive accuracy, but its role in propagating errors along the wind-energy-revenue chain has received limited attention. To address this gap, the relationship is qualitatively evaluated at a tropical site on the Yucatán Peninsula using 10-minute SoDAR measurements collected over a year of climatic variability. Representative forecasting approaches with different learning mechanisms, including persistence, support vector regression (SVR), and light gradient boosting machine (LGBM), were assessed across various forecast horizons, lag structures, and predictor configurations that included turbulence-related variables. The findings indicate that atmospheric stability significantly influences both prediction performance and the consequences of prediction errors. LGBM consistently outperforms both persistence and SVR, particularly in multi-step forecasting scenarios, and exhibits low sensitivity to lags and predictors. In contrast, SVR shows a stronger dependence on model configuration and improves when turbulent variables (turbulence intensity, turbulent kinetic energy, and vertical velocity) are incorporated, especially under stable conditions. Propagating forecast errors through a representative power curve and actual market prices shows that small biases near rated power can cause revenue deviations of 2.3-4.8%. These results provide a systematic assessment of how atmospheric stability shapes forecasting skill and the operational and economic risk of ultra-short-term wind power forecasting.

1 Introduction

Accurate wind forecasting is essential for the safe and cost-effective integration of wind power into modern electricity systems with high renewable generation rates. Forecast errors affect how operations are planned, reserves are set aside, energy is scheduled, and markets are involved, and these effects become more important as the use of renewable energy increases (Shoaei et al., 2024). Although forecasting is often evaluated using statistical measures such as RMSE or MAE, these do not always capture the real operational or economic impact of forecast errors (Aslam et al., 2025). Even small errors in wind speed

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

(WS) forecasts can lead to much larger differences in energy production and revenue since wind power conversion is inherently nonlinear and electricity prices vary over time.

Numerous forecasting approaches have been developed for wind energy applications, including persistence models, numerical weather prediction (NWP), statistical time-series methods such as ARIMA, and more recently, machine learning (ML) techniques (Singh and Rizwan, 2021). ML algorithms, such as artificial neural networks (ANN), tree-based algorithms, fuzzy systems, evolutionary approaches, and deep learning (DL) architectures, have demonstrated strong predictive performance in areas such as grid management, WS and power forecasting, and predictive maintenance (Singh and Rizwan, 2021; Aslam et al., 2025; Manohar et al., 2025; Dörterler et al., 2024). Despite these advances, many forecasting frameworks continue to rely on simplified representations of atmospheric processes, limiting their ability to capture variability in atmospheric boundary layer (ABL) dynamics.

Atmospheric stability is the ability of the atmosphere to resist vertical motion in the ABL, and is classified as stable, unstable, or neutral. Stable conditions suppress convection and turbulence, unstable conditions feature strong vertical mixing driven by surface heating, and neutral conditions lie between these two regimes (Stull, 1988). Although this classification is well established in boundary-layer meteorology, many operational forecasting approaches and engineering applications still assume neutral atmospheric conditions or homogeneous turbulence (Veers et al., 2023). Such simplifications are increasingly limiting for modern wind turbines, whose large rotor diameters interact with a significant portion of the ABL under variable or combined stability regimes. Atmospheric stability modifies the vertical wind profile and the intensity of turbulent fluctuations, directly affecting the operational efficiency of utility-scale wind turbines and wind parks.

Research indicates that the wind turbine power curve is highly sensitive to these conditions, with the most pronounced biases occurring in the transient region between cut-in and rated wind speeds. Enhanced wind shear and frequent low-level jets can increase power production by 15-24% for a given WS hub height under stable stratification, opposed to unstable regimes (Liu and Stevens, 2021). Conversely, unstable conditions typical of tropical environments promote vertical mixing, flattening the velocity profile and increasing turbulence intensity (TI). In turn, this could reduce power capture while introducing an asymmetric torque distribution across the rotor due to buoyant-driven upward flows (Pérez Albornoz et al., 2022).

Several studies have incorporated atmospheric stability and turbulence variables into ML models for wind forecasting, with outcomes that differ according to the selected predictor. Nielson et al. (2020) demonstrated that atmospheric stability conditions influence the performance of the prediction of ANN-based wind power, with the highest forecast errors observed under stable conditions. Optis and Perr-Sauer (2019) reported that the integration of variables related to turbulence into the model, especially turbulent kinetic energy (TKE), improves production forecasts rather than simply changing the learning algorithm. Vassallo et al. (2021) found that variables associated with turbulence and vertical transport enhance short-term WS forecasting. Kim and Kim (2023) reported that classical stability parameters, such as the Richardson number, do not consistently improve performance when used directly as input variables due to the difficulty models have in learning abrupt transitions between atmospheric regimes. Instead, TKE and TI proved to be more useful predictors. More recently, Ho et al. (2023) and Wan et al. (2025) demonstrated that physically significant predictors associated with atmospheric stratification improve the forecasting of the WS and the estimation of the vertical wind profile on various temporal and spatial scales. This is especially significant, as

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

wind farm power production has been found to vary by up to 15%, depending on the atmospheric stability regime (Wharton and Lundquist, 2012), which could result in energy production deviations of approximately 8-16% (Pérez et al., 2023).

Key studies that include atmospheric stability or turbulence in WS or power forecasts are presented in Table 1. These studies consistently indicate that stability-related variables improve forecasting accuracy, particularly under stable or highly stratified conditions, rather than increasing model complexity. However, quantifying the impact of stability-dependent WS forecast errors on power and revenue remains underexplored.

Table 1. Studies on the incorporation of atmospheric stability in wind energy forecasting

Study	Target variable	Stability/turbulence variables	Model
Nielson et al. (2020)	Wind power	WS, air density, TI, Richardson number, wind shear	ANN Feed Forward Back Propagation (FFBP)
Optis and Perr-Sauer (2019)	Wind power	TKE, TI, temperature, air pressure, potential temperature gradient, wind shear, Monin–Obukhov length	Generalized additive model, ANN, GBM (Gradient Boosting Machine), extremely randomized trees, SVM (Support Vector Machine)
Vassallo et al. (2021)	WS	TI, vertical velocity, turbulent heat flux, wind direction	ARIMA-Random Forest (RF)
Optis et al. (2021)	Vertical WS profile	Air–sea temperature difference, surface wind speed, air pressure, wind direction	RF
Kim and Kim (2023)	Power / daily energy	Richardson number, TKE, TI, dissipation rate	SVM, RF, LightGBM (LGBM), ANN
Ho et al. (2023)	WS	Air–sea thermal gradients	RF, ANN, persistence
Wan et al. (2025)	Vertical WS profile	WS at 10 m, temperature and relative humidity	ANN

Addressing this gap could improve decision-making and financial outcomes, as a scarce amount of studies have examined how forecast errors under different atmospheric stability regimes propagate throughout the chain, from the WS to energy production and revenue. Moreover, it remains unclear whether the improvements associated with stability-aware modeling translate into profitable reductions in energy deviations and economic risk. This topic is important in the energy sector, where the value of forecasts depends on their impact on system operation, costs, and market performance. It also contributes to the reliable and cost-effective integration of wind energy into future low-carbon electrical systems. Its importance is increasing as climate change alters atmospheric stratification, wind patterns, and boundary-layer dynamics, thereby changing the availability and variability of renewable resources and their economic value (Zhu et al., 2025).

The main objective of this study is to quantify how atmospheric stability influences the propagation of short-term WS forecast errors along the wind-energy-revenue chain. For this purpose, representative forecasting approaches with different learning mechanisms are evaluated: Persistence, Support Vector Regression (SVR), and Light Gradient Boosting Machine (LGBM). Ten-minute SoDAR measurements at 30 m and 80 m in a tropical environment are used to analyze different forecast

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

horizons, lag structures, and predictive configurations, including turbulence variables, and to assess how performance varies across atmospheric stability regimes. Forecast errors are propagated through a representative turbine power curve and actual electricity market prices to quantify their impact on energy estimation and revenue variability. The study covers the June-October 2023 period, which coincided with the transition from La Niña to El Niño, which led to an unusually long rainy season and altered the frequency of tropical waves and cyclones (Coordinación General del Servicio Meteorológico Nacional et al., 2024), allowing the evaluation of model robustness under particularly complex atmospheric conditions.

The article is organized as follows: Section 2 details the study site and data preparation; Section 3 describes the modeling methodology and energy estimation; Section 4 discusses forecast performance across stability regimes and associated economic risks; and Section 5 summarizes the main conclusions and presents the future work.

2 Data description

This section presents the five-stage WS forecasting methodology: (i) data collection, (ii) preprocessing, (iii) model construction, (iv) forecast evaluation, and (v) economic risk analysis for power generation. The methodology is illustrated in Fig. 1 and detailed in the following sections.

2.1 Measuring site

Data were collected at the Autonomous University of Yucatan (21.049726°N, 89.644566°W), located in the northwestern Yucatan Peninsula, 30 km from the coast (Fig. 2). The regional topography is predominantly flat, with elevations mostly below 50 m a.m.s.l. (above mean sea level). Tropical moist, semidry, warm and dry, and semiarid climates coexist in the region (Rivero et al., 2017).

Regional atmospheric dynamics are strongly modulated by large-scale synoptic forcings, particularly *"Nortes"* or cold surges, associated with cold fronts crossing the Gulf of Mexico. Frequent from autumn to spring, these events cause abrupt changes in wind intensity, direction, and atmospheric stability (Allende-Arandía et al., 2020). Consequently, the site provides a representative environment for evaluating forecast sensitivity and its implications in energy production under different stability regimes in a tropical–subtropical context, which has been scarcely studied.

2.2 Data collection

A Scintec SFAS (Flat Array Sodar) acoustic profiler (hereafter SoDAR) was employed to characterize vertical wind and turbulence profiles in the ABL. Remotely measure winds up to approximately 500 m altitude, operates from 2525 to 4850 Hz with up to nine beams at different angles, and offers a minimum 5 m vertical resolution to reconstruct three-dimensional wind and turbulence-related parameters.

The instrument is installed on a building roof, 3 m above the ground, in an urban area with scattered vegetation (see inset in Fig. 2). It recorded data every 10 minutes from June to December 2023 at heights from 15 to 200 m. The vertical resolution was 5 m from $15-35$ m and 15 m from $50-200$ m, with WS resolution of $0.1-0.3\ \mathrm{ms^{-1}}$, respectively. The research used data at

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

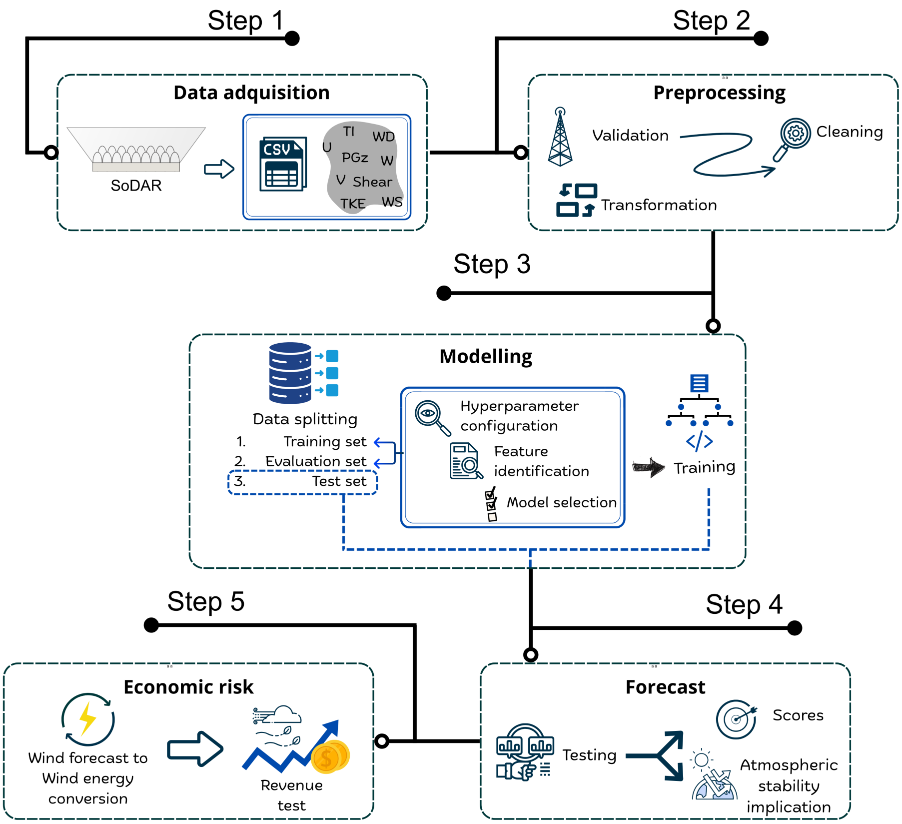

Figure 1. Methodological diagram.

two heights to capture different ABL dynamics driving mechanisms and stability regimes. The lower height, located within the surface layer, is pertinent for small and urban wind turbines, while the upper height minimizes direct influences from surface mechanical and thermal forcing, aligning more with utility-scale wind turbines. Thus, the proposed methodology is evaluated under two prevailing conditions and atmospheric stability regimes, providing a robust framework to assess its performance. Moreover, both heights must be selected considering data availability and completeness.

2.3 Data preprocessing and atmospheric stability classification

Reliable forecasting requires proper data handling. Thus, a preprocessing method must ensure data quality and consistency before any model implementation. The methodology used in this work is detailed in the following subsections.

2.3.1 Data quality control

The SoDAR operated continuously from June to December 2023. These measurements were cross-validated with a nearby meteorological tower (≈ 19 m) equipped with ultrasonic anemometers. The consistency between both systems was evaluated

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

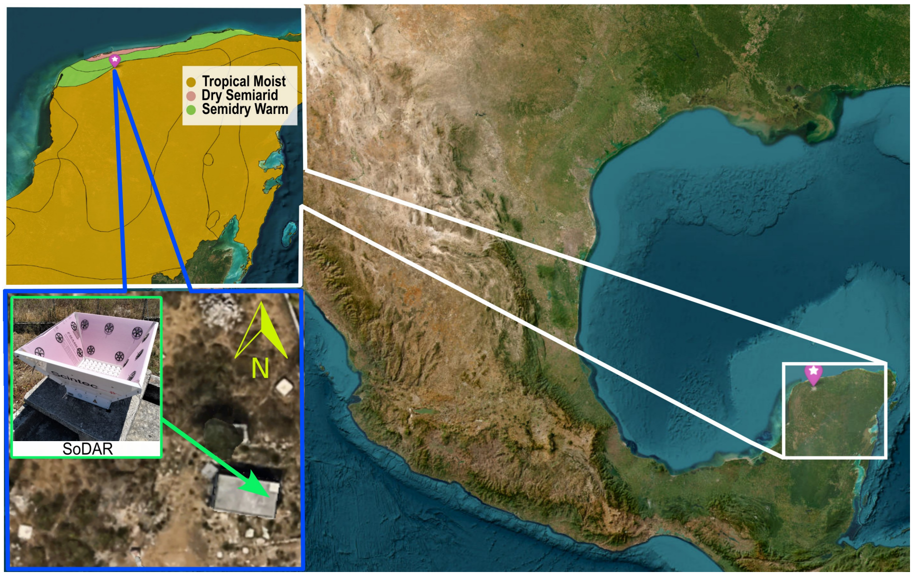

Figure 2. Location of the monitoring station. Maps data: © OpenStreetMap 2026.

using monthly correlations, and those months with coefficients greater than 0.70 were considered reliable. Only the period from June to October met this criterion. The final dataset comprises 153 days (22,032 10-minute records). The correlation plots are provided in Figs. S1 and S2 in the Supplementary Material (SM). Table 2 summarizes the recorded meteorological variables.

Table 2. SoDAR measurements at 30 m and 80 m, including their minimum and maximum values

Feature	Abrev.	Min/Max	Unit
Wind speed	WS	$0.01/12.91$	m s^{-1}
Wind direction	WD	$0/360$	Degree
Easterly wind speed	U	$-11.51/9.69$	m s^{-1}
Northerly wind speed	V	$-12.80/10.32$	m s^{-1}
Vertical wind speed	W	$-0.89/1.07$	m s^{-1}
Turbulence intensity	TI	$5.30 \times 10^{-7}/4.42$	–
Turbulent kinetic energy	TKE	$0.01/2.05$	m^2 s^{-2}
Shear	WSH	$9.72 \times 10^{-4}/0.28$	m s^{-1}
Eddy dissipation rate	EDR	$0/0.02$	m^2 s^{-3}
Pasquill-Gifford stability class	PG	From A to F	Categorical

Heights at 30 and 80 m were selected due to their high data completion rates and distinct atmospheric regimes, providing a robust framework to assess the influence of atmospheric stability on the prediction of WS and revenue estimation. Figure 3 shows data availability at 30 and 80 m for each variable. Although imputation is generally reliable when less than 30% of data

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

are missing (Mantuano et al., 2025), this study adopted a conservative threshold of 20%. The total data loss was $\approx 12\%$ at both heights. Only the turbulence intensity (TI) exceeded the 20% threshold (red line), since TI relies on higher-order statistics and is more sensitive to low signal-to-noise ratios (Bradley, 2007).

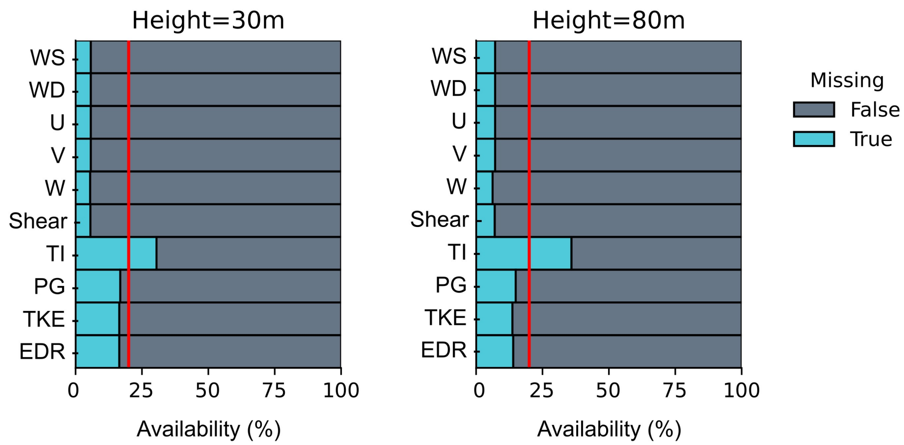

Figure 3. Data availability from the SoDAR at 30 m and 80 m for each measured variable during June-October 2023. The red line marks the 20% missing data threshold.

The models implemented in this work require complete datasets. To maintain data integrity for the models, missing values were restored using two complementary procedures: (i) iterative multivariate imputation via Random Forest (RF), applied when at least three variables were available per record ($\approx 6.4\%$ of data at 30 m and $\approx 6.38\%$ at 80 m); and (ii) two-dimensional interpolation in the hour–day space used for gaps between 10 minutes and 17 hours ($\approx 5.09\%$ for 30 m and $\approx 5.65\%$ for 80 m). These imputed values were marked as *reconstructed* and weighted during training to reduce their influence.

The processed dataset (original and reconstructed records) was chronologically divided into three subsets: June–September for training (80%), the first 20 days of October for validation (13%) and the final 11 days for testing (7%). The test dataset remained independent, with imputation rates below 0.4% for wind variables and 1–3% for turbulent variables. Previous benchmarks show that RF imputation yields low Mean Absolute Errors (MAE) ($0.09 - 0.15$ ms^{-1}) for similar missing data rates (Canché-Cab et al., 2026). For 2D interpolation, the MAE ranged from $0.70 - 2.30$ ms^{-1}, depending on the temporal gap and atmospheric conditions, aligns with sophisticated AI models (Boujoudar et al., 2025). These low error margins and the weighting strategy ensure that imputation does not bias the final results.

2.3.2 Transformation and codification

The variables in the dataset (Table 2) have different scales, so we standardized them to avoid bias, balance their influence, and support stable model convergence. Standardization was performed using the *StandardScaler* from *scikit-learn*, which centers the data at zero mean and scales it to unit variance.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

Atmospheric stability was measured directly by the SoDAR system using the Pasquill-Gifford (PG) classification derived from the standard deviation of the wind elevation angle (ScintecAG, 2017). This classification delineates six categories, from extremely unstable (A) to extremely stable (F) conditions. Although stability schemes and classification criteria vary (Pérez Albornoz et al., 2022), this work adopts the SoDAR-derived PG classification as the reference, supported by empirical validation and widespread regulatory adoption (U.S. Environmental Protection Agency, 2000).

To mitigate class imbalance typical of atmospheric data, categories were regrouped into three classes: unstable (U: A, B, C), neutral (N: D), and stable (S: E, F). This reclassification (Fig. 4) was used exclusively for performance stratification and interpretation of forecast errors. It was not included as a predictor in any forecasting model, allowing the influence of atmospheric stability to be evaluated independently of model training.

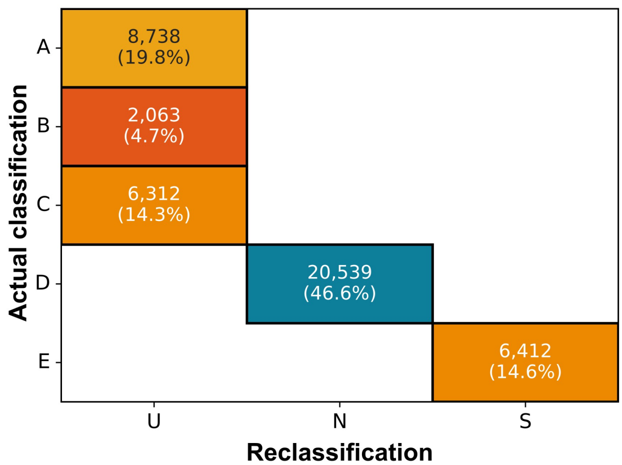

Figure 4. Correspondence between the original Pasquill–Gifford classification (A–F) and the reassignment into three categories (U, N, S) at both heights. Each cell indicates the number of records and their relative proportion (%) with respect to the total number of cases.

Figure 5 illustrates the relationship between WS and atmospheric stability. At 30 m, weak winds (< 2 ms^{-1}) are mainly associated to unstable conditions. As the WS increases, the flow becomes mostly neutral up to about 8 ms^{-1}, and for higher values the stable condition predominates. In turn, at 80 m, the unstable condition persists for < 2 ms^{-1}, as for 30 m. However, for $4 \gtrsim$ WS $\gtrsim 6$ms^{-1} stable and neutral conditions were found, and for higher wind speeds the atmosphere is predominantly neutral. These profiles align with marine and coastal observations, such as the Yucatan region (Sathe et al., 2013; Nybø et al., 2020), which supports the selection of these two heights. Figure 5 also illustrates the WS distribution (solid black line). As expected, the mean WS at 80 m is higher than at 30 m due to mechanical shear, which controls how WS increases with height (Pérez Albornoz et al., 2022).

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

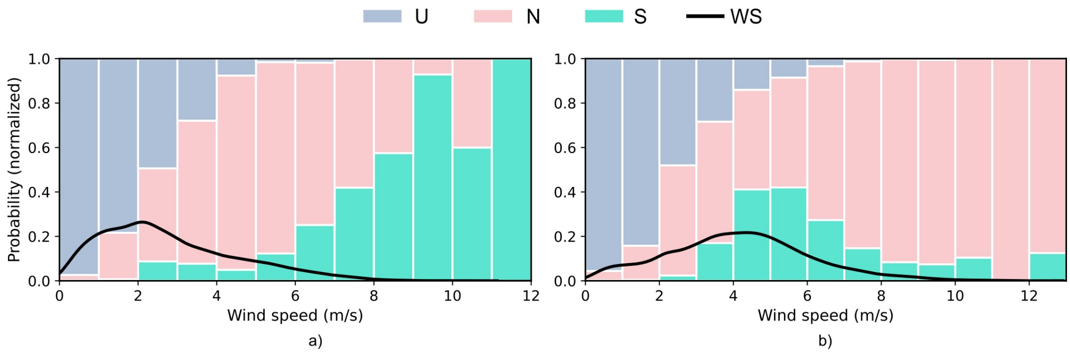

Figure 5. Stability distributions based on PG classification measured at (a) 30 m and (b) 80 m, and the corresponding WS distribution (solid black line).

3 Model construction

This section details the model-building process. It first describes the selected algorithms, then the methods for hyperparameter optimization and feature importance, and finally introduces the multi-step-ahead forecasting strategy.

3.1 Algorithms

This study analyzes two ML algorithms widely used in wind forecasting tasks: SVR and LGBM (Arslan Tuncar et al., 2024). *Persistence* is taken as the reference model. Below are the main characteristics of these algorithms.

– **Persistence.** The persistence method is the simplest forecasting model commonly implemented in atmospheric sciences. It assumes that the current weather conditions will *persist* in the immediate future (Wilks, 2020). Thus, the forecasted WS $v(t)$ can be expressed as follows:

$$v(t) = v(t - \Delta t) \tag{1}$$

where $v(t - \Delta t)$ is the velocity measured at the previous instant Δt. Since this model exhibits good performance over ultra-short-term horizons (Singh and Rizwan, 2021), it is used as a benchmark to evaluate the ML models.

– **Support vector regression (SVR).** SVM allow to model complex relationships through kernel functions that project the data into higher-dimensional spaces where optimal boundaries can be identified. SVR extends its applicability to continuous problems and applies the same principle to estimate numerical variables using kernels such as linear, polynomial, or radial basis function (RBF), which facilitates capturing nonlinear patterns with a structurally simple model (Hearst et al., 1998). SVR aims to find a function that stays within a tolerance margin ϵ, where small deviations are not penalized. This approach, known as *epsilon-insensitive loss*, promotes more stable models with a lower risk of overfitting by focusing optimization only on the points that exceed that margin (Géron, 2023).

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

- **Light gradient boosting machine (LGBM).** Gradient boosting is an ensemble technique that combines multiple weak models, typically decision trees (DT), to iteratively improve predictive accuracy. Each tree is trained to correct the residual errors of the previous one, allowing the ensemble to progressively learn from its mistakes. Although this approach achieves high precision, traditional implementations can be computationally demanding for large or high-dimensional datasets (Géron, 2023; Hastie et al., 2009). LGBM is an optimized version designed to increase training speed and reduce memory usage while maintaining performance. It employs a leaf-wise growth strategy that expands leaves with the largest loss, and incorporates techniques such as Gradient-based One-side Sampling (GOSS), which retains the most informative instances, and Exclusive Feature Bundling (EFB), which reduces dimensionality by grouping mutually exclusive features (Ke et al., 2017).

3.2 Hyperparameter selection

Hyperparameter selection was conducted using bayesian optimization. Bayesian optimization represents the objective function (e.g., validation error) with a probabilistic surrogate that approximates its behavior and quantifies uncertainty. From previous evaluations, this model estimates the expected performance and uncertainty for each hyperparameter configuration. An acquisition function then selects the next hyperparameters by balancing exploration of uncertain regions and leveraging those with the best estimated performance (Feurer and Hutter, 2019). This iterative process reduces the number of expensive training evaluations required to approach an optimal configuration. The optimal values obtained for each algorithm are presented in Table 3.

Table 3. Bayesian optimization results for SVR and LGBM hyperparameters

Algorithm	Hyperparameter
SVR	kernel=rbf, c=1.249, epsilon=0.716, max_iter=110, gamma=scale
LGBM	learning_rate= 0.0884, max_depth= 28, n_estimators=201

Hyperparameter optimization was performed using all input variables to identify a robust hyperparameter configuration in the most complex scenario. The resulting hyperparameters were consistently applied to all evaluated models, including those trained with a reduced number of predictors. This approach allows us to directly compare the different scenarios and evaluate the impact of adding variables, while reducing the influence of variations introduced by repeated hyperparameter re-optimization.

3.3 Feature importance

A feature selection procedure was implemented to establish an order for incorporating the predictors into the model to analyze how sensitive the model's performance is to the inclusion of different atmospheric variables. Feature selection methods are commonly classified as filter, wrapper, and embedded methods (Guyon and Elisseeff, 2003). In this work, a combined strategy

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

based on filter and wrapper methods was adopted. A correlation analysis was applied as a filtering method to identify variables weakly related to the target variable and to detect potential redundancies among the predictors. This analysis enabled the elimination of highly correlated attributes, reducing multicollinearity and simplifying the initial set of variables. Subsequently, Recursive Feature Elimination with Cross-Validation (RFECV) was applied using LGBM as the estimator. RFECV iteratively removes the least relevant predictor according to cross-validated performance ($cv = 10$), using MSE as the optimization criterion. The resulting ranking was used only to guide the sequential inclusion of variables in the forecasting experiments; it was not intended to identify a single optimal feature subset.

3.4 Multi-step-ahead forecasting strategies

Forecast horizons in power systems are usually grouped into ultra-short term (seconds to 30 min), short term (30 min to 6 h), medium term (6 h to 1 day), and long term (several days to months), each associated with different system operation and planning needs (Kumar et al., 2026; Rivero et al., 2020). This study focuses on ultra-short-term horizons at 10 and 30 minutes (one- and three-steps-ahead), intervals critical for real-time turbine control and grid stability (Singh and Rizwan, 2021; Arslan Tuncar et al., 2024).

To address multi-step forecasting, this work adopts a direct strategy in which an independent model is trained for each future step. This avoids error propagation, enables unbiased evaluation across horizons, and yields more reliable results in highly variable, nonlinear wind regimes (Omidkar et al., 2025; Vassallo et al., 2020). Furthermore, to capture multi-scale atmospheric dynamics, we implemented a lag structure with six time windows: 1, 2, 4, 6, 12, and 24 hours (corresponding to 6, 12, 24, 36, 72, and 144 observations). The shorter lags capture rapid variations relevant to real-time control, while the longer lags incorporate information associated with the diurnal cycle.

3.5 Model performance

This subsection presents the methodological framework for assessing and comparing forecasting model performance, including the temporal validation strategy, evaluation metrics, and statistical tests for significant model differences.

3.5.1 Temporal validation strategy

A time-based validation was conducted using backtesting to evaluate the models' performance. Backtesting is a variant of cross-validation for time series, but instead of splitting the data into random subsamples, it uses a temporal cut-off point, as illustrated in Fig. 6. The adopted configuration was based on retraining with a fixed rolling window. Namely, all forecasted values are obtained with a model trained with the n previous observations, with n being constant. In this study, the *Skforecast* Python library (Amat Rodrigo and Escobar Ortiz, 2025) was used for hyperparameter optimization, model training, and backtesting. There was variable standardization embedded in the forecasting pipeline and updated in the training set at each iteration to prevent temporal information leakage.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

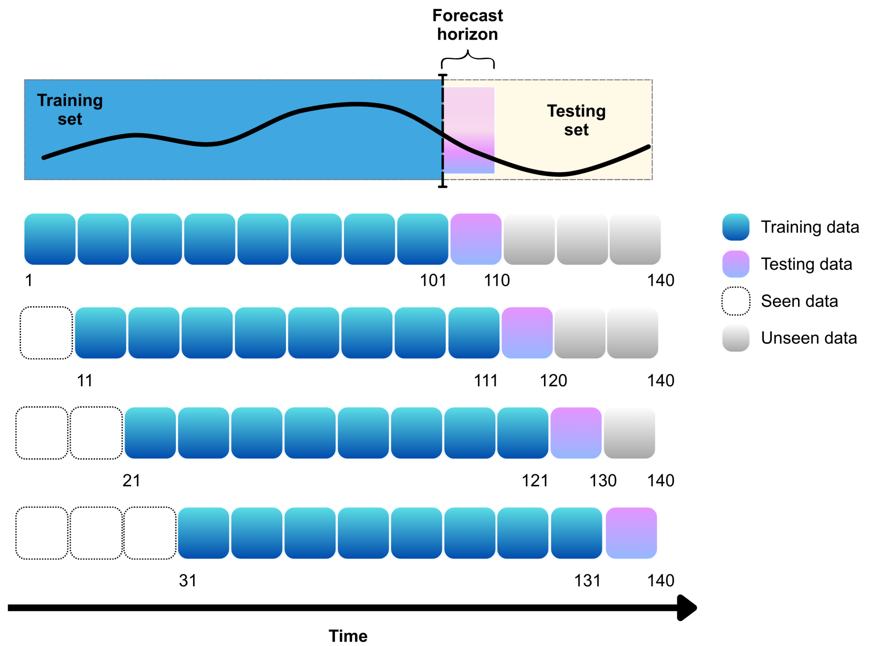

Figure 6. Backtesting with a fixed rolling window: at each step, the model is retrained on data up to a cut-off date and tested on subsequent periods. The cut-off date then moves forward and the process repeats.

3.5.2 Evaluation metrics

235 Table 4 summarizes the metrics used to quantify the residuals between the forecasts and observations. This table also includes general comments for each metric.

Table 4. Evaluation metrics

Abbr.	Full name	Definition	Comments
ME	Mean error	$\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)$	It measures the direction of bias, and a poor forecast can still get a perfect score if its errors cancel out
RMSE	Root mean square error	$\sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}$	Sensitive to outliers, comparable to the target variable
NRMSE	Normalized RMSE	$\frac{RMSE}{y_{max} - y_{min}}$	Normalizes error so values from different data scales can be compared
MAE	Mean absolute error	$\frac{1}{N}\sum_{i=1}^{N}\lvert(\hat{y}_i - y_i)\rvert$	Does not strongly penalize major errors
R^2	Coefficient of determination	$1 - \frac{\sum\limits_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum\limits_{i=1}^{N}(y_i - \overline{y})^2}$	It measures explained variance, not accuracy. A high value does not guarantee accurate predictions

N : total number of data, y : actual value, $\hat{y}_i$: forecasted value, $\overline{y}$: average of actual values

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

WIND ENERGY SCIENCE DISCUSSIONS
Open Access
european academy of wind energy

Conventional metrics offer a comprehensive evaluation of performance, but have limitations in operational forecasting contexts (Rivero et al., 2020). Specifically, they only provide information on error distribution, which may be less relevant to operators and stakeholders. To mitigate these limitations, we also compute the forecasting skill (FS) to assess the ML model's improvement relative to a reference model. The forecast skill is defined as

$$FS = 1 - \frac{\text{NRMSE}}{\text{NRMSE}_{\text{ref}}} \tag{2}$$

where NRMSE is the error of the evaluated model and $\text{NRMSE}_{\text{ref}}$ is the error of the reference model. FS ranges in $(-\infty, 1]$, with $FS = 0$ indicating performance equal to reference and positive values indicating improvement.

3.5.3 Diebold-Mariano test

The Diebold–Mariano test (*DM*) is widely used to compare the accuracy of wind power forecasting models (Carrera and Kim, 2024; Chen et al., 2014). It enables us to assess whether a forecasting model exhibits a statistically significant improvement over another. This test is particularly suitable for situations in which forecast errors exhibit temporal dependence. The test is based on the null hypothesis that two models have the same predictive ability. Let x_t be the observed series and $x_{(A,t)}$ and $x_{(B,t)}$ the forecasts generated by models A and B, respectively. Given a loss function $g\left(x_t, \hat{x}_{i,t}\right)$, the loss differential is defined as:

$$d_t = g\left(x_t, \hat{x}_{A,t}\right) - g\left(x_t, \hat{x}_{B,t}\right) \tag{3}$$

The sample mean of the loss differentials is calculated as:

$$\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t \qquad t = 1, 2, \ldots, n \tag{4}$$

The *DM* test statistic is defined as:

$$S = \frac{\bar{d}}{\sqrt{\frac{\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k}{n}}} \cong N(0,1) \tag{5}$$

where γ_k is the lag-k autocovariance of the loss differential, γ_0 is the lag-0 autocovariance, and *h* is the horizon. The inclusion of these autocovariances allows us to correct the variance for the temporal dependence of the forecast errors. The null hypothesis is rejected when the *p*-value associated with the *DM* statistic is below a significance level, indicating that one of the models exhibits significantly superior predictive performance compared to the other.

3.6 Power generation

This work evaluates wind power and revenue as downstream measures of predictive accuracy, clarifying how WS forecasting affects operational and economic outcomes, an area barely explored in the literature. We propagate forecast errors through the power curve ($P \propto v^3$) and published prices by the national market regulator to connect meteorological precision to grid-level financial outcomes. Consequently, a 500 kW turbine power curve and reported market prices were used as a simplified, physically interpretable basis to compare forecast configurations across atmospheric stability regimes.

For this purpose, wind power at 80 m (hub height) was estimated using the LTW90 power curve, with a 500 kW rated power (P_r), a cut-in speed (v_{ci}) of 3 ms^{-1}, a rated speed (v_r) of 8 ms^{-1}, and a cut-out speed (v_{co}) of 26 ms^{-1}. The power curve describes the nonlinear relationship between WS and the electrical power generated by the turbine. Since the manufacturer provides the power curve as discrete WS–power pairs, a continuous piecewise polynomial approximation ($p(v)$) was fitted: a fifth-order polynomial over the partial-load region (3-7 ms^{-1}) and a third-order polynomial in the transition to rated power (7-8 ms^{-1}). The turbine power output P_{wt} can be expressed piecewise as:

$$P_{wt} = \begin{cases} 0 & v < v_{ci},\ v > v_{co} \\ p(v) & v_{ci} < v < v_r \\ P_r & v_r \leqslant v \leqslant v_{co} \end{cases} \tag{6}$$

The output energy E_{out} (in watt-hours) for a series of N WS observations v_i with time interval Δt was calculated as:

$$E_{out} = \sum_{i=1}^{N} P_{wt} \cdot (v_i) \cdot \Delta t \tag{7}$$

Once the energy was estimated from the measured (baseline case) and the WS forecast at 80 m, the associated revenue was quantified using the hourly energy price published by CENACE (2025) for the Mexican electricity market. The energy estimate was initially obtained at a resolution of 10-minutes, consistent with the frequency of the wind data. Since energy prices are available at an hourly resolution, the 10-minute energy values were resampled to an hourly scale by adding consecutive 10-minute intervals for each hour.

4 Results and discussion

This section presents the performance of the forecasting models described above. The evaluation uses temporal validation during the test period.

4.1 Feature importance

Figure 7 illustrates the correlation between the variables in the dataset. This analysis revealed that WS has a significant relationship with variables associated with atmospheric stability, TI, and TKE. We identified strong correlations among certain predictors, notably between WD and the U component of the horizontal wind. However, as WD does not provide additional information beyond the U and V components, it was excluded from the final predictor set. In contrast, although wind shear and EDR exhibited correlation indicative of their shared physical processes, turbulence generation, and dissipation, they were retained for the analysis. These variables are central to the study, as they explicitly aim to assess their influence on forecast performance across different stability regimes.

The hierarchical ranking of the features (from one, most important, to six, least important) shows a consistent structure across both heights and data configurations (Fig. 8). In both cases, the horizontal wind components (U and V) were identified

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

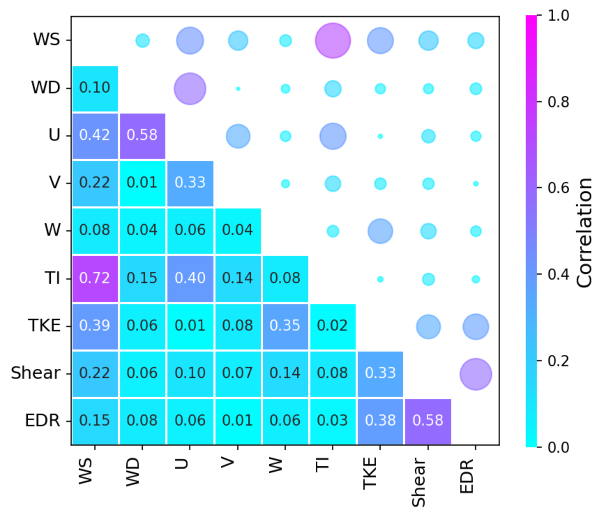

Figure 7. Correlation matrix.

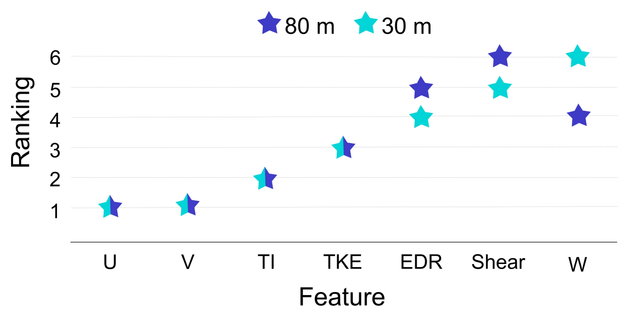

Figure 8. Ranked importance of characteristics at 30 m and 80 m. Scores range from 1 to 6, with 1 indicating highest importance.

as the dominant predictors, consistently ranking highest in importance. TI and TKE consistently occupied the next places in the ranking, respectively. These positions highlight the relevance of including turbulence-related variables in prediction. These findings align with Optis and Perr-Sauer (2019), who found that stability-related variables, especially TKE, are statistically relevant to explain power production in wind farms. Conversely, W, Shear, and EDR showed lower importance and greater differences with height. At 30 m, near the surface layer, surface interactions amplify turbulence, increasing the relevance of EDR compared to W and Shear. At 80 m, lower surface friction and more stable turbulence enhance the role of the vertical flow component (W). This can also be attributed to ABL dynamics, where phenomena such as Ekman pumping or vertical wind shear induce more pronounced vertical flows. A detailed investigation of these mechanisms extends beyond the scope of this work.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

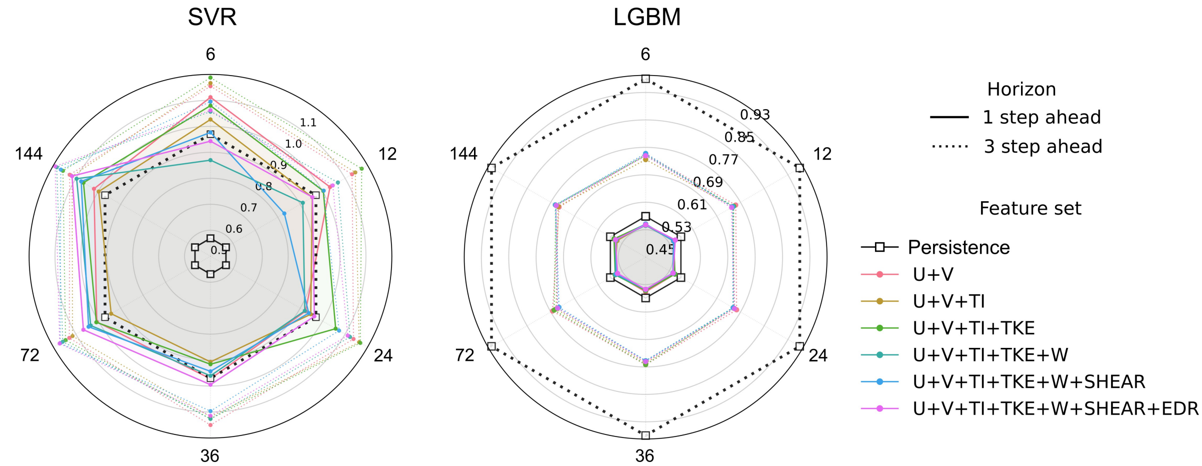

Figure 9. Radial plot of the MAE error corresponding to the forecasts at 80 m height for both models, all considered time windows (lags, $6-144$ steps), forecast horizon (10-min in solid line and 30-min dashed), and input variable configuration ($2-7$ predictors). Persistence (black square markers, same in both plots) is included as a reference (it is not defined for lags).

Based on this feature hierarchy, we define a sequential model configuration for the following analysis, adhering to the incremental approach outlined in Table 5. It is worth noting that, given the influence of the diurnal cycle on wind patterns, we also incorporate the hour variable, encoded using sine and cosine functions to capture its periodicity.

Table 5. Combinations of features used as inputs in the models

Set number	Feature set order
1	U+V
2	U+V+TI
3	U+V+TI+TKE
4	U+V+TI+TKE+W
5	U+V+TI+TKE+W+Shear
6	U+V+TI+TKE+W+Shear+EDR

4.2 Evaluation of model performance

Figure 9 presents the MAE errors at 80 m for both models, time windows, forecast horizon, and input variable configuration. The corresponding results for 30 m can be found in Tables S1 and S2 in SM. In the one-step-ahead (10-min) scenario, LGBM shows outstanding and consistent performance at both heights. At 80 m, LGBM achieved reductions of $\approx 8.12\%$ in RMSE, $\approx 4.74\%$ in MAE, and $\approx 8.60\%$ in NRMSE compared to persistence (RMSE = 0.86 ms^{-1}, MAE = 0.57 ms^{-1}, NRMSE = 0.09). These gains were even more pronounced at 30 m, where LGBM surpassed persistence by $\approx 10\%$ in RMSE and NRMSE, and $\approx 8.67\%$ in MAE. The stability of these metrics across various configurations suggests that LGBM effectively captures wind variability with minimal tuning. In contrast, SVR exhibited higher errors and significant sensitivity to configuration changes. At both heights, SVR underperformed persistence in the one-step-ahead scenario, with NRMSE up to 24.73% higher at 80 m and 5.71% higher at 30 m. This indicates a limited ability to model the nonlinear transitions characteristic of turbulent regimes at high temporal resolutions.

As the forecasting horizon extends, all models exhibit a degradation in performance. For example, the NRMSE at 30m increases by 28% from one-step to three-step forecasts, reflecting the inherent challenge of multi-step-ahead predictions (Aslam et al., 2025). LGBM maintains a significant advantage over the persistence baseline, with reductions of $\approx 24.44\%$ (RMSE), 24.12% (MAE) and 24.29% (NRMSE) at 30 m and similar reductions at 80 m. This advantage is also evident in the positive FS values (see Fig. S3 in the SM), with positive and increasing values (0.07 to 0.24) as the horizon extended. Conversely, SVR shows modest and inconsistent gains. At 30 m, it surpasses persistence by 14.95% in RMSE and NRMSE, and 7.25% in MAE. At 80 m, the gains are marginal, and MAE exceeds persistence. The FS is negative for one-step-ahead forecasts, slightly positive for three-step-ahead and 30 m forecasts, and inconsistent at 80 m, indicating configuration-dependent performance and weak robustness in multi-step forecasting.

The errors obtained in this study align with literature ranges for ultra-short-term wind forecasting. At 10-minute resolution, our LGBM model (RMSE ≈ 0.69–0.70 ms^{-1}; MAE ≈ 0.44–0.46 ms^{-1} at 30 m) performs similarly to the hybrid ARIMA–RF model (RMSE ≈ 0.71 ms^{-1}; MAE ≈ 0.51 ms^{-1}) (Vassallo et al., 2021). It has also been reported that SVM models can be competitive with hourly data (Zabihi et al., 2025); but their 10-minute results indicate that SVR exhibits larger errors and even performs worse than persistence, especially at greater heights, probably due to the kernel sensitivity to the non-stationary signal (Vassallo et al., 2020).

Operating at a 10-minute resolution situates this study within the Van der Hoven spectral gap, a transition region between mesoscale motions and high-frequency turbulence, where wind variability increases and signal persistence decreases (Escalante Soberanis and Mérida, 2015), making forecasting inherently more challenging than at coarser resolutions such as hourly data Vassallo et al. (2021). Although advanced deep learning may offer higher precision (Jiang et al., 2024), LGBM remains highly competitive due to its lower methodological complexity and reduced data requirements, supporting its practical utility in tropical environments.

4.2.1 Effect of lag (time window) and input features

The lag analysis highlights distinct algorithmic behaviors (full metrics in Tables S2 and S3 in the SM). LGBM is insensitive to window length, while SVR exhibited greater sensitivity and achieved better performance with short to intermediate lags (6-12 observations, or 12 hours). This suggests that for 10- to 30 minute horizons, recent data capture the main predictive signal and limit overfitting (Galarza-Chavez et al., 2025). This is consistent with the ABL dynamics driven by surface conditions on hourly timescales (Stull, 1988). Consequently, longer lags do not provide performance gains but increase the computational cost (see Figs. S4 in the SM).

Performance for a 6-observation lag was summarized using Taylor diagrams in Fig. 10. This diagram compares correlation, standard deviation (STD), and RMSE. Radial lines show the correlation coefficient from 0 to 1; points closer to the horizontal

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

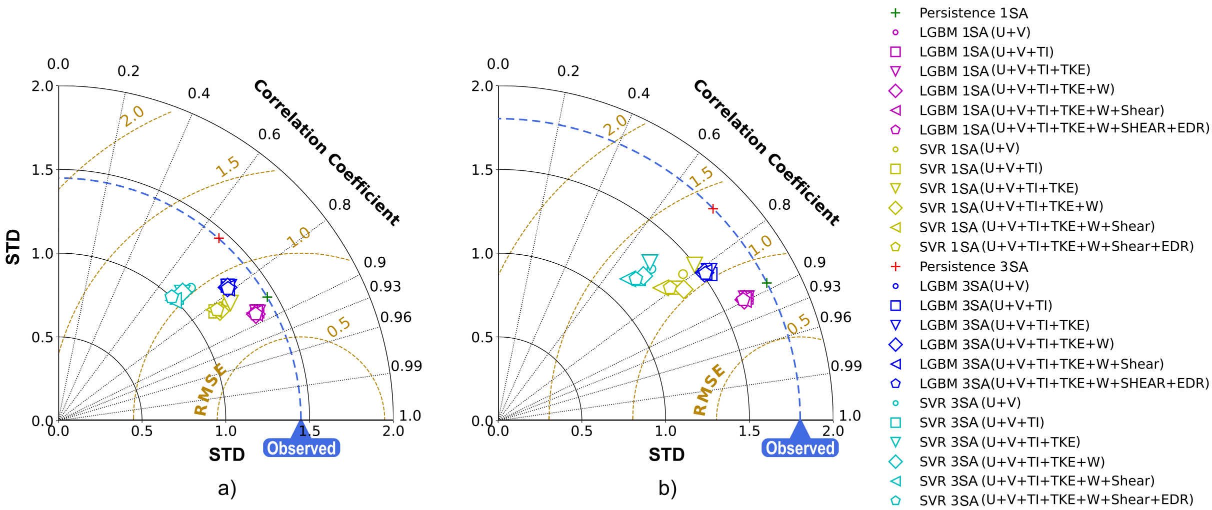

Figure 10. Taylor diagram of forecast WS using LGBM and SVR relative to persistence at a) 30 m and b) 80 m for one-step-ahead (1SA) and three-step-ahead (3SA) forecasts. Time window = 6.

line indicate a better model for that variable. Concentric arcs around the origin show STD. The blue dotted arc is the reference STD of the measured data ("Observed"). Concentric arcs around the reference point (blue solid triangle) indicate RMSE. In all cases, LGBM was consistently closer to the reference point than SVR and persistence, with the most pronounced improvements occurring at the 30-min horizon. In contrast, SVR exhibited only marginal improvements in persistence at the 30-min horizon, depending on the specific configuration of the predictor.

These findings are corroborated by the DM test, using the $U+V$ configuration as reference. We evaluated 30 combinations (5 variable sets $\times$ 6 lags) for each model and horizon, identifying configurations with significant predictive superiority ($p < 0.05$ and positive S statistic). These are visualized as colored ribbons in Fig. 11, with purple ribbons denoting statistically significant improvements. The complete *DM* test statistics are presented in Table S3 in the SM. Figure 11 reveals that SVM models show gains under specific conditions: the $U+V$-only benchmark outperforms expanded predictor sets, especially at lags of 6, 12 and 36. For other lags, the performance improves with added predictors, especially with *TKE*. This suggests that forecast performance depends not only on the algorithm or the time horizon but also on how effectively the predictor set represents the underlying state of the ABL. In particular, turbulence-related variables appear to provide complementary information on mixing and momentum transport processes that are not fully described by the mean horizontal wind components alone.

For LGBM, no statistically significant differences were observed in any configuration. This indicates that the minimal set predictor $U+V$ captures the dominant structure of the series, confirming the robustness of boosting-based models against redundant or weakly informative predictors (Ke et al., 2017). This agrees with Carrera and Kim (2024), who found that boosting algorithms remained competitive using only target and temporal variables. Other studies have also found that DT-based models barely benefit from variables linked to atmospheric stability at 10-minute temporal resolution. The influence of variables related

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

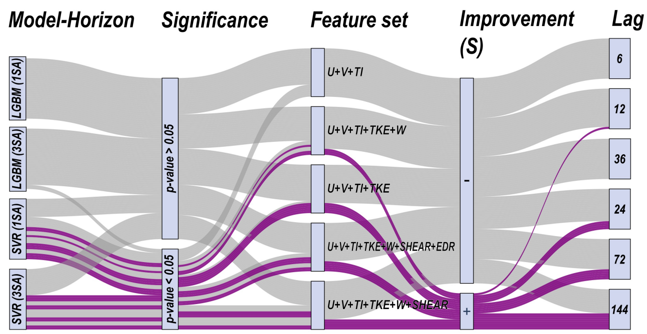

Figure 11. Comparative results based on the DM test, considering the corresponding model trained with $U+V$ as the reference. Significant difference for $p < 0.05$. Model enhancement with respect to the reference model: $S > 0$.

to atmospheric stability increases when analyzing hourly or daily averages, as these processes are allowed to develop (Kim and Kim, 2023; Vassallo et al., 2021), consistent with the patterns observed in this study.

4.3 Performance under different atmospheric stability regimes

Figure 12 illustrates the absolute error distribution at 80 m for a lag of 6 observations, disaggregated by stability regimes and predictor sets. The distribution is represented through a violin plot.

For SVR at the three-step-ahead horizon, class S yields the highest errors (MAE: $1.10 - 1.40$ ms^{-1}) exceeding the neutral (MAE: $0.98 - 1.13$ ms^{-1}) and unstable (MAE: $1.00 - 1.20$ ms^{-1}) regimes. Reducing the forecasting horizon lower errors across all classes (by $\approx$ 18% in S, 8% in N, and 16% in U). Moreover, S remains the most challenging scenario at 80 m, with a MAE up to 20% higher than in N and U. The larger errors under stable conditions reflect the complexity of the stable boundary layer, where intermittent turbulence, strong wind shear, and evening transitions create nonlinear dynamics that are difficult to capture. This shows that the accuracy of the SVR depends not only on the forecast horizon but also on the structure of the ABL, consistent with reports of higher power prediction errors under strongly stable regimes (Nielson et al., 2020).

In contrast, LGBM delivers more consistent performance across all stability regimes, with significantly lower and more constrained errors than SVR. Under S conditions, it achieves its lowest error for both horizons (MAE: ≈ 0.40 ms^{-1} for 10–min and ≈ 0.62 ms^{-1} for 30–min horizons), remaining below the overall mean error (dashed line). Errors increase slightly in the U and N regimes, but stay below those of SVR.

Similar findings appear at 30 m (Fig. S5 in SM). SVR shows the highest errors under U conditions (22% and 24% higher than under neutral and stable regimes, respectively), driven by surface roughness and obstacles that promote atmospheric instability. In contrast, LGBM remains relatively robust across heights (errors up to 2% and 8% under N and U in the three-step horizon, and 11% and 15% in the one-step horizon), indicating lower sensitivity to variations in the vertical wind profile and greater stability across heights.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

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

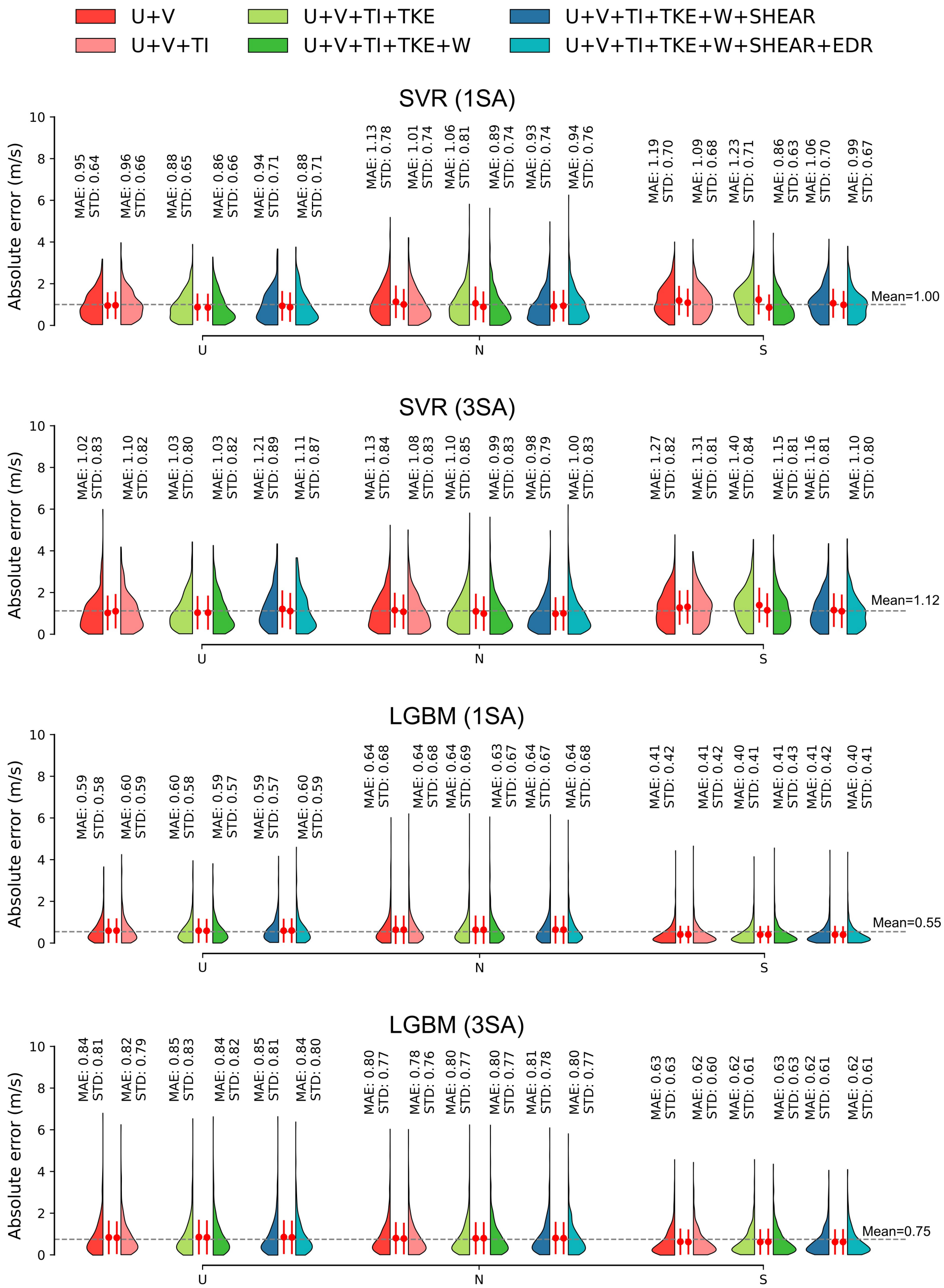

Figure 12. Violin plots of absolute error at 80 m for each configuration, with mean and standard deviation (red circle, vertical bar). The dotted line represents the average of the mean error values.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

The impact of predictor sets also varies by algorithm. For LGBM, expanding the input to $U+V+TI$ offers a marginal 2% improvement. However, SVR benefits substantially from turbulence-related variables, a configuration including $TI+TKE+W$ reduces MAE by $6-14\%$ at 30 m and up to 28% at 80 m compared to baseline $U+V$, with greater benefits under stable and neutral conditions (see Tables S4 and S5 in SM).

The sensitivity of the algorithms to turbulence-related predictors can be attributed to the fundamental physics of the ABL. Although horizontal wind components effectively describe mean advection and flow direction, they provide limited insight into the stochastic high-frequency turbulent processes that govern short-term wind variability (Pérez Albornoz et al., 2022). In this direction, predictors such as TI, TKE, and W characterize the dynamic state of the ABL by explicitly describing the intensity of fluctuations and the energy available for vertical mixing and momentum transport between atmospheric layers.

Under stable conditions, the physical interplay between buoyancy and shear forces becomes critical as they often reach the same order of magnitude. In this regime, buoyancy suppresses vertical mixing, and the evolution of the wind is dominated by intermittent shear-driven turbulence and localized phenomena, such as low-level jets and rapid transition events (Menken and Wildmann, 2026; Liu and Stevens, 2021; Barthelmie et al., 2015). These processes are inherently difficult to predict because atmospheric states with nearly identical mean WS can evolve along divergent trajectories. Consequently, turbulence variables provide essential predictive information not captured by the mean flow (namely, only by $U+V$). In contrast, unstable conditions promote intense vertical mixing, resulting in a more homogeneous and predictable wind field. This could explain why stable stratification is often the most challenging regime to forecast, particularly at higher altitudes like 80 m, where shear and stratification are more pronounced.

The significant performance gain observed in the SVR suggests that the kernel-based model benefits from a clear and explicit physical description of turbulence through variables related to turbulence. Conversely, the marginal improvement in LGBM indicates that tree-based ensembles may partially recover this information through complex non-linear interactions between basic predictors. This enhancement is consistent with previous research that identified TKE as one of the most relevant predictors for short-term wind and power forecasts (Optis and Perr-Sauer, 2019; Kim and Kim, 2023; Nielson et al., 2020).

Figure 13 illustrates the hourly distribution of the forecast error alongside the average energy price profile (dashed gray line). The maximum deviations for both heights occur between 16–18 hours, coinciding with transitions in the atmosphere state (N to U at 30 m; N to S at 80 m) and align with peak energy prices and high demand. The prevalence of neutral stratification at this site, driven by humidity and urban heat island effects that suppress daytime convection and nighttime stability (Emeis, 2018), creates a complex forecast environment during sunset transitions. LGBM demonstrates superior robustness during these periods, outperforming persistence at the three-step-ahead horizon and effectively capturing regime transitions. Conversely, SVR exhibits pronounced sensitivity and substantially larger errors under stable conditions (i.e., 1 to 7 hours). However, because energy prices are lower during these early hours, the financial impact of SVR inaccuracy is limited.

This interplay between model performance and varying energy prices motivates a dedicated assessment of the forecast impact on revenue, presented in the following section.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

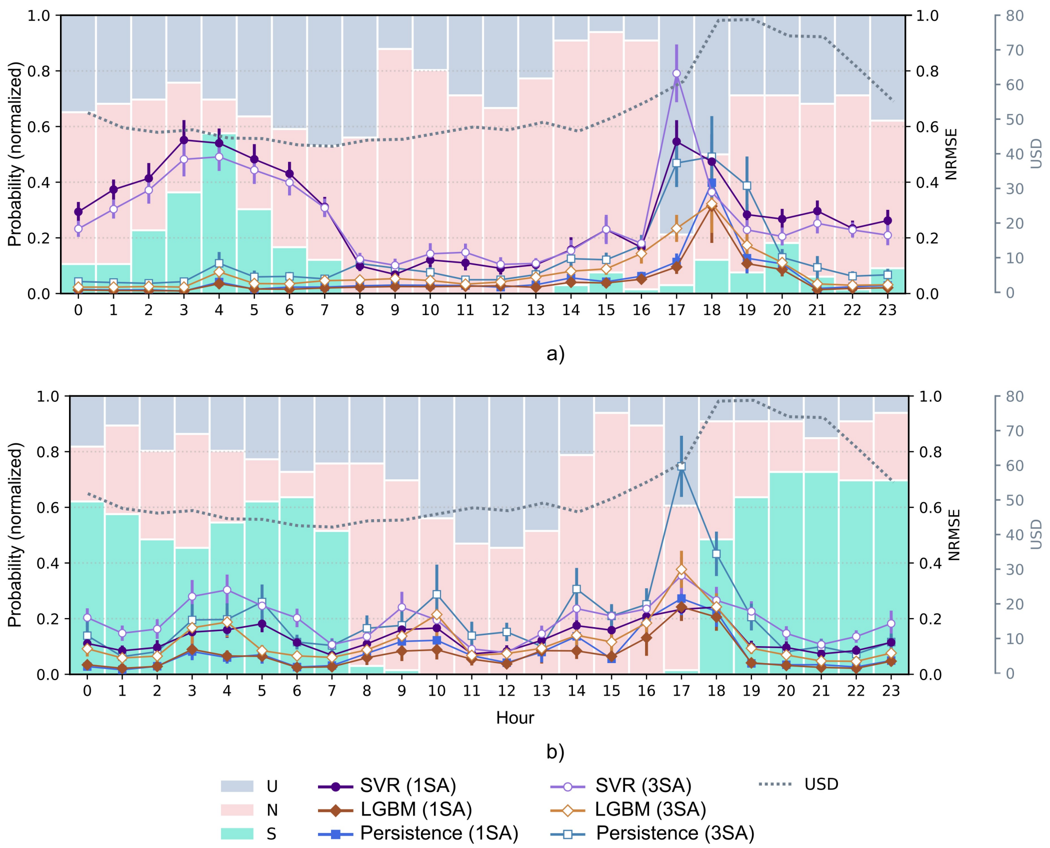

Figure 13. Hourly NRMSE of the ML models and the persistence model at 1SA and 3SA, for the best-performing configurations: $U+V$ in LGBM and $U+V+TI+TKE+W$ in SVR, at (a) 30 m and (b) 80 m. In the background, the hourly probability of occurrence of each stability class is shown. The gray dashed line indicates the average hourly energy price.

4.4 Impact of forecasting errors on energy revenues

This subsection analyzes the economic implications of uncertainties in WS forecasts, an aspect that is often overlooked in wind power assessment. To visualize the error distribution, Fig. 14 shows the frequency matrix of measured versus forecasted WS in speed bins, overlaid with the power curve used in this work. A perfect forecast places all points on the main diagonal; greater spread away from it indicates poorer model performance. For the one-step-ahead horizon, both Persistence and LGBM show high concentration along this diagonal, particularly within the $3-7$ ms^{-1} range. In contrast, SVR exhibits significant distortion, especially at the extremes.

For three-step-ahead horizon, this distortion is amplified for all models, with deviations of up to 3 ms^{-1} at low WS. Although these errors usually occur below cut-in speed, overestimating WS near cut-in can wrongly indicate power production when the turbine is idle, while underestimating WS at high speeds suggests rated operation despite a lower predicted output. Modest

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

Open Access
WIND ENERGY SCIENCE
DISCUSSIONS

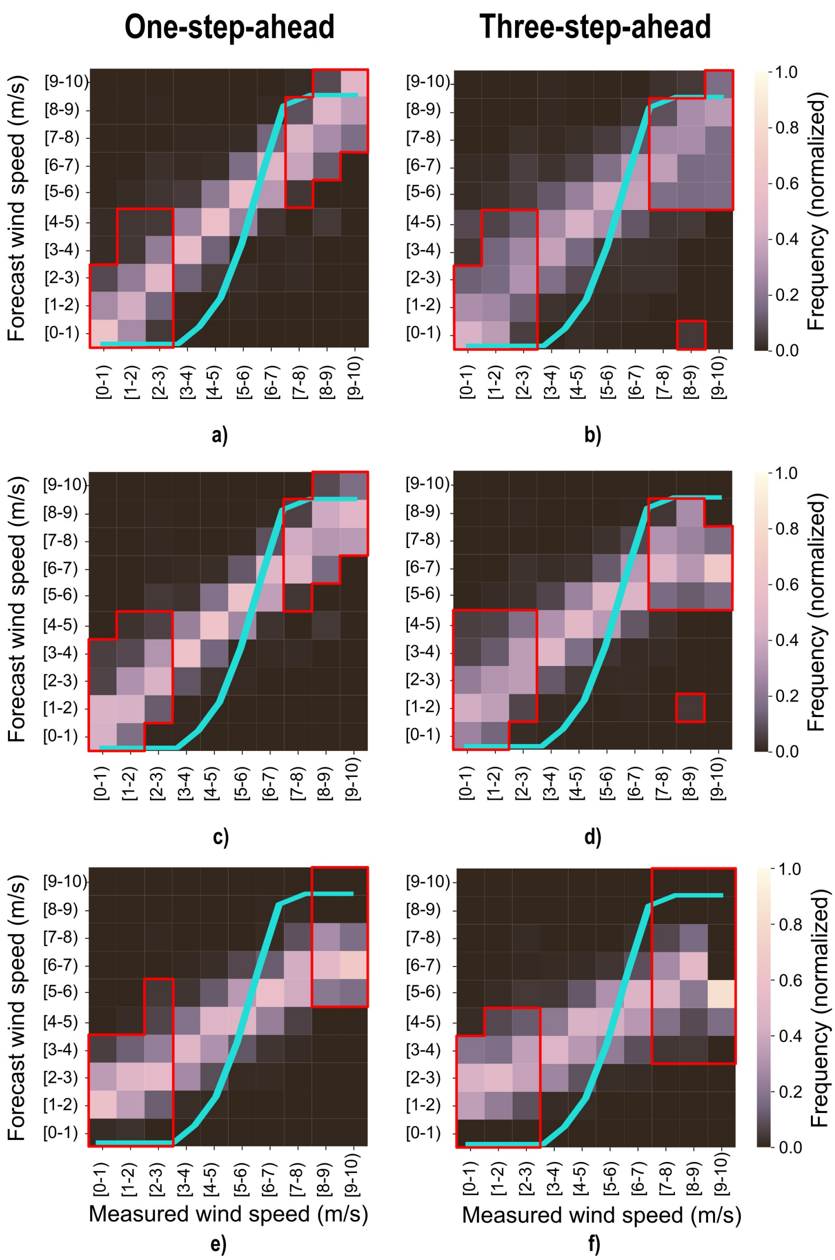

Figure 14. Relationship between the observed winds and forecast winds with one-step-ahead forecast (left column) and three-step-ahead forecast (right column), grouped by ranges. a) and b) correspond to persistence, c) and d) to LGBM, and e) and f) to SVR. The color bar shows the frequency of occurrence. The turquoise line indicates the power curve. The red boxes highlight the ranges where the forecast winds deviate the most from the observed winds.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

atmospheric characterization errors can cause large annual energy production differences; Pérez et al. (2023) found deviations of 8–16% depending on the stability regime.

Figure 15 shows the percentage differences in revenue compared to the baseline as a function of the lag size and atmospheric stability predictors for the different LGBM horizons. The baseline is the estimated energy production in the test period. The persistence model (not shown) yielded revenue differences of 0.16% for the one-step-ahead horizon and 0.31% for the three-step horizon.

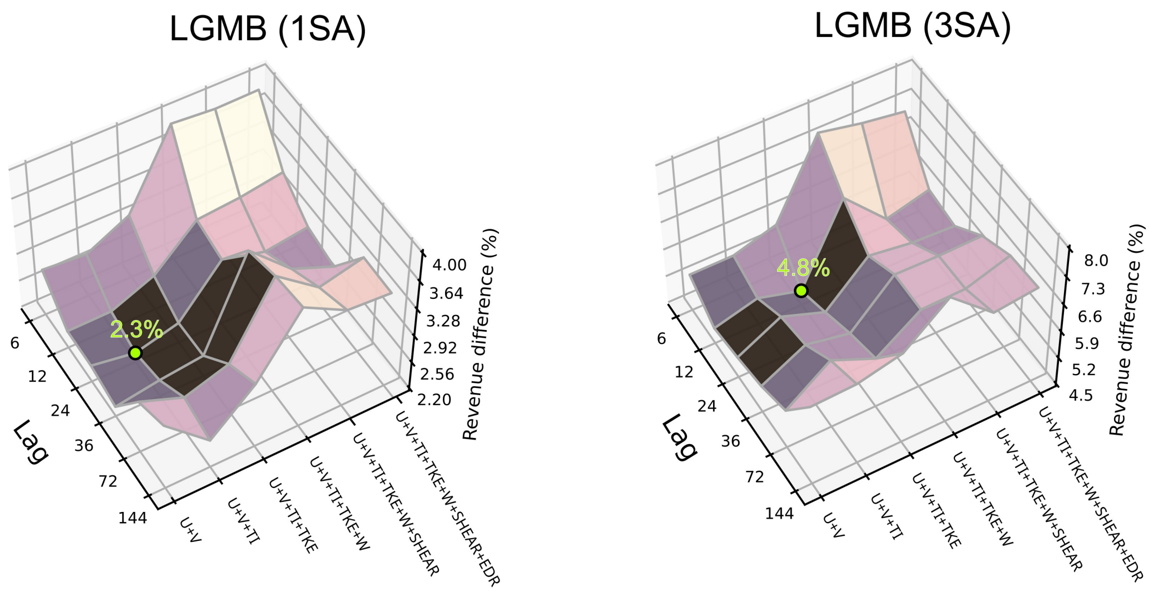

Figure 15. Revenue differences estimated by the LGBM model for both horizons relative to the base case as a function of the number of predictors and lag (1 lag = 10 min). The green dot indicates the configuration with the smallest revenue difference.

For one-step-ahead forecasts, LGBM exhibited the smallest revenue deviations from baseline when using lag 24 (4 h of observations) with the $U+V+TI$ set, yielding a difference of about 2.3%. For forecasts in three-steps, the deviations increase to 4.8% with $U+V+TI+TKE$ and 12 lags (2 h of observations). These configurations differ from the simplest model (six lags with $U+V$) by 10% and 17% for the one- and three-step horizons, respectively. These findings do not perfectly reflect the WS forecast behavior. LGBM produced its best WS forecasts with simpler features and short lags, while energy estimation performs best with intermediate lags $(24-36)$ and at least TI. However, despite reducing WS forecast error, LGBM's energy estimation errors can match or exceed persistence at the three-step horizon, particularly for WS > 9 ms^{-1} near the turbine's rated power, as shown in Fig. 16. This occurs because persistence concentrates its largest WS errors in ranges where the power curve is less sensitive and maintains a lower bias at high WS (Fig. 14a–b), limiting its impact on generation estimates. In contrast, LGBM tends to underestimate at high WS, amplifying energy deviations through the power curve conversion, while

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

SVM errors increase substantially, up to 15 times with respect to LGBM and persistence at one-step horizon for higher WS, and by factor $2-2.5$ at the three-step horizon.

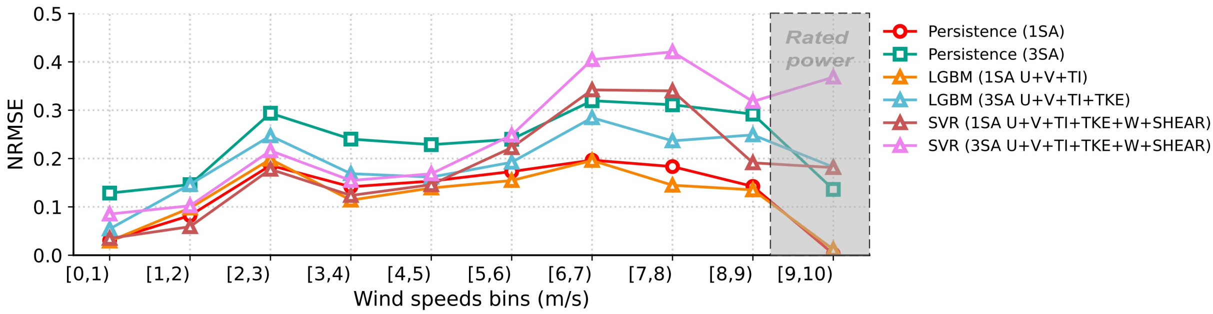

Figure 16. Comparison of the NRMSE errors of the estimated energy as a function of the WS bins. The corresponding best performing models were used for each horizon.

These results show that minimizing mean WS forecast error does not necessarily reduce energy or economic impacts, because small WS deviations can cause large differences in estimated energy and revenue. We examine this effect by comparing the model's estimated daily revenue to the base case, normalized by the base case's average daily revenue and disaggregated by atmospheric stability, as shown in Fig. 17. Unstable conditions prevail at low speeds ($\lesssim 6.5\ \mathrm{ms}^{-1}$), while neutral and stable conditions dominate the $3-10\ \mathrm{ms}^{-1}$ range. At the one-step horizon, LGBM tends to underestimate the energy under stable conditions ($5-8\ \mathrm{ms}^{-1}$). By the three-step horizon, errors increase by a factor of $2-3$, with LGBM maintaining a negative bias under stable conditions and asymmetry under neutral regimes. In contrast, persistence shows a more symmetric error distribution, but with larger absolute deviations exceeding $\pm 1.5\ \mathrm{ms}^{-1}$. These patterns align with operational wind farm data that suggest that stable regimes are associated with reduced energy production and increased variability (Kim and Kim, 2022).

Temporal analysis reveals that revenue discrepancies are not uniform throughout the day. Mean hourly errors peak between 17:00 and 19:00, coinciding with the transition from neutral to stable regimes and peak energy price levels. This overlap suggests that the largest financial risks occur during sunset transitions. Additionally, forecast performance degraded significantly during synoptic events of high variability, such as Hurricane "Otis" and a "Norte" event (Figs. S6 and S7 in SM). These results indicate that revenue differences depend on the distribution of biases across WS ranges and stability regimes, where errors at medium-high WS are amplified by both the power curve and time-varying energy prices. Finally, Fig. 18 summarizes the NRMSE assessment for WS, energy, and revenue. The results show that economic deviations do not maintain a linear dependence on WS error. At the one-step horizon, LGBM and persistence perform comparably, with LGBM showing marginal gains in specific configurations. However, at the 3SA horizon, LGBM's superiority in WS and energy forecasting does not translate into proportional revenue improvements, as similar results were observed for both models. SVR consistently exhibits the highest errors and sensitivity to predictor configurations across all horizons, performing similarly to persistence only in WS forecasting at three-step-ahead horizon, but failing to match its economic reliability.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

WIND ENERGY SCIENCE DISCUSSIONS

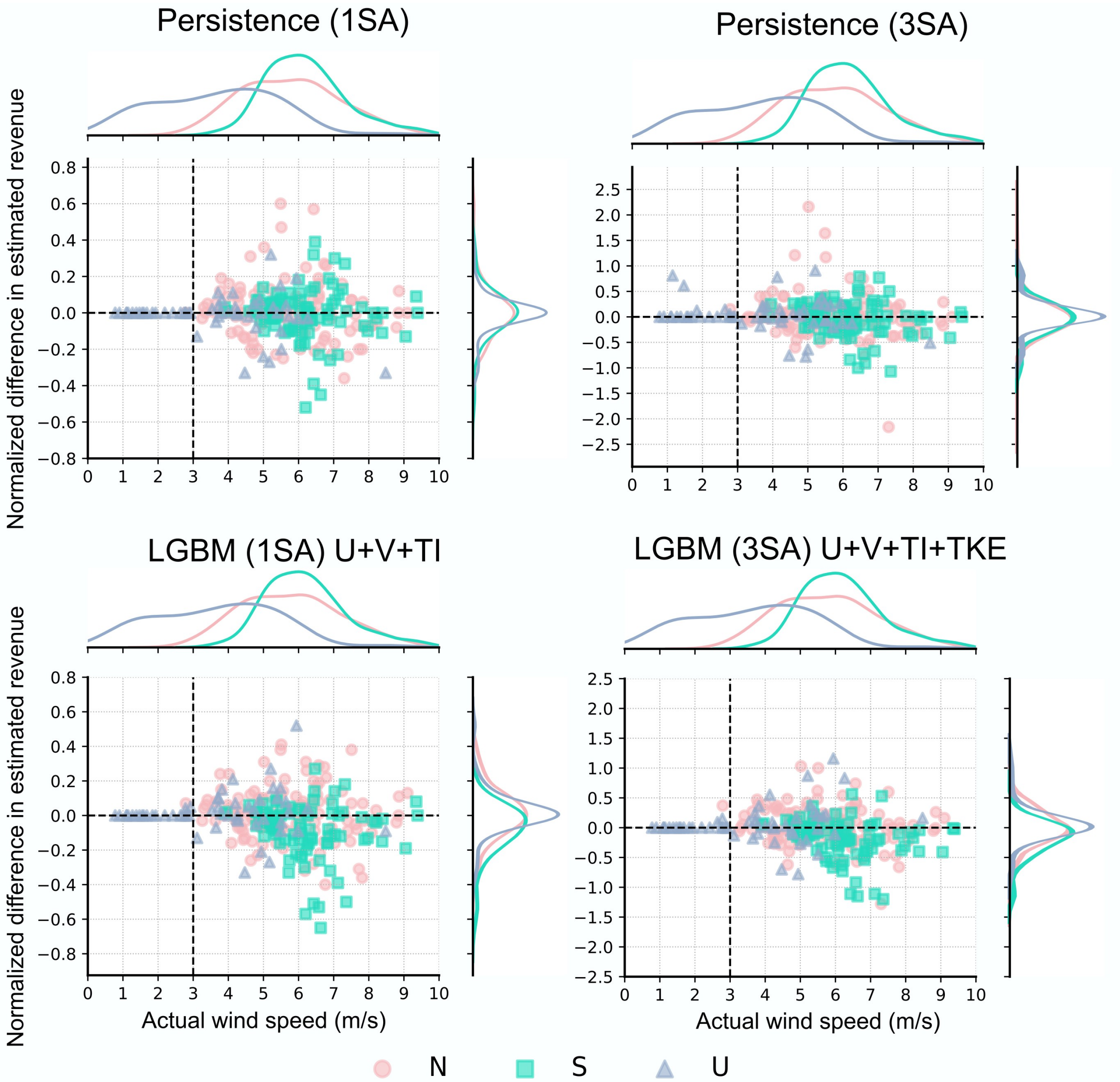

Figure 17. Scatter plot of the relationship between the observed WS and the bias of the estimated daily revenue. The vertical dashed line indicates the cut-in speed of the wind turbine.

5 Conclusions

This exploratory study investigated how atmospheric stability influences the propagation of ultra-short-term WS forecast errors along the wind–power–revenue chain at a tropical site on the Yucatan Peninsula. The results indicate that atmospheric stability

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

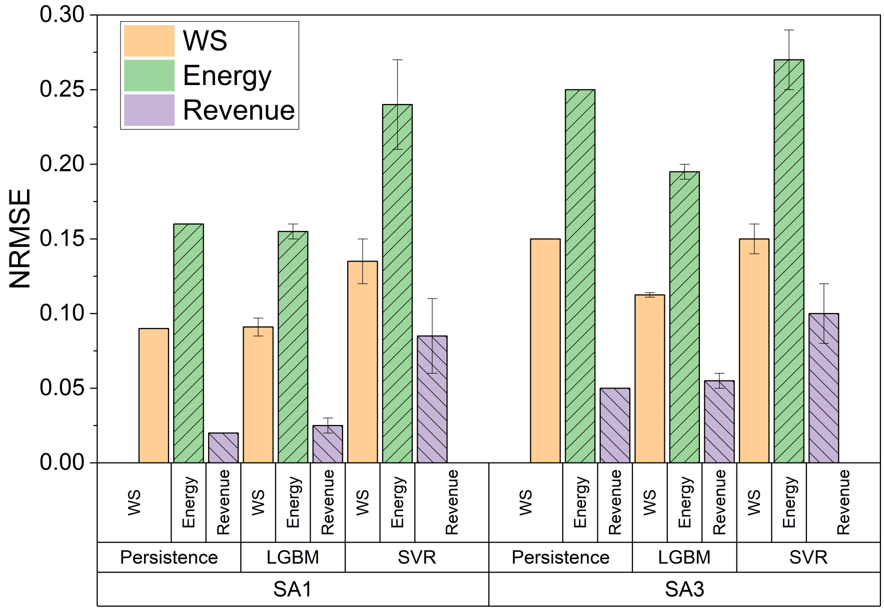

Figure 18. Normalized RMSE for WS, energy, and revenue forecasts at one-step-ahead (1SA) and three-step-ahead (3SA) horizons for the three evaluated models. WS and energy were estimated every 10 minutes, and revenue hourly using available energy prices. For LGMB and SVR, the reported range denotes the minimum and maximum NRMSE obtained across models trained with different numbers of predictors.

influences not only the accuracy of forecasts, but also the extent to which WS prediction errors propagate into deviations in energy production and associated revenue.

LGBM consistently provided the most accurate and robust WS forecasts, exhibiting minimal sensitivity to lag length and predictor set. In contrast, SVR was more strongly affected by atmospheric conditions and benefited markedly from including variables related to turbulence (TI, TKE, W), particularly under stable and neutral conditions. These findings indicate that the value of atmospheric predictors depends on the prevailing boundary-layer structure and that turbulence descriptors become increasingly important when vertical mixing is suppressed.

A key finding was that improvements in the accuracy of the WS forecast do not necessarily translate into proportional economic benefits. Due to the nonlinear response of the turbine power curve, small biases in wind speed produced variations in daily revenues of around 2.3% to 4.8%, especially during stability changes, electricity price peaks and synoptic situations with strong atmospheric variability. Consequently, evaluating WS forecast models using only conventional error metrics can give a partial picture of their operational value.

Although the findings of this work are relevant, they also have several limitations that should be considered when interpreting the estimated power and revenue errors from the WS forecasts, as in the implementation itself. The analysis is restricted to a single tropical site with simple orography, a specific SoDAR configuration, a generic power curve, and a simplified market model. Consequently, the exact magnitude of the reported power and revenue deviations may differ in complex terrain, offshore

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

settings, or other climatic regimes. In addition, economic analysis omits wake effects, storage, imbalance costs, and alternative market structures. Within these controlled assumptions, our results provide a lower bound on the economic impact of short-term forecast errors and a transferable framework that can be extended to climate-change assessments, hybrid renewable systems, and more realistic market conditions. These represent challenges to be addressed in future work.

Supplement. Supplementary material for this article can be found in the attached document.

Data availability. The complete dataset used in this study is subject to institutional data-sharing restrictions and therefore cannot be made publicly available through an open repository. Researchers interested in accessing the data may request it from the corresponding author.

Author contributions. LCC.: conceptualization, formal analysis, investigation, methodology, data curation, software, visualization, writing – original draft, writing – review and editing. MR, MAES: conceptualization, writing - review & editing, formal analysis, visualization, methodology, supervision. LS, AB: formal analysis, writing - review & editing.

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

Acknowledgements. This work has been supported by Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI), Mexico, through the PhD research grant entitled to Linda Canché-Cab, with CVU number 926185. During the preparation of this paper, AI tools (ChatGPT, OpenAI) were used for grammatical correction and to improve readability; all authors developed, reviewed, and approved all technical content.

https://doi.org/10.5194/wes-2026-122
Preprint. Discussion started: 21 August 2026

References

Allende-Arandía, M. E., Zavala-Hidalgo, J., Torres-Freyermuth, A., Appendini, C. M., Cerezo-Mota, R., and Taylor-Espinosa, N.: Sea-land breeze diurnal component and its interaction with a cold front on the coast of Sisal, Yucatan: A case study, Atmospheric Research, 244, 105 051, https://doi.org/10.1016/j.atmosres.2020.105051, 2020.

Amat Rodrigo, J. and Escobar Ortiz, J.: skforecast, https://doi.org/10.5281/zenodo.8382788, 2025.

Arslan Tuncar, E., Şafak Sağlam, and Oral, B.: A review of short-term wind power generation forecasting methods in recent technological trends, Energy Reports, 12, 197–209, https://doi.org/10.1016/j.egyr.2024.06.006, 2024.

Aslam, S., Aung, P. P., Rafsanjani, A. S., and Majeed, A. P. P. A.: Machine learning applications in energy systems: current trends, challenges, and research directions, Energy Informatics, 8, 62, https://doi.org/10.1186/s42162-025-00524-6, 2025.

Barthelmie, R. J., Churchfield, M. J., Moriarty, P. J., Lundquist, J. K., Oxley, G. S., Hahn, S., and Pryor, S. C.: The role of atmospheric stability/turbulence on wakes at the Egmond aan Zee offshore wind farm, Journal of Physics: Conference Series, 625, 012 002, https://doi.org/10.1088/1742-6596/625/1/012002, 2015.

Boujoudar, M., Ydrissi, M. E., Abraim, M., Bouarfa, I., Alani, O. E., Ghennioui, H., and Bennouna, E. G.: Comparing machine learning algorithms for imputation of missing time series in meteorological data, Neural Computing and Applications, 37, 23 773–23 787, https://doi.org/10.1007/s00521-024-10601-8, 2025.

Bradley, S.: Sound in the Atmosphere, chap. 3, pp. 27–53, CRC Press, https://doi.org/10.1201/9781420005288, 2007.

Canché-Cab, L., Rivero, M., and Escalante, M.: Revenue discrepancies in wind power: How data imputation methods influence yield estimates, manuscript under review, 2026.

Carrera, B. and Kim, K.: Comparative Analysis of Machine Learning Techniques in Predicting Wind Power Generation: A Case Study of 2018–2021 Data from Guatemala, Energies, 17, 3158, https://doi.org/10.3390/en17133158, 2024.

CENACE: Precios de Energía y Servicios Conexos MDA, https://www.cenace.gob.mx/Paginas/SIM/Reportes/PreEnerServConMDA.aspx, accessed: 2025-12-28, 2025.

Chen, H., Wan, Q., and Wang, Y.: Refined Diebold-Mariano Test Methods for the Evaluation of Wind Power Forecasting Models, Energies, 7, 4185–4198, https://doi.org/10.3390/en7074185, 2014.

Coordinación General del Servicio Meteorológico Nacional, Gerencia de Meteorología y Climatología, and Subgerencia de Climatología y Servicios Climáticos: REPORTE ANUAL DEL CLIMA EN MÉXICO 2023, Tech. rep., https://smn.conagua.gob.mx/tools/DATA/Climatolog%C3%ADa/Diagn%C3%B3stico%20Atmosf%C3%A9rico/Reporte%20del%20Clima%20en%20M%C3%A9xico/Anual2023.pdf, 2024.

Dörterler, S., Arslan, S., and Özdemir, D.: Unlocking the potential: A review of artificial intelligence applications in wind energy, Expert Systems, 41, e13 716, https://doi.org/10.1111/exsy.13716, 2024.

Emeis, S.: Vertical Profiles Over Flat Terrain, pp. 31–89, Springer International Publishing, Cham, ISBN 978-3-319-72859-9, https://doi.org/10.1007/978-3-319-72859-9_3, 2018.

Escalante Soberanis, M. and Mérida, W.: Regarding the influence of the Van der Hoven spectrum on wind energy applications in the meteorological mesoscale and microscale, Renewable Energy, 81, 286–292, https://doi.org/https://doi.org/10.1016/j.renene.2015.03.048, 2015.

Feurer, M. and Hutter, F.: Hyperparameter Optimization, pp. 3–33, Springer International Publishing, Cham, ISBN 978-3-030-05318-5, https://doi.org/10.1007/978-3-030-05318-5_1, 2019.

Galarza-Chavez, A. A., Martinez-Rodriguez, J. L., Domínguez-Cruz, R. F., López-Garza, E., and Rios-Alvarado, A. B.: Multi-step wind energy forecasting in the Mexican Isthmus using machine and deep learning, Energy Reports, 13, 1–15, https://doi.org/10.1016/j.egyr.2024.11.074, 2025.

Géron, A.: Hands-on machine learning with Scikit-Learn, Keras and TensorFlow : concepts, tools, and techniques to build intelligent systems, O'Reilly Media, Inc., 3 edn., ISBN 9781098125974, 2023.

Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, J. Mach. Learn. Res., 3, 1157–1182, 2003.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, Springer New York, 2 edn., ISBN 978-0-387-84857-0, https://doi.org/10.1007/978-0-387-84858-7, 2009.

Hearst, M., Dumais, S., Osuna, E., Platt, J., and Scholkopf, B.: Support vector machines, IEEE Intelligent Systems and their Applications, 13, 18–28, https://doi.org/10.1109/5254.708428, 1998.

Ho, C.-Y., Cheng, K.-S., and Ang, C.-H.: Utilizing the Random Forest Method for Short-Term Wind Speed Forecasting in the Coastal Area of Central Taiwan, Energies, 16, https://doi.org/10.3390/en16031374, 2023.

Jiang, W., Liu, B., Liang, Y., Gao, H., Lin, P., Zhang, D., and Hu, G.: Applicability analysis of transformer to wind speed forecasting by a novel deep learning framework with multiple atmospheric variables, Applied Energy, 353, 122 155, https://doi.org/10.1016/j.apenergy.2023.122155, 2024.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: LightGBM: a highly efficient gradient boosting decision tree, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, p. 3149–3157, Curran Associates Inc., Red Hook, NY, USA, ISBN 9781510860964, 2017.

Kim, D.-Y. and Kim, B.-S.: Differences in wind farm energy production based on the atmospheric stability dissipation rate: Case study of a 30 MW onshore wind farm, Energy, 239, 122 380, https://doi.org/10.1016/j.energy.2021.122380, 2022.

Kim, D.-Y. and Kim, B.-S.: Contribution of meteorological factors based on explainable artificial intelligence in predicting wind farm power production using machine learning algorithms, Journal of Renewable and Sustainable Energy, 15, https://doi.org/10.1063/5.0127519, 2023.

Kumar, A., Singh, A., and Kumar, S.: Forecasting wind speed over multiple horizons: Superiority of deep learning and decomposition-driven hybrid models, Unconventional Resources, 9, 100 296, https://doi.org/10.1016/j.uncres.2025.100296, 2026.

Liu, L. and Stevens, R. J.: Effects of atmospheric stability on the performance of a wind turbine located behind a three-dimensional hill, Renewable Energy, 175, 926–935, https://doi.org/https://doi.org/10.1016/j.renene.2021.05.035, 2021.

Manohar, N., Chaurasia, M. A., Mozar, S., and Juang, C.-F.: Artificial Intelligence and IoT in Renewable Energy, chap. 5, pp. 135–145, Springer Nature Singapore, Singapore, ISBN 978-981-97-9939-8, https://doi.org/10.1007/978-981-97-9939-8_5, 2025.

Mantuano, C., Omoyele, O., Hoffmann, M., Weinand, J. M., Panella, M., and Stolten, D.: Data imputation methods for intermittent renewable energy sources: Implications for energy system modeling, Energy Conversion and Management, 339, 119 857, https://doi.org/10.1016/j.enconman.2025.119857, 2025.

Menken, J. and Wildmann, N.: Impact of atmospheric stability and turbulence on wind turbine wake characteristics: a nacelle lidar study, Wind Energy Science Discussions, 2026, 1–27, https://doi.org/10.5194/wes-2026-8, 2026.

Nielson, J., Bhaganagar, K., Meka, R., and Alaeddini, A.: Using atmospheric inputs for Artificial Neural Networks to improve wind turbine power prediction, Energy, 190, 116 273, https://doi.org/10.1016/j.energy.2019.116273, 2020.

Nybø, A., Nielsen, F. G., Reuder, J., Churchfield, M. J., and Godvik, M.: Evaluation of different wind fields for the investigation of the dynamic response of offshore wind turbines, Wind Energy, 23, 1810–1830, https://doi.org/10.1002/we.2518, 2020.

Omidkar, A., Es'haghian, R., and Song, H.: Using machine learning methods for long-term technical and economic evaluation of wind power plants, Green Energy and Resources, 3, 100 115, https://doi.org/10.1016/j.gerr.2025.100115, 2025.

Optis, M. and Perr-Sauer, J.: The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production, Renewable and Sustainable Energy Reviews, 112, 27–41, https://doi.org/10.1016/j.rser.2019.05.031, 2019.

Optis, M., Bodini, N., Debnath, M., and Doubrawa, P.: New methods to improve the vertical extrapolation of near-surface offshore wind speeds, Wind Energy Science, 6, 935–948, https://doi.org/10.5194/wes-6-935-2021, 2021.

Pérez, C., Rivero, M., Escalante, M., Ramirez, V., and Guilbert, D.: Influence of Atmospheric Stability on Wind Turbine Energy Production: A Case Study of the Coastal Region of Yucatan, Energies, 16, https://doi.org/10.3390/en16104134, 2023.

Pérez Albornoz, C., Escalante Soberanis, M., Ramírez Rivera, V., and Rivero, M.: Review of atmospheric stability estimations for wind power applications, Renewable and Sustainable Energy Reviews, 163, 112 505, https://doi.org/10.1016/j.rser.2022.112505, 2022.

Rivero, M., Orozco, S., Sellschopp, F., and Loera-Palomo, R.: A new methodology to extend the validity of the Hargreaves-Samani model to estimate global solar radiation in different climates: Case study Mexico, Renewable Energy, 114, 1340–1352, https://doi.org/10.1016/j.renene.2017.08.003, 2017.

Rivero, M., Reyes, A., Escalante, M., and Probst, O.: Forecasting of renewable energy generation for grid integration, chap. Chapter 9, pp. 219–257, https://doi.org/10.1049/PBPO159E_ch9, 2020.

Sathe, A., Mann, J., Barlas, T., Bierbooms, W., and van Bussel, G.: Influence of atmospheric stability on wind turbine loads, Wind Energy, 16, 1013–1032, https://doi.org/10.1002/we.1528, 2013.

ScintecAG: Scintec Flat Array Sodars Theory Manual SFAS, MFAS, XFAS including RASS RAE2 and windRASS, Scintec, Rottenburg, Germany, version 1.03 edn., www.scintec.com, 2017.

Shoaei, M., Noorollahi, Y., Hajinezhad, A., and Moosavian, S. F.: A review of the applications of artificial intelligence in renewable energy systems: An approach-based study, Energy Conversion and Management, 306, 118 207, https://doi.org/10.1016/j.enconman.2024.118207, 2024.

Singh, U. and Rizwan, M.: A Systematic Review on Selected Applications and Approaches of Wind Energy Forecasting and Integration, Journal of The Institution of Engineers (India): Series B, 102, 1061–1078, https://doi.org/10.1007/s40031-021-00618-1, 2021.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Springer Netherlands, ISBN 978-90-277-2769-5, https://doi.org/10.1007/978-94-009-3027-8, 1988.

U.S. Environmental Protection Agency: Meteorological monitoring guidance for regulatory modeling applications, Tech. Rep. EPA-454/R-99-005, Office of Air Quality Planning and Standards, Research Triangle Park, NC. 2000, https://www.epa.gov/sites/default/files/2020-10/documents/mmgrma_0.pdf, 2000.

Vassallo, D., Krishnamurthy, R., Sherman, T., and Fernando, H. J. S.: Analysis of Random Forest Modeling Strategies for Multi-Step Wind Speed Forecasting, Energies, 13, https://doi.org/10.3390/en13205488, 2020.

Vassallo, D., Krishnamurthy, R., and Fernando, H. J. S.: Utilizing physics-based input features within a machine learning model to predict wind speed forecasting error, Wind Energy Science, 6, 295–309, https://doi.org/10.5194/wes-6-295-2021, 2021.

Veers, P., Bottasso, C. L., Manuel, L., Naughton, J., Pao, L., Paquette, J., Robertson, A., Robinson, M., Ananthan, S., Barlas, T., Bianchini, A., Bredmose, H., Horcas, S. G., Keller, J., Madsen, H. A., Manwell, J., Moriarty, P., Nolet, S., and Rinker, J.: Grand challenges in the design, manufacture, and operation of future wind turbine systems, Wind Energy Science, 8, 1071–1131, https://doi.org/10.5194/wes-8-1071-2023, 2023.

Wan, S., Chen, L., Jin, S., Yin, W., Cui, Y., and Zhang, F.: Estimates of wind speed profiles from surface observations: Machine learning versus Monin–obukhov approach, Boundary Layer Meteorol., 191, https://doi.org/10.1007/s10546-025-00903-2, 2025.

Wharton, S. and Lundquist, J. K.: Atmospheric stability affects wind turbine power collection, Environmental Research Letters, 7, 014 005, https://doi.org/10.1088/1748-9326/7/1/014005, 2012.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Elsevier, ISBN 9780128158234, https://doi.org/10.1016/C2017-0-03921-6, 2020.

Zabihi, A., Manur, V. B., Dintakurthy, Y., Prasad, K., Rene, P., and Krishna, V.: Machine learning-based wind speed forecasting: a comparative study, Scientific Reports, 15, 40 866, https://doi.org/10.1038/s41598-025-24640-1, 2025.

Zhu, X.-R., Tian, J.-H., and Wang, Y.: The impact of climate change on wind and solar photovoltaic power potential and distribution in China, Advances in Climate Change Research, 16, 730–746, https://doi.org/https://doi.org/10.1016/j.accre.2025.04.015, special Issue on Technological Innovation for Adaptation to Climate Change, 2025.