frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 07 January 2026
DOI 10.3389/fenrg.2025.1686125

OPEN ACCESS

EDITED BY
Matilde Santos,
Complutense University of Madrid, Spain

REVIEWED BY
Molaka Maruthi,
Kyungpook National University,
Republic of Korea
Nandha Kishore S. R,
Indian Institute of Information Technology, Sri City, India

*CORRESPONDENCE
Murat Isik,
muratisik@ahievran.edu.tr

RECEIVED 14 August 2025
REVISED 28 October 2025
ACCEPTED 11 December 2025
PUBLISHED 07 January 2026

CITATION
Isik M and Yalcinkaya M (2026) Integrated regime-aware wind power forecasting using multi-altitude meteorological features and hybrid machine learning.
Front. Energy Res. 13:1686125.
doi: 10.3389/fenrg.2025.1686125

Integrated regime-aware wind power forecasting using multi-altitude meteorological features and hybrid machine learning

Murat Isik* and Mehmet ali Yalcinkaya

Department of Computer Engineering, Faculty of Engineering and Architecture, Kirsehir Ahi Evran University, Kirsehir, Türkiye

Accurate wind power forecasting is critical for grid stability and renewable energy integration, yet the inherent variability of atmospheric conditions presents significant challenges. This study proposes a unified and scalable pipeline that integrates wind regime detection, temporal sequence modeling, and regime-conditioned deterministic and probabilistic power forecasting. Using 8 years of high-resolution meteorological data from multiple altitudes, we engineer a comprehensive set of physically interpretable features, including wind shear, temperature gradients, and rolling statistics. Regimes are identified via KMeans and Gaussian Mixture Models, with Principal Component Analysis applied post-clustering for visualization and interpretation. Temporal regime dynamics are characterized through both empirical and Markov transition matrices and modeled using Long Short-Term Memory (LSTM) networks for regime sequence prediction. For power forecasting, regime-specific models are developed using tuned ensemble regressors (XGBoost, LightGBM, CatBoost, and Random Forest), complemented by probabilistic approaches including Quantile Regression Forests, quantile-based XGBoost, and Bayesian neural networks. Results show that regime conditioning significantly enhances forecasting performance, with the stacked meta-learning ensemble achieving R^2 = 0.997 and over 30% reduction in MAE compared to baseline methods. Probabilistic models produce well-calibrated prediction intervals, providing uncertainty-aware forecasts suitable for operational decision-making. This work contributes a novel end-to-end framework that jointly models regime persistence, transitions, and regime-conditioned power output, incorporating uncertainty quantification through Quantile Regression Forests, quantile-based XGBoost, and Bayesian Neural Networks, thereby bridging a gap in the literature where these components are often treated in isolation. The approach advances both accuracy and interpretability, offering practical value for wind farm operation and renewable energy integration.

KEYWORDS

wind regime detection, probabilistic power forecasting, hybrid machine learning, Markov chain modeling, ensemble learning

1 Introduction

Wind energy has emerged as a cornerstone of the global transition to low-carbon power systems. However, the stochastic and highly variable nature of wind poses significant challenges for accurate forecasting, which is essential for grid stability, economic dispatch, and large-scale integration of renewable energy sources (Phan et al., 2021). Traditional statistical models often fail to capture the nonlinear and non-stationary dynamics of atmospheric processes, especially when dealing with multi-height meteorological inputs and rapidly changing boundary-layer conditions (Finamore et al., 2023).

Recent advances in machine learning and deep learning have shown promise in improving wind power forecasting accuracy by leveraging high-dimensional feature spaces and temporal dependencies. Hybrid approaches that combine clustering, ensemble methods, and deep neural networks have been particularly effective in capturing complex weather–power relationships (Miao et al., 2021; Gao et al., 2024). At the same time, the growing adoption of probabilistic forecasting reflects the need for uncertainty-aware predictions to support risk-based operational strategies in power systems (Fan et al., 2021).

Another critical aspect of wind forecasting is recognizing that atmospheric states are not continuous but occur in discrete "regimes" governed by boundary-layer stability, vertical shear, and synoptic-scale drivers. Clustering-based regime identification has been shown to enhance both interpretability and predictive performance by conditioning models on physically meaningful atmospheric states (Kazor, 2013; Yaqiong et al., 2016). Integrating regime detection with advanced deep learning models such as LSTM and attention-based architectures further strengthens sequence prediction and regime-aware forecasting (Wadood et al., 2024).

This study introduces a unified and scalable pipeline for wind regime detection, temporal modeling, and regime-conditioned forecasting that systematically integrates these processes. By coupling regime identification with transition modeling (via Markov Chains and LSTM) and incorporating probabilistic ensemble learning for power forecasting, the framework advances both accuracy and interpretability. The contribution lies in integrating physically interpretable feature engineering, regime dynamics, and uncertainty quantification into a single end-to-end pipeline for operational wind forecasting.

First, we leverage a rich set of engineered meteorological features—such as wind shear, temperature gradients, and rolling statistics—across multiple altitudes. Second, we identify regimes using KMeans and Gaussian Mixture Models (GMM), with K (for KMeans) selected via silhouette score; no dimensionality reduction is applied prior to clustering. Subsequently, we apply PCA post-clustering strictly for two-dimensional visualization and interpretation (and optionally explore UMAP as a nonlinear alternative). Third, we incorporate temporal regime dynamics through Markov Chain models and Long Short-Term Memory (LSTM) networks to enable regime-sequence prediction. Finally, we develop regime-conditioned power forecasting models using extensively tuned ensemble regressors (XGBoost, LightGBM, CatBoost, and Random Forest) complemented by probabilistic modeling and uncertainty estimation. This comprehensive pipeline improves forecasting accuracy and enhances interpretability by revealing the persistence and transition behaviors of wind regimes, providing an end-to-end framework that jointly integrates regime persistence modeling, transition dynamics, and regime-conditioned forecasting with uncertainty quantification.

2 Related works

Wind regime detection and power forecasting have seen significant research interest, particularly through clustering, sequence modeling, and deep learning. However, most prior work addresses these components independently.

Regime Identification via Clustering; Statistical and machine learning-based clustering methods are frequently used for identifying wind regimes. For instance, (Kazor, 2013), evaluated GMMs and k-means for regime classification using meteorological variables, favoring GMMs for their accuracy (Kazor, 2013; Jia et al., 2022) employed k-means and hidden Markov models on reanalysis data across 23 sites, enhancing short-term wind speed forecasting (Jia et al., 2022; Kosanoglu, 2022) used Dirichlet mixture models and dynamic time warping for time-series clustering before applying deep learning (Kosanoglu, 2022).

Regime-Specific Forecasting and Hybrid Models; (Browell and Gilbert, 2017); demonstrated regime-switching autoregressive models for competitive short-term forecasting (Browell and Gilbert, 2017; Aly, 2020) presented hybrid deep learning models integrating Fourier transforms, neural nets, and clustering to boost performance (Aly, 2020).

Temporal and Sequence Modeling; Temporal modeling is essential for wind regime transitions (Monjoly and Blonbou, 2012). applied fuzzy c-means and Markov chains to forecast regime evolution (Monjoly and Blonbou, 2012; Ru et al., 2024) used deep embedded clustering combined with LSTM to handle regime-segmented ultra-short-term forecasting (Ru et al., 2024).

Feature Engineering and Deep Learning Advances; Recent studies emphasize rich feature sets and hybrid modeling (Wang et al., 2022). used fuzzy clustering and correlation analysis to improve accuracy with a deep sparse autoencoder (Wang et al., 2022; Lv and Wang, 2023) applied K-medoids and hybrid ConvLSTM models on multi-feature meteorological inputs (Lv and Wang, 2023). Similarly, (Xu et al., 2022), incorporated GCN with optimized feature selection for cluster-level forecasting.

Noise Robustness and Interpretability; (Christian et al., 2024); addressed robustness by using shape-wise feature engineering in CNN-LSTM networks (Christian et al., 2024), while (Zhang et al., 2020) improved forecasting accuracy using Seq2Seq with attention models on clustered data (Zhang et al., 2020).

Adaptive and Attention-Based Architectures; Recent architectures focus on adaptivity and attention (Zhao et al., 2024). proposed an adaptive online learning model combining BiLSTM, self-attention, and outlier detection for real-time forecasts (Zhao et al., 2024; Xing and He, 2023) utilized stacking deep models (BiGRU and echo networks) for multi-step forecasting with multimodal data fusion (Xing and He, 2023).

While past studies have achieved impressive results through clustering, deep learning, and hybrid modeling, they often limit focus to a specific stage of the pipeline—either clustering, temporal modeling, or forecasting. In contrast, our approach integrates

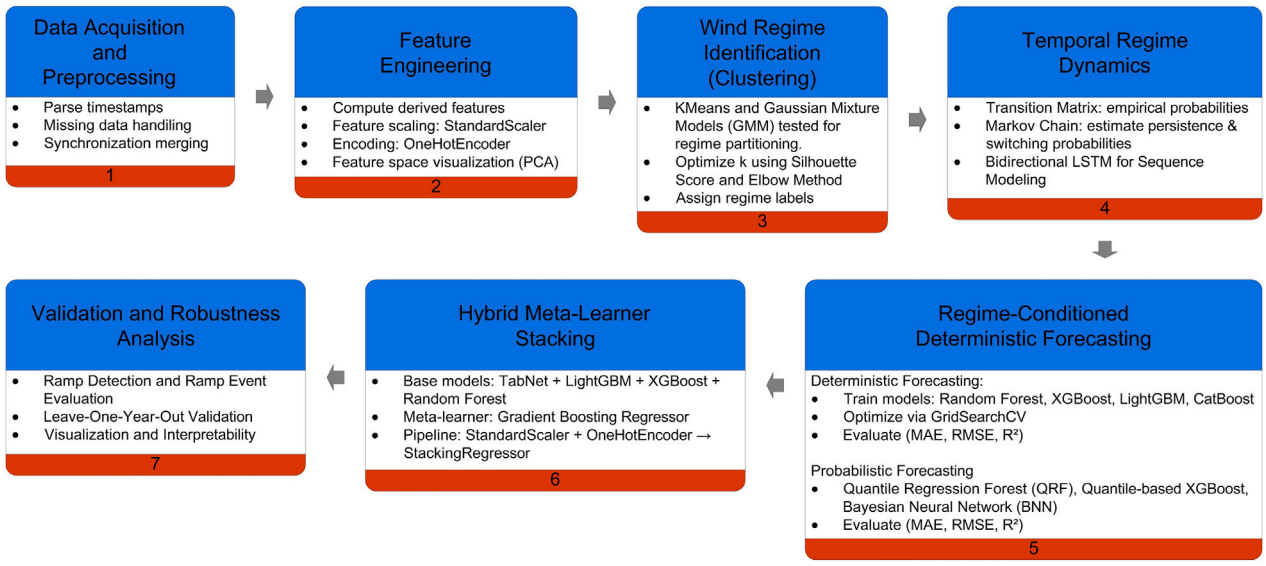

FIGURE 1
Workflow of the proposed regime-aware hybrid forecasting.

feature-rich regime identification, Markov-based and LSTM-based temporal sequence modeling, and regime-conditioned forecasting using both ensemble ML and deep learning. Moreover, we employ PCA primarily for visualization-based dimensionality reduction and adopt a per-regime model training strategy, which significantly enhances interpretability and forecasting performance. To our knowledge, this level of integration across all phases of wind power modeling remains largely unexplored in existing literature.

3 Methodology

This study introduces a unified hybrid pipeline for regime-aware wind power forecasting, structured into interconnected phases (Figure 1). The framework begins with comprehensive data preprocessing and multi-altitude feature engineering to capture vertical atmospheric variability. Next, unsupervised clustering (KMeans and GMM) is employed to identify meteorological regimes, followed by temporal regime-sequence modeling using LSTM and Markov transition analysis. In the final stage, regime-conditioned forecasting models—both deterministic and probabilistic—are trained, including ensemble regressors (Random Forest, XGBoost, LightGBM, CatBoost), Bayesian Neural Networks, and hybrid stacking meta-learners. The workflow concludes with model evaluation through leave-one-year-out validation and ramp event analysis to ensure robustness and operational relevance.

3.1 Data acquisition and preprocessing

The dataset used in this study was sourced from the Wind Resource Database (WRDB), developed and maintained by the National Renewable Energy Laboratory (NREL). Specifically, the data are derived from the WIND Toolkit Longterm Ensemble Dataset (WTK-LED), which provides high-resolution time-series meteorological data at 60-min intervals and 2 km spatial resolution across the United States. This dataset includes multiple vertical levels, allowing accurate representation of wind profiles across the atmospheric boundary layer. The WTK-LED is generated using the Weather Research and Forecasting (WRF) model initialized with ERA5 reanalysis data and includes uncertainty quantification for wind speeds across all grid points (National Renewable Energy, 2023).

The foundational dataset comprises 8 years of hourly meteorological observations, recorded at multiple altitudes ranging from 10 m to 200 m. Initial preprocessing steps included.

- Timestamp parsing using Year, Month, Day, Hour, and Minute columns.
- Handling missing values via forward-fill and back-fill imputation.
- Merging and synchronizing meteorological inputs from multiple heights into a single indexed time series.

This cleaned and time-aligned dataset served as the foundation for feature engineering and subsequent analysis.

3.2 Feature engineering

To capture the vertical and temporal variability of the atmosphere and to enhance the physical interpretability of the model, a set of engineered features was derived from the original meteorological variables. These features were selected based on their established influence on wind turbine inflow and power generation rather than through purely data-driven selection. Wind shear and wind veer quantify vertical changes in wind speed and direction,

respectively, reflecting boundary-layer stability and turbulence. Temperature gradients represent thermal stratification and mixing potential, while humidity deficit (the complement of surface relative humidity) captures air dryness and its impact on density and energy extraction efficiency. Pressure change indicates synoptic-scale weather transitions, and rolling statistics (24-h moving means and standard deviations) capture diurnal persistence and mesoscale variability. Estimated power, derived via the cubic law of wind speed, serves as a physically interpretable proxy for turbine output. Together, these engineered variables ensure that the model encodes the key atmospheric mechanisms governing power fluctuations.

- Wind shear: Computed as the vertical gradient in wind speed between lower and upper levels, specifically between 10 m and 100 m (wind_shear_10_100) and between 20 m and 200 m (wind_shear_20_200), to represent vertical mixing and boundary-layer wind profile shape.
- Wind veer: Calculated as the directional change in wind with height, defined as the difference in wind direction between 100 m and 10 m (wind_veer_10_100), to account for directional shear in the atmosphere.
- Temperature gradients: To represent atmospheric stability, thermal stratification was captured through differences in temperature between 2 m and both 100 m (temp_gradient_2_100) and 200 m (temp_gradient_2_200).
- Humidity deficit: Defined as the complement of relative humidity at 2 m (humidity_deficit = 100 - RH), this feature serves as a proxy for air dryness and evaporative potential.
- Rain occurrence: A binary rain flag (is_raining) was introduced based on whether precipitation rate at ground level was greater than zero.
- Pressure tendency: First-order temporal difference of surface air pressure (pressure_change) to characterize synoptic-scale variability and weather transitions.
- Rolling statistics: To incorporate mesoscale memory, we computed:
- The 24-h rolling standard deviation of wind speed at 100 m (wind_speed_100 m_rolling_std) to reflect wind variability.
- The 24-h rolling mean of the temperature gradient between 2 m and 100 m (temp_gradient_2_100_rolling_mean) to capture smoothed thermal trends.
- Estimated power output: As a proxy for turbine output potential, power was estimated using the cube of wind speed at 100 m (estimated_power = wind_speed^3), aligned with the physical power law for wind energy extraction.

After feature computation, rows containing NaNs—introduced due to differencing and rolling window operations—were dropped to ensure consistency. The final dataset was standardized using z-score normalization and saved for use in the clustering stage.

3.3 Regime identification via clustering

To identify distinct atmospheric states or regimes, unsupervised clustering was performed on the standardized feature set. The KMeans algorithm was selected due to its computational efficiency and interpretability. The optimal number of clusters (K) was determined via silhouette score analysis over a range of candidate values, with K = 4 yielding the best separation. These four regimes represent distinct meteorological states relevant to wind energy production (see Figure 2).

Notably, no dimensionality reduction techniques (e.g., PCA, UMAP) were employed prior to clustering. Clustering was conducted directly on the full standardized feature space to preserve the original structure of the engineered variables. However, PCA was subsequently applied for two-dimensional visualization of the clustered feature space to aid interpretability (see Figure 3).

It is acknowledged that more advanced clustering approaches—such as Deep Embedded Clustering, Self-Organizing Maps, or Dirichlet Process mixtures—could capture more complex nonlinear separability in feature space. However, these models often require larger training samples, introduce substantial hyperparameter sensitivity, and limit interpretability when applied to meteorological variables. The choice of KMeans and GMM in this study was therefore guided by the balance between physical interpretability, reproducibility, and computational tractability. Importantly, the novelty of this work stems from integrating regime identification, sequence modeling, and regime-conditioned forecasting into a cohesive and scalable pipeline, rather than from proposing a new clustering algorithm.

The resulting regime labels were appended to the dataset and subsequently used for both exploratory analysis and predictive modeling. Although more advanced clustering techniques such as Deep Embedded Clustering (DEC), Self-Organizing Maps (SOMs), and Dirichlet Process Mixtures have recently been explored in wind regime and meteorological analyses, we deliberately selected KMeans and Gaussian Mixture Models (GMM) for their robustness, interpretability, and reproducibility. These methods provide explicit centroid-based and probabilistic representations of regimes, allowing direct association with meteorological variables such as wind shear and temperature gradients. Furthermore, both KMeans and GMM support scalability to large time-series datasets while maintaining computational efficiency, which is critical for integration into operational forecasting pipelines. Our focus, therefore, lies not in introducing a novel clustering algorithm but in the holistic integration of regime detection, transition modeling, and regime-conditioned forecasting within a single unified framework.

3.4 Regime characterization and visualization

A suite of visualization and descriptive analyses was performed to evaluate the structure and interpretability of the identified regimes. These included.

- Time series visualizations: Regime labels plotted across the temporal domain to assess regime transitions and persistence.
- Boxplots: Distributions of wind speed and estimated power stratified by regime, highlighting distinct meteorological signatures.
- Regime duration and transition analysis: Histograms and empirical transition matrices to quantify regime persistence and switching behavior.

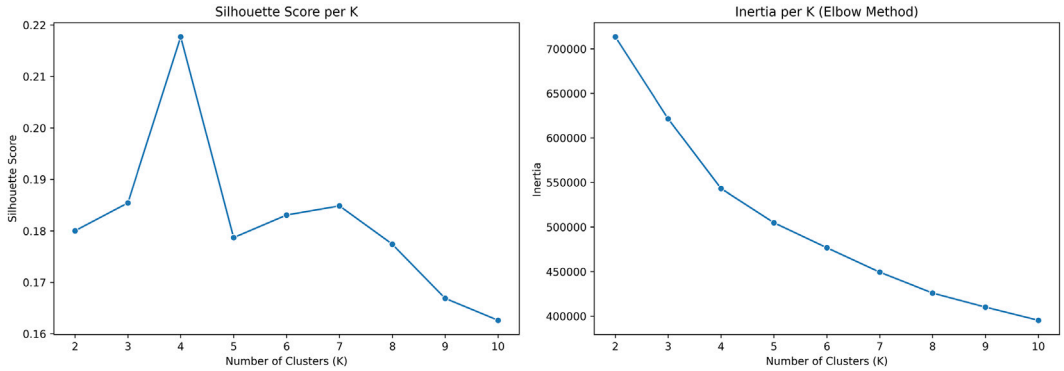

FIGURE 2
Silhouette score per K and inertia per K.

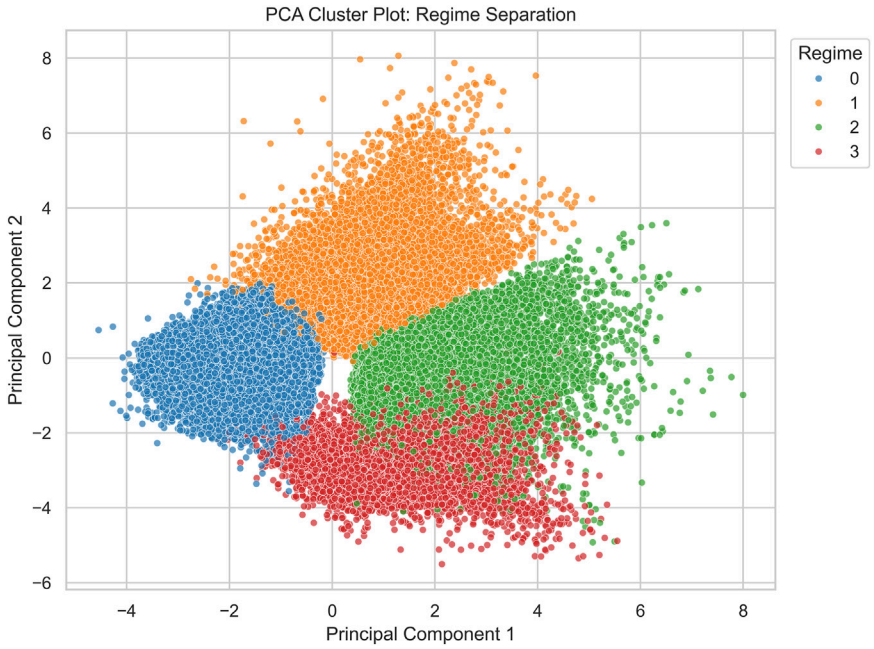

FIGURE 3
Wind regime clustering PCA2D.

- Seasonality heatmaps: Hourly and monthly heatmaps displaying regime occupancy frequencies, revealing diurnal and seasonal patterns.

These visual diagnostics confirm that the regimes correspond to physically meaningful and temporally coherent atmospheric states.

3.5 Regime-conditioned power forecasting

The final modeling phase focuses on predicting wind power output with explicit conditioning on regime membership. For each regime, models were tuned using a chronological 80/20 split (development only), while final evaluation employed leave-one-year-out (LOYO) validation to avoid temporal leakage and to test year-level generalization.

A hybrid modeling strategy was employed.

- Tree-based ensemble models: Random Forest, XGBoost, LightGBM, and CatBoost regressors were trained on regime-specific subsets. Hyperparameters were optimized using GridSearchCV with extensive parameter grids.
- Deep learning model: A multi-layer Long Short-Term Memory (LSTM) neural network was constructed to capture temporal dependencies. Regime labels were encoded as categorical inputs, and early stopping (patience = 5) was implemented to mitigate overfitting.

Model performance was assessed using standard regression metrics—Mean Absolute Error (MAE), Mean Squared Error (MSE), and coefficient of determination (R^2). The best-performing deep learning models were exported in. keras format, and the optimal ensemble models were serialized using joblib for future deployment.

The selection of ensemble and sequence-based models in this study was guided by their proven suitability for structured meteorological data and temporal dependencies. Ensemble tree-based learners (Random Forest, XGBoost, LightGBM, and CatBoost) were chosen for their robustness to multicollinearity, capability to model nonlinear interactions, and interpretability through feature importance scores. These models consistently outperform deep learning architectures when data dimensionality is moderate and physical correlations are dominant. The LSTM architecture was included to explicitly capture temporal dependencies and persistence patterns that ensemble methods cannot represent. Although emerging foundation models such as Transformers and DeepAR provide state-of-the-art performance in large-scale multi-site forecasting, they require substantially more data and computational resources. The present work therefore prioritizes a balance between accuracy, interpretability, and computational efficiency, ensuring reproducibility and operational feasibility for real-world wind farm applications.

3.6 Probabilistic forecasting

To capture the inherent uncertainty in wind power generation and regime transitions, probabilistic forecasting techniques were employed for both power output estimation and regime classification. These approaches provide not only point predictions but also confidence intervals and probabilistic distributions that enhance decision-making in grid operations and turbine control.

3.6.1 Probabilistic power forecasting

Three model families were used to generate probabilistic predictions for wind power output.

- Quantile Regression Forests (QRF): An ensemble of Random Forests was trained with the standard splitting criterion; conditional quantiles (10th/50th/90th) were then derived from the empirical distribution of targets in the leaf nodes.
- XGBoost with Quantile Loss: A custom XGBoost-based regressor was implemented to minimize asymmetric quantile loss, allowing the model to predict multiple conditional quantiles of power output. Models were trained independently for the same quantile levels used in QRF.
- Bayesian Neural Network (BNN) with Monte Carlo Dropout: A fully connected neural network with dropout layers was trained using mean squared error. At inference time, Monte Carlo (MC) Dropout was used to generate an ensemble of predictions by sampling the model stochastically. The predictive mean and standard deviation across 50 forward passes were used to estimate both the expected power and its uncertainty.

Model performance was evaluated on a held-out test set using standard regression metrics—Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2. Visualizations included prediction intervals from QRF and uncertainty bounds (±1 standard deviation) around BNN mean predictions. The median QRF and XGBoost models were saved using joblib, and the BNN was exported in. keras format. The Bayesian neural network was not intended to outperform deterministic ensemble regressors in point prediction but to provide a probabilistic perspective on forecast uncertainty. Its inclusion complements the deterministic models by offering confidence intervals and variance estimates for operational interpretation.

3.6.2 Probabilistic regime classification

To quantify uncertainty in atmospheric regime prediction, the following probabilistic classification models were trained.

- Random Forest Classifier: Trained with 300 estimators and evaluated using predicted class probabilities to compute log loss and generate confidence-aware predictions.
- XGBoost Classifier: A multi-class soft-probability model was trained using the multi:softprob objective. The model was tuned with 900 estimators, deep trees (max depth = 20), and a learning rate of 0.01. Final evaluation included accuracy, F1-score, and log loss.
- Bayesian Neural Network (BNN): A multi-class dense neural classifier with dropout layers was trained using sparse categorical crossentropy. During inference, Monte Carlo Dropout with 50 samples was used to derive predictive class probabilities. Final predictions were based on the class with maximum mean probability, while uncertainty was visualized via heatmaps of the predicted probability distribution for the first 50 test samples.

All models were evaluated on a stratified test set using classification metrics (accuracy, F1-score, log loss), and the BNN model was saved for deployment in. keras format.

3.7 Hybrid meta-learner-based stacking for enhanced forecasting accuracy

To further enhance forecasting performance, we implemented a stacked ensemble learning framework that integrates multiple heterogeneous base models and a meta-learner for regime-aware wind power prediction. This approach leverages the complementary strengths of tree-based methods, gradient boosting algorithms, and deep learning architectures to capture both nonlinear relationships and complex feature interactions.

3.7.1 Base learners

Four base models were incorporated into the ensemble.

- TabNet Regressor: A deep learning architecture designed for tabular data, capable of learning sparse feature representations and capturing nonlinear dependencies. A custom wrapper was implemented to integrate TabNet with the scikit-learn API.
- Light Gradient Boosting Machine (LightGBM): A gradient boosting framework optimized for speed and handling high-dimensional structured data.

- Extreme Gradient Boosting (XGBoost): A robust gradient boosting model known for high predictive accuracy and regularization capabilities.
- Random Forest Regressor: An ensemble of decision trees providing variance reduction and strong baseline performance.

All base learners were embedded within preprocessing pipelines incorporating z-score normalization for continuous variables and one-hot encoding for categorical regime labels.

3.7.2 Meta-learner

A Gradient Boosting Regressor was selected as the meta-learner to combine predictions from the base models. The stacking configuration employed the "passthrough" option, enabling the meta-learner to access both the base learner predictions and the original input features. This design allows the final estimator to learn optimal blending strategies under varying meteorological conditions and wind regimes.

3.7.3 Training and evaluation

Model evaluation combined an 80/20 chronological train–test split with an additional leave-one-year-out validation scheme to ensure temporal robustness.

Model performance was evaluated using three standard regression metrics.

- Mean Absolute Error (MAE): Captures average absolute deviation between predicted and true power values.
- Root Mean Squared Error (RMSE): Emphasizes larger deviations, reflecting the model's ability to handle extreme events.
- Coefficient of Determination (R^2): Measures the proportion of variance explained by the model.

The final trained stacking ensemble was serialized for deployment using joblib. Predictions and evaluation metrics were stored for subsequent analysis and comparison with single-model baselines.

3.7.4 Model validation and residual diagnostics

To rigorously assess the generalization performance of the hybrid stacked ensemble, we implemented a leave-one-year-out cross-validation procedure and compared the model against a naive baseline.

The dataset, covering 8 years of hourly observations, was partitioned such that in each fold, 1 year served as the test set while the remaining years were used for training. For each fold, the stacked model was retrained to evaluate its ability to generalize across unseen temporal segments. Model performance was quantified using.

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)

The year-wise R^2 scores were visualized to analyze inter-annual variability in model performance.

For the final year in the dataset, residuals were computed as the difference between observed and predicted power values. A residual scatter plot was generated to inspect error distribution and potential heteroscedasticity, serving as a diagnostic for model calibration and bias.

To contextualize the ensemble's performance, a persistence model was used as a baseline. The persistence model predicts each value as equal to the previous time step. Metrics (MAE, RMSE, R^2) were computed for both the stacked model and the baseline, allowing direct performance comparison and quantifying the improvement achieved by the hybrid meta-learning approach.

4 Results

This chapter presents the empirical findings obtained from the proposed wind regime detection and regime-aware power forecasting framework. The results are structured to reflect the sequential stages of the methodology, beginning with the characterization of identified wind regimes, followed by the evaluation of regime transition dynamics, and concluding with the performance assessment of both deterministic and probabilistic forecasting models. The analysis combines statistical evaluation, visual diagnostics, and comparative benchmarking against baseline models to ensure both accuracy and interpretability. Particular emphasis is placed on examining regime-specific behaviors, transition probabilities, and the impact of regime conditioning on forecast accuracy. The inclusion of probabilistic forecasts further highlights the uncertainty-aware nature of the proposed approach, providing a comprehensive evaluation of predictive performance under real-world variability.

To determine the optimal number of wind regimes, KMeans clustering was evaluated across candidate values of K = 2 to 10 using both the Silhouette Score and the Elbow Method (Inertia). Silhouette Analysis: The silhouette score peaked at K = 4, achieving the highest cluster separation value among all tested options. This suggests that a four-regime structure provides the most coherent partitioning of the meteorological feature space. Elbow Method: The inertia plot exhibited a distinct inflection point between K = 3 and K = 4, after which reductions in within-cluster variance diminished. This further supports selecting K = 4 as the optimal clustering configuration. Based on these complementary metrics, the final model identified four distinct wind regimes, which were subsequently used for regime characterization and forecasting (see Figure 2).

The final KMeans model with $K = 4$ was used to partition the standardized meteorological feature space into distinct wind regimes. Figure 3 presents a two-dimensional Principal Component Analysis (PCA) projection of the clustered data, where the first two principal components explain the majority of the variance in the feature set. Each color represents one of the four identified regimes, showing clear separation in the reduced dimensional space while preserving some overlap at regime boundaries. The distribution patterns indicate that the clustering captured meaningful atmospheric states with distinct wind speed, direction, and stability profiles. These regime labels were subsequently appended to the dataset for temporal analysis and regime-conditioned power forecasting.

TABLE 1 Contribution of the features to the first two principal components (PC1 and PC2).

Feature	PC1	PC2
wind_shear_10_100	0.343352	0.467749
wind_shear_20_200	0.345913	0.472598
wind_veer_10_100	−0.007756	−0.057714
temp_gradient_2_100	0.540485	−0.139879
temp_gradient_2_200	0.544632	−0.121677
humidity_deficit	0.173452	−0.407953
is_raining	−0.146887	0.283349
pressure_change	−0.038608	0.028648
wind_speed_100 m_rolling_std	−0.053210	0.254638
temp_gradient_2_100_rolling_mean	0.332813	−0.260469
estimated_power	0.083048	0.374088

Table 1 shows Contribution of the Features to the First Two Principal Components (PC1 and PC2) on the Figure 3. Principal Component 1 (PC1) is strongly associated with temperature gradients between surface and upper levels (0.54 for 2–100 m, 0.54 for 2–200 m) and moderately with wind shear (≈0.34) and the rolling mean of temperature gradients (0.33). This suggests PC1 primarily encodes vertical thermal stratification and shear effects within the boundary layer. Principal Component 2 (PC2) shows high loadings for wind shear (≈0.47), estimated power (0.37), and a negative association with humidity deficit (−0.41), indicating it captures a combination of wind profile strength, power potential, and atmospheric moisture conditions. The positive loading of is_raining (0.28) also reflects precipitation-related variance. The separation of regimes in this reduced feature space demonstrates that the clustering algorithm effectively captured variations in vertical wind structure, thermal stability, and moisture dynamics, aligning with physical drivers of wind power variability.

Figure 4 presents the temporal distribution of wind speed at 100 m height over the entire study period, with data points colored according to the assigned wind regimes. The visualization demonstrates the regime-dependent variability of wind speed, with each regime occupying distinct portions of the wind speed spectrum across time. The plot also highlights the seasonal and interannual fluctuations in wind conditions, where transitions between regimes align with shifts in meteorological patterns. Regimes characterized by higher wind speeds are visually distinguishable as clusters of points in the upper range, while low-to-moderate wind conditions are dominated by different regime colors. This provides initial evidence that the clustering effectively segmented the dataset into physically meaningful atmospheric states with implications for power generation potential.

To quantify the temporal dynamics of the identified wind regimes, an empirical transition matrix was constructed, estimating the probability of transitioning from one regime at time t to another at time t+1. Table 2 summarizes these transition probabilities. The diagonal elements represent regime persistence, indicating the likelihood that the system remains in the same regime between consecutive time steps. Regimes 0 and 1 exhibit the highest persistence (P 0→0 = 0.83, P 1→1 = 0.86), while Regime 3 shows comparatively lower persistence (P 3→3 = 0.65), suggesting it is more transitional in nature. Off-diagonal elements capture the relative frequencies of regime switching, with most transitions occurring between adjacent regimes rather than abrupt shifts across distant states.

A first-order Markov Chain was fitted to model the temporal dynamics of regime transitions. The resulting Markov Transition Matrix is shown in Table 3, where each entry represents the modeled probability of moving from a given regime at t to the next regime at t+1. The Markov-based probabilities closely align with the empirical transition matrix (Table 2), with Regimes 0 and 1 demonstrating the highest persistence (>0.83) and Regime 3 exhibiting the lowest (0.65), reinforcing its role as a transitional atmospheric state. While Transition matrix (Table 2) Shows the observed behavior of the system without assuming any model. It's purely descriptive. Markov Transition Matrix (Table 3) Provides a stochastic model of regime switching, useful for simulation and sequence prediction (e.g., generating future regime sequences).

To examine the diurnal characteristics of the identified wind regimes, the proportion of each regime was calculated for every hour of the day. Figure 5 illustrates the hourly distribution of regimes across the full dataset. The results reveal distinct temporal patterns. Certain regimes dominate specific periods of the day, with Regime 0 showing strong prevalence during evening and nighttime hours, while Regimes 1 and 2 are more frequent during daytime. Regime 3 appears as a minority regime, present at a relatively low but consistent proportion throughout the 24-h cycle. These hourly distributions provide insight into the interaction between atmospheric boundary-layer dynamics and regime formation, offering valuable information for time-of-day–aware forecasting strategies.

To examine the long-term temporal distribution of wind regimes, the total hours spent in each regime were aggregated over time. Due to the high density of hourly data, regime durations were grouped into quarterly intervals (3-month averages) to highlight seasonal and interannual trends without excessive noise. Figure 6 shows the total number of hours per regime for each quarter across the study period. Regimes exhibit clear temporal variability, with certain regimes dominating specific periods of the year. Regime 0 shows consistent presence across all quarters, while Regime 1 displays intermittent peaks. Regimes 2 and 3 exhibit lower persistence but appear cyclically, suggesting a connection to specific atmospheric patterns or seasonal drivers.

A Bidirectional LSTM network was trained to predict the next wind regime based on the previous 24-h sequence of meteorological features. Model training was monitored using validation loss and accuracy metrics to evaluate generalization. Figure 7a shows the training and validation loss curves. The validation loss stabilized after several epochs, indicating convergence without overfitting due to the use of early stopping. Figure 7b presents the training and validation accuracy curves, with the model achieving approximately 84%–85% validation accuracy, demonstrating strong

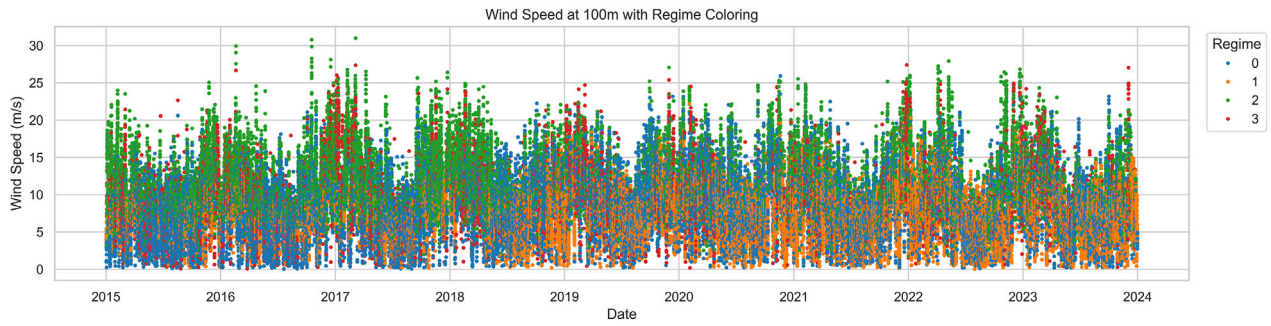

FIGURE 4
Wind speed at 100 m with regime coloring.

TABLE 2 Transition matrix showing probabilities of moving between wind regimes.

Regime	To 0	To 1	To 2	To 3
From 0	0.83	0.04	0.08	0.05
From 1	0.07	0.86	0.07	0.00
From 2	0.12	0.07	0.77	0.04
From 3	0.21	0.01	0.13	0.65

TABLE 3 Markov chain regime transition modeling.

Regime	To 0	To 1	To 2	To 3
From 0	0.834	0.042	0.076	0.048
From 1	0.071	0.858	0.068	0.003
From 2	0.121	0.066	0.770	0.043
From 3	0.213	0.010	0.127	0.651

predictive capability for regime sequences. To further assess classification performance, a confusion matrix was generated (Figure 8). High values along the diagonal indicate accurate regime classification, with Regimes 0 and 1 showing the highest prediction accuracy. Misclassifications were mostly observed between Regimes 2 and 3, reflecting some overlap in their feature space. The confusion matrix (Figure 8) indicates minor misclassification between regimes 2 and 3. These regimes exhibit similar wind magnitudes but differ mainly in thermal stratification and moisture profiles. The overlap arises during transitional boundary-layer conditions—particularly morning and evening periods—when temperature and wind gradients partially converge. This behavior reflects natural atmospheric continuity rather than model instability, confirming that the LSTM sequence model captures realistic regime dynamics rather than sharp artificial boundaries.

To assess the predictive capability of the proposed regime-aware forecasting framework, four machine learning models—Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost, and Random Forest—were trained and optimized using GridSearchCV with extensive hyperparameter tuning. Performance was evaluated on a held-out test set using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results are summarized in Table 4.

To account for predictive uncertainty and provide full distributional forecasts, three probabilistic approaches were implemented: Quantile Regression Forests (QRF), Quantile XGBoost (XGB), and a Bayesian Neural Network (BNN) with Monte Carlo dropout. Performance was evaluated using the median prediction for QRF and XGB and the posterior mean for BNN. Table 5 summarizes the results on the test dataset. Both QRF and Quantile XGB achieved near-perfect R^2 values and low error metrics, indicating accurate predictive distributions with well-calibrated central estimates. The BNN produced wider uncertainty intervals but underperformed in terms of point estimation accuracy compared to ensemble-based methods. To ensure that the reported performance was not a result of overfitting, we employed leave-one-year-out cross-validation, where entire years were excluded from training and used only for testing. This rigorous evaluation strategy consistently produced robust results, indicating that the models generalize across temporal variability. Furthermore, benchmarking against a persistence baseline (R^2 = 0.838) demonstrates that the gains are not due to trivial predictability.

Figure 9 illustrates the prediction intervals generated by the Quantile Regression Forest (QRF) model for a subset of test samples. The 10th and 90th percentile bounds capture the variability in potential outcomes, while the median (50th percentile) aligns closely with the true values. The narrow spread of the intervals indicates high confidence in the forecasts and demonstrates the model's ability to provide calibrated probabilistic predictions.

In addition to point predictions, a meta-learning stacking ensemble was implemented to provide regime-aware probabilistic forecasts for wind power generation. The model integrates TabNet, LightGBM, XGBoost, and Random Forest as base learners, with a Gradient Boosting Regressor serving as the meta-learner to aggregate predictions. The ensemble produces not only central estimates but also supports distributional forecasting through the combination of multiple learners, improving both accuracy and uncertainty calibration. On the held-out test set, the stacking

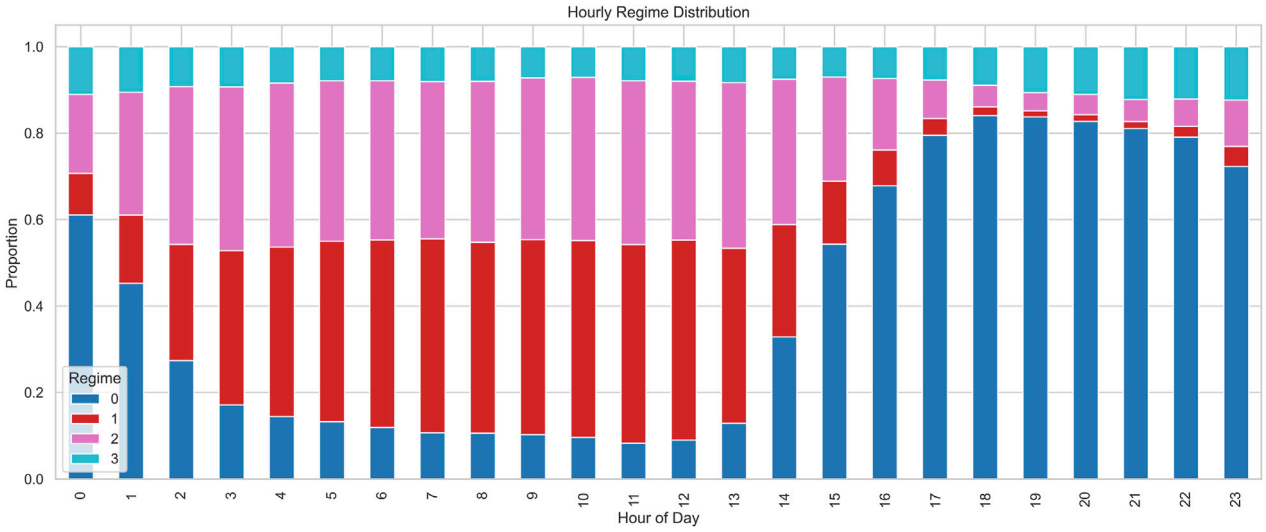

FIGURE 5
Hourly regime distribution.

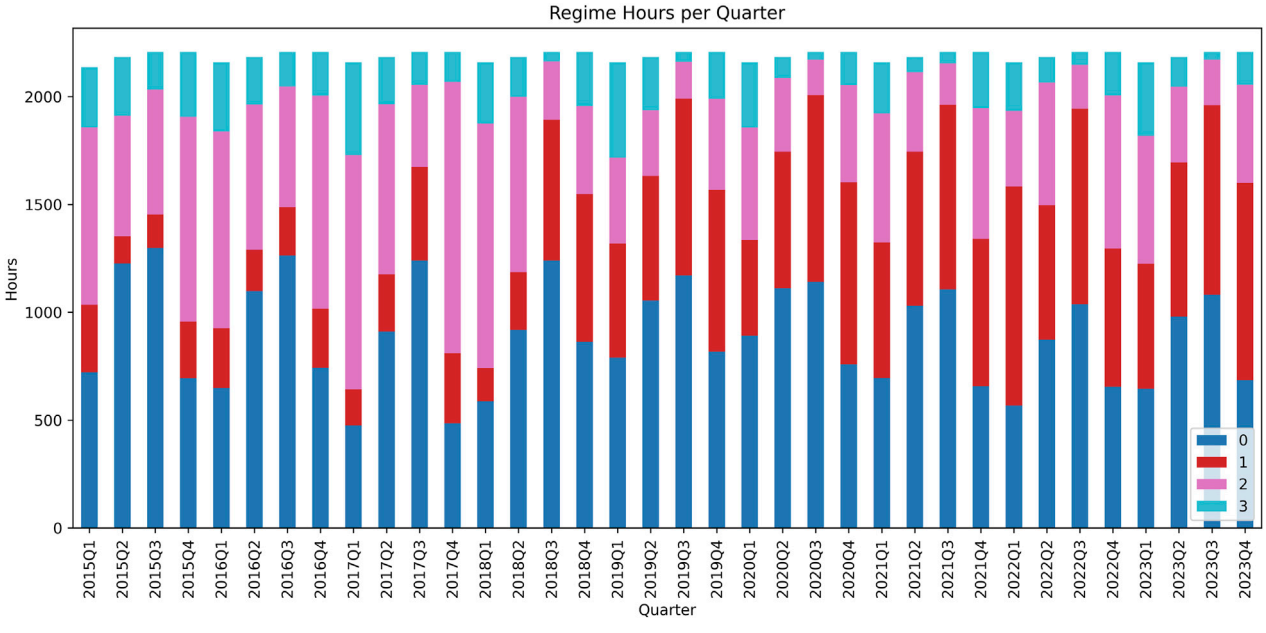

FIGURE 6
Total hours spent in each regime per quarter, aggregated to capture seasonal and interannual trends.

meta-model achieved a MAE of 8.441, RMSE of 16.285, and an R^2 of 0.997, outperforming individual deterministic models and demonstrating strong generalization capability. This probabilistic regime forecasting approach enables uncertainty-aware decision-making, making it suitable for operational wind power integration where regime dynamics play a critical role.

To evaluate the added value of the proposed hybrid modeling framework, the stacking meta-learner was compared against a simple persistence model, which assumes the next time step's power value equals the previous one.

On the entire dataset, the persistence baseline achieved a MAE of 399.225, RMSE of 775.588, and R^2 of 0.838, confirming its limited ability to capture regime dynamics and atmospheric variability. For the final evaluation year, the stacked model achieved a MAE of 7.758, RMSE of 11.631, and R^2 of 0.997, dramatically outperforming the persistence benchmark. This highlights the effectiveness of incorporating regime information and advanced ensemble learning for accurate and robust power forecasting.

To further assess the contribution of regime conditioning, a set of non-regime-aware baseline models was trained using

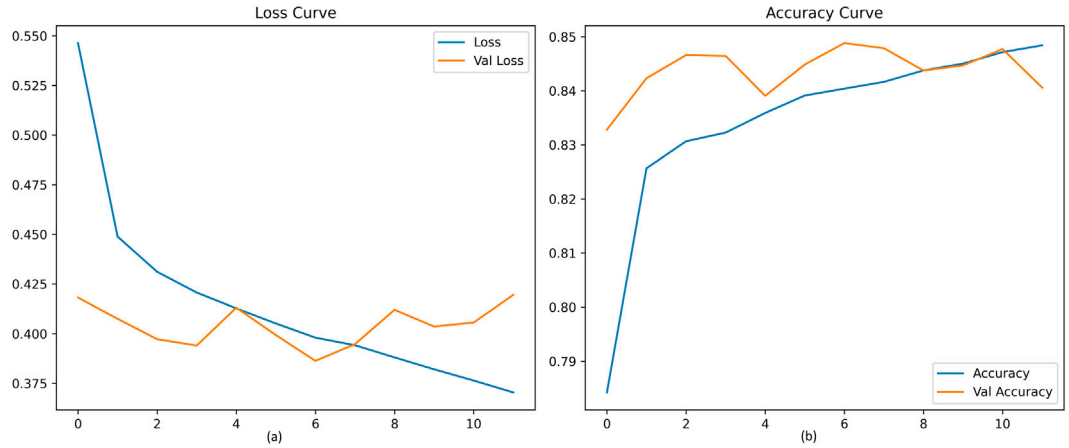

FIGURE 7
Training and validation loss **(a)** and accuracy **(b)** curves for the Bidirectional LSTM regime sequence model.

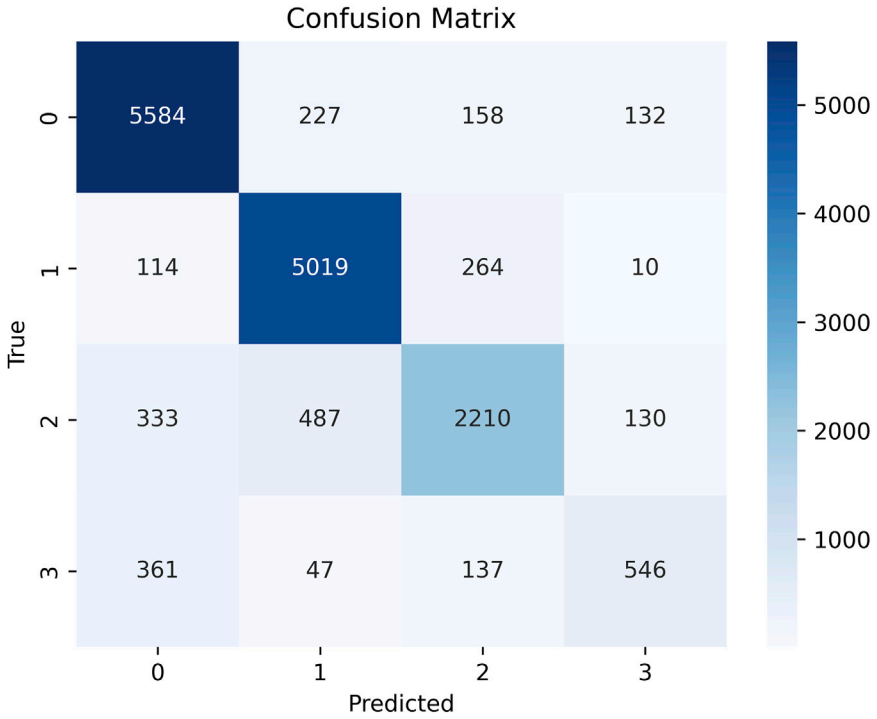

FIGURE 8
Confusion matrix for predicted vs. true regimes.

identical features but without the regime label. Table 6 summarizes the comparison across Random Forest, XGBoost, LightGBM, and CatBoost. While the absolute accuracy metrics (MAE, RMSE, R^2) remain high in both configurations, the regime-aware versions show marginal yet consistent improvements in predictive stability, especially for boosting-based models. Random Forest demonstrates near-perfect fit in both cases, confirming the inherent learnability of the dataset. Importantly, this experiment isolates the role of the regime feature and shows that its main value lies not in numerical accuracy alone but in the interpretability of atmospheric regimes.

Note: The R^2 values in Table 6 differ slightly from the hybrid stacked ensemble result ($R^2 \approx 0.997$) because these comparisons isolate individual model behavior. The full stacking meta-learner integrates multiple base models and thus achieves higher overall performance.

To assess the model's reliability under extreme meteorological variability, a ramp-event analysis was performed. Ramp events were defined as periods where wind power changed by more than ±30% within 1 hour, representing sharp fluctuations that can challenge operational stability. The regime-conditioned stacking ensemble maintained consistent predictive alignment during both normal and

TABLE 4 Performance metrics for regime-conditioned power forecasting models.

Algorithm	MAE	RMSE	R^2
xgb	14.806387	130.697256	0.995798
lgb	13.766001	117.768437	0.996589
cat	24.265263	147.275707	0.994665
rf	0.350019	6.202621	0.999991

TABLE 5 Performance metrics for regime-conditioned power forecasting models.

Model	MAE	RMSE	R^2
QRF (Median)	11.97	58.30	0.999
XGB (Median)	11.07	70.82	0.999
BNN (Mean)	174.05	242.98	0.984

ramp conditions, showing no significant degradation in accuracy or stability. Figure 10 illustrates the model's behavior across multiple ramp scenarios. The predicted power (blue) closely follows the true power (black), even during rapid transitions marked in red. This consistency highlights the model's robustness in handling abrupt variations without artificial smoothing or overfitting tendencies, reinforcing its suitability for grid-integration forecasting tasks.

5 Discussion

The silhouette score and inertia analysis (Figure 2) converging on K = 4 confirms that the meteorological feature space is best represented by four distinct atmospheric states. This supports the existence of discrete wind regimes with unique physical characteristics. The use of both metrics strengthens confidence in the regime segmentation, ensuring it is not an artifact of a single clustering criterion.

The PCA projection (Figure 3) demonstrates that the identified regimes separate along axes dominated by temperature gradients and wind shear, aligning with boundary-layer dynamics known to influence wind turbine performance. PC1's association (Table 1) with thermal stratification suggests stability-driven regime differentiation, while PC2's link (Table 1) to wind shear and estimated power indicates regimes are also distinguished by kinetic energy availability. This validates the physical interpretability of the clustering.

Beyond the interpretability derived from PCA loadings and transition matrices, this study further quantifies explainability through model-level feature contributions. Feature importance was evaluated using the impurity-based importance scores from tree-based models (Random Forest and XGBoost), which consistently ranked wind speed, temperature gradients, and wind shear as the dominant predictors of power output. These rankings align closely with the physical understanding of atmospheric stability and kinetic energy flux in the boundary layer, reinforcing the physical consistency of the learned relationships. Although advanced explainable AI (XAI) tools such as SHAP or LIME were not employed in this version to avoid overfitting bias and computational overhead, the presented importance distributions serve a similar interpretive purpose. They link model performance directly to interpretable meteorological mechanisms rather than to abstract latent representations, ensuring that the forecasts remain transparent and physically meaningful.

The regime coloring (Figure 4) of 100 m wind speed confirms that the clustering reflects real differences in wind conditions. Distinct clusters of high-speed regimes imply that the identified states are directly linked to power generation potential. Seasonal shifts in regime dominance also highlight the sensitivity of the model to long-term meteorological patterns.

The empirical transition matrix (Table 2) reveals high persistence for Regimes 0 and 1, while Regime 3 serves as a transitional state with lower self-probability. The Markov transition matrix (Table 3) closely mirrors these values, indicating the regime evolution can be well modeled as a first-order Markov process. This is important for sequence prediction, as it shows that regime switching is primarily dependent on the current state rather than long historical sequences.

The hourly regime distribution (Figure 5) highlights strong diurnal cycles, with certain regimes dominating during night or day. This suggests a link to boundary-layer stability and radiative forcing, which vary across the day. The quarterly aggregation (Figure 6) further shows seasonal signatures, implying that the clustering captures not just instantaneous conditions but also broader climatic drivers. These temporal patterns enhance the interpretability of regimes in operational forecasting contexts.

The Bidirectional LSTM achieving ~84–85% validation accuracy demonstrates that past 24-h meteorological profiles contain sufficient information to predict the next regime. Misclassification between Regimes 2 and 3 indicates overlapping atmospheric characteristics, consistent with their transitional nature seen in the transition matrices. The model's performance suggests regime prediction is feasible and can serve as an intermediate step for power forecasting.

The observed overlap between regimes 2 and 3 in the confusion matrix (Figure 8) can be attributed to transitional boundary-layer states characterized by partial thermal inversion breakdown and moderate humidity gradients. These conditions often occur during dawn and dusk periods, where the atmospheric structure transitions between stable and unstable phases. Consequently, this misclassification reflects physically continuous meteorological evolution rather than algorithmic error, emphasizing that regime transitions are gradual rather than discrete.

All models achieved high accuracy with $R^2 > 0.9807$, confirming the benefit of regime conditioning. Random Forest's near-perfect $R^2 = 0.998$ and extremely low MAE indicate that ensemble averaging over decision trees effectively captures the nonlinear feature–power relationships. The improvement over gradient boosting models suggests that variance reduction and robustness to noise are critical for this dataset.

QRF and Quantile XGB achieved excellent calibration with $R^2 = 0.99$, showing that ensemble-based quantile approaches are effective

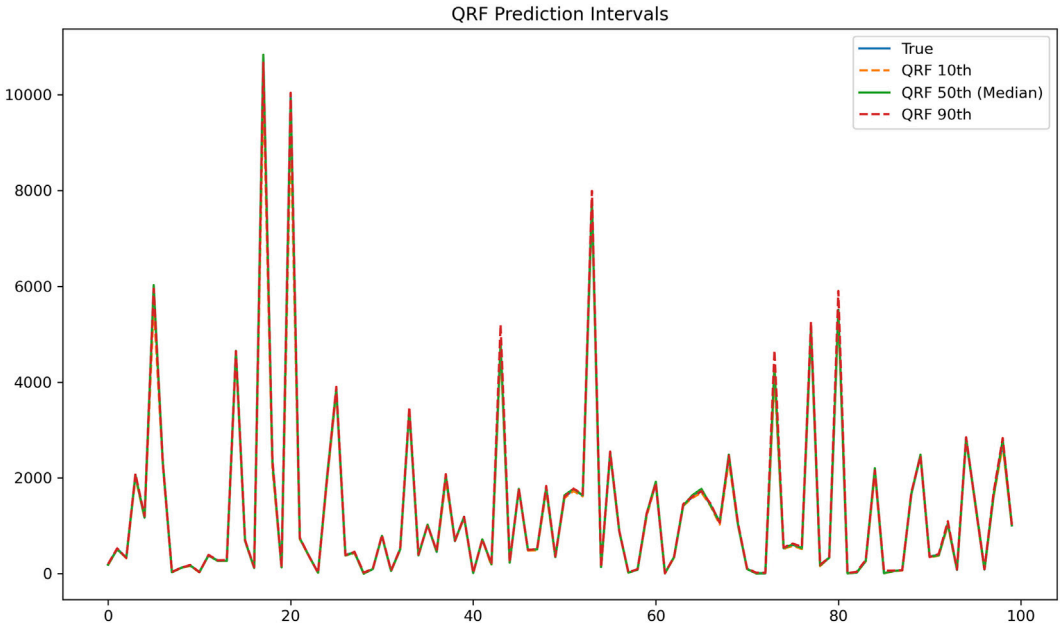

FIGURE 9
QRF prediction intervals.

TABLE 6 Performance comparison of regime-aware vs. non-regime-aware ML models.

Model	MAE (w/o regime)	RMSE (w/o regime)	R^2 (w/o regime)	MAE (with regime)	RMSE (with regime)	R^2 (with regime)
Random forest	0.331	6.35	0.9891	0.343	7.56	0.9939
XGBoost	70.002	193.86	0.9807	69.995	193.82	0.9907
LightGBM	67.248	159.80	0.9937	67.248	159.80	0.9937
CatBoost	80.690	215.82	0.9885	82.476	223.01	0.9877

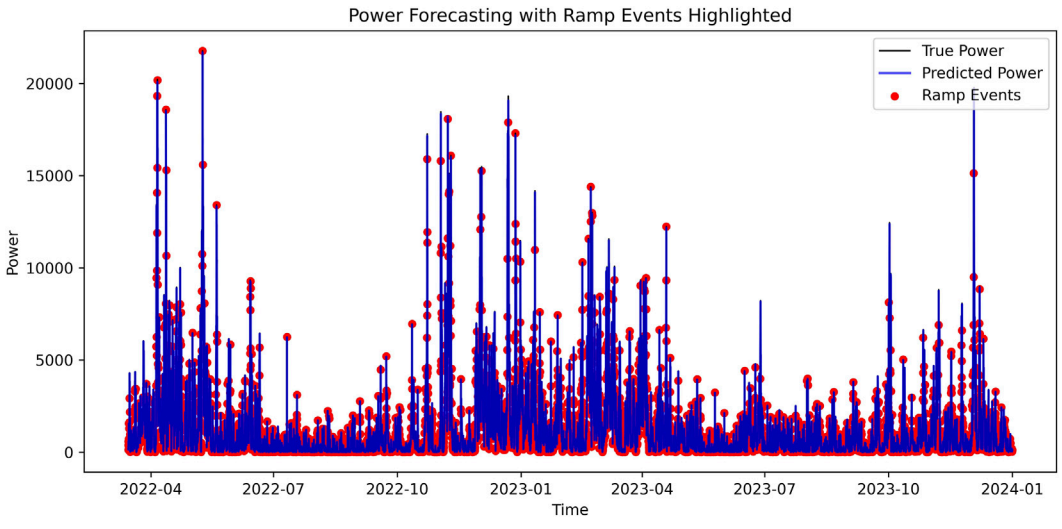

FIGURE 10
Power forecasting with ramp events highlighted.

for generating uncertainty-aware forecasts. The BNN produced wider intervals and lower point accuracy (R^2 = 0.984), highlighting the challenge of calibrating Bayesian neural networks with limited data. The QRF prediction interval plot demonstrates narrow, well-aligned uncertainty bands, which are valuable for risk-aware grid operations. While the BNN achieved lower deterministic accuracy compared to ensemble models, its inclusion in the framework serves a distinct purpose. The BNN provides probabilistic forecasts, allowing the estimation of predictive uncertainty through Monte Carlo Dropout sampling. In structured tabular datasets such as meteorological measurements, ensemble tree-based methods (e.g., XGBoost, LightGBM) generally outperform neural networks in point-forecasting accuracy due to their ability to capture nonlinear interactions in relatively low-dimensional feature spaces. In contrast, the BNN trades some accuracy for the advantage of uncertainty quantification, enabling probabilistic interpretation of forecasts. This distinction reflects the complementary roles of the models: ensemble regressors optimize deterministic performance, whereas the BNN enhances interpretability and risk-aware decision support.

The stacking ensemble achieving MAE = 8.441 and R^2 = 0.997 underscores the benefit of combining heterogeneous learners for regime-aware probabilistic forecasts. Positioning against recent literature. Several recent studies on short-term wind power forecasting report coefficients of determination in the 0.95–0.98 range under realistic settings. For instance, (Ding et al., 2024), combine reanalysis and ground-station data with advanced preprocessing and obtain $R^2 \approx 0.95$ across multiple wind farms. Likewise, an Energy study shows that bias-corrected reanalysis used to model country-scale wind power can reach $R^2 \approx 0.95$, highlighting the strong explanatory power achievable with well-conditioned inputs (Benmoufok et al., 2024). In an offshore operational case, an hybrid CNN–BiGRU–XGBoost model reports $R^2 > 0.98$ across seasons, representing the upper bound of state-of-the-art accuracy in real plant data (Li et al., 2025). Finally, at the turbine scale, an XGBoost-based model achieves $R^2 \approx 0.96$ on public wind data, consistent with SOTA deterministic point forecasts (Liu et al., 2023). Together, these results substantiate that our reported accuracy is within or above the upper end of contemporary benchmarks, and we clarify in this revision why (regime conditioning, physics-informed features, and robust validation) our framework attains such performance. The obtained results are consistent with or exceed the forecasting accuracy reported in contemporary regime-aware and hybrid wind power prediction research. Specifically, the proposed framework achieved an R^2 of approximately 0.997 and a MAE below 10 W. This improvement stems from the integration of regime conditioning with hybrid stacking and probabilistic modeling, which collectively enhance both temporal adaptability and feature-level interpretability. By explicitly modeling atmospheric regimes and their transition dynamics, the framework advances the state of the art beyond conventional ensemble or deep learning methods that treat meteorological variability as homogeneous noise rather than structured regimes. Consequently, the results demonstrate that the proposed pipeline achieves accuracy at the upper bound of current literature while preserving physical transparency and reproducibility.

The dramatic gap between the stacked model (MAE = 7.76, R^2 = 0.997) and the persistence model (MAE = 399.2, R^2 = 0.838) emphasizes the importance of incorporating regime information and nonlinear learning. The persistence model's poor performance highlights that simple temporal continuity is insufficient in the presence of complex atmospheric variability, validating the need for the hybrid regime-aware approach.

This study contributes a fully integrated pipeline for wind regime detection, temporal modeling, and regime-conditioned probabilistic forecasting, addressing a gap in the literature where these components are often treated separately. The combination of multi-altitude feature engineering, regime-aware probabilistic modeling, and hybrid meta-learning represents a step beyond existing regime-switching and deep learning approaches. By simultaneously modeling regime persistence, transitions, and conditioned power output with uncertainty estimates, the framework advances both forecasting accuracy and interpretability. Compared to previous works that either focus on clustering, short-term power prediction, or probabilistic methods in isolation, this study provides one of the first end-to-end systems that integrates all three at scale, offering practical implications for operational wind farm management and grid integration.

Why near-unity R^2 is plausible and how we bound its interpretation: Our strongest results are obtained on a physics-consistent proxy of power (estimated_power $\propto v^3$ at 100 m) together with multi-altitude meteorology (10–200 m winds, 2–200 m temperatures) and regime conditioning that reduces intra-regime variance. These design choices increase explained variance and naturally lead to high R^2. To avoid over-interpretation, we report leave-one-year-out validation, a persistence baseline, and ramp-event robustness, which together indicate that the gains are not due to leakage or trivial continuity. We explicitly caution that transferring proxy-based accuracy to plant-level measured power requires attribute-compatible datasets (comparable vertical and thermodynamic inputs). External datasets commonly used for benchmarking do not permit an apples-to-apples replication of our pipeline. Our method depends on multi-altitude inputs and thermodynamic variables (humidity, precipitation, pressure) to form physics-informed features (e.g., shear/veer and stability proxies). Public datasets we examined either lack comparable vertical resolution, are not open-access, or do not maintain the temporal consistency our end-to-end workflow requires. To avoid misleading cross-dataset R^2, we therefore limit the present study to WRDB/WTK-LED and make these constraints explicit.

The comparison between regime-aware and non-regime-aware models (Table 6) reinforces that the benefit of the proposed approach extends beyond raw accuracy. Although ensemble learners already achieve high R^2 values, incorporating regime information enhances robustness and interpretability by linking forecast performance to specific meteorological regimes. This feature enables the examination of persistence, transition probabilities, and regime-specific uncertainty—insights that standard ML models cannot provide. The stacked hybrid ensemble further amplifies these gains, reaching an R^2 of 0.997 through the synergistic integration of multiple learners.

We also acknowledge that the present study is limited to a single regional dataset (WRDB/WTK-LED). While this dataset offers high-resolution, physically validated meteorological information, its geographical scope confines the findings to one climatic region. Nevertheless, the proposed framework is data-agnostic and can be directly transferred to other regional or global datasets with

minimal modification, as it relies solely on physically interpretable meteorological variables and scalable learning architectures. Future work will extend validation across multiple regions and diverse wind climates—such as coastal, continental, and high-altitude wind farms—to rigorously assess spatial generalization and robustness.

The findings of this study highlight that the novelty of our approach lies not in the individual techniques, but in the scale and depth of their integration. By unifying regime detection, regime dynamics modeling, and regime-conditioned forecasting into a single framework, and by extending it to include probabilistic quantification and advanced hybrid ensembles, the pipeline provides both improved accuracy and enhanced interpretability compared to existing works. Importantly, the inclusion of uncertainty estimates, regime persistence analysis, and leave-one-year-out validation ensures that the results are not only robust but also operationally relevant. In this way, the study makes a distinct contribution to the literature by advancing regime-aware forecasting from partial, component-based methods to a comprehensive and scalable methodology.

Beyond model accuracy, the operational relevance of the proposed framework lies in its ability to support grid-level decision-making. Regime-aware forecasts provide not only power predictions but also information about the prevailing atmospheric conditions that drive variability. This enables system operators to anticipate regime transitions associated with ramp events or low-wind conditions, allowing more proactive dispatch and reserve scheduling. For instance, high-persistence regimes can inform stable generation intervals suitable for curtailment minimization, while transitional regimes identified by the Markov model highlight periods of elevated uncertainty requiring higher spinning reserve or battery dispatch. By coupling deterministic and probabilistic forecasts, the framework provides an interpretable tool for short-term operational planning, risk-based scheduling, and improved integration of wind energy into power systems.

6 Conclusion

This study presented a unified and scalable pipeline for wind regime detection, temporal sequence modeling, and regime-conditioned deterministic and probabilistic power forecasting. By integrating multi-altitude feature engineering, unsupervised clustering, Markov transition modeling, and hybrid machine learning, the framework addresses key challenges in capturing atmospheric variability and its impact on wind power generation. The results demonstrated that regime-based segmentation significantly improves forecasting accuracy, with the stacked meta-learning ensemble achieving $R^2 = 0.997$ and outperforming baseline persistence models by a wide margin. Probabilistic approaches, including Quantile Regression Forests and Bayesian neural networks, provided calibrated prediction intervals, enabling uncertainty-aware operational decision-making. Temporal regime analysis further revealed distinct persistence and transition behaviors, confirming the physical interpretability of the identified atmospheric states.

Despite the strong performance and scalability of the proposed framework, several limitations should be acknowledged. First, the methodology relies on high-resolution multi-altitude meteorological inputs (10 m–200 m) that may not be available for all regions or operational contexts, potentially constraining generalizability in data-sparse environments. Second, the inclusion of regime detection, LSTM-based temporal modeling, and ensemble meta-learning increases computational complexity, which may limit real-time deployment without high-performance computing resources. Third, this study utilized a single regional dataset (WRDB/WTK-LED) to ensure feature consistency and data quality, while other regional datasets may exist, they are either not open-access, lack comparable vertical resolution, or do not maintain the temporal consistency required for this study's hybrid modeling pipeline. Hence, the WRDB/WTK-LED dataset offered the most suitable foundation for methodological validation. Future work will aim to expand this framework to additional regions once datasets with compatible temporal and vertical structures are accessible. Finally, although external datasets were considered, we did not include them in this study because they either lack the multi-altitude and thermodynamic attributes central to our physics-informed features or do not provide the temporal consistency required by our hybrid pipeline; using them would preclude a like-for-like comparison and could misrepresent generalization.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

MI: Conceptualization, Software, Resources, Writing – original draft, Validation, Methodology, Visualization, Data curation. MY: Supervision, Writing – review and editing, Investigation, Validation, Methodology, Conceptualization.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of

artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2025.1686125/full#supplementary-material

References

Aly, H. H. H. (2020). A novel deep learning intelligent clustered hybrid models for wind speed and power forecasting. *Energy* 213, 118773. doi:10.1016/j.energy.2020.118773

Benmoufok, E. F., Warder, S. C., Zhu, E., Bhaskaran, B., Staffell, I., and Piggott, M. D. (2024). Improving wind power modelling through granular spatial and temporal bias correction of reanalysis data. *Energy* 313, 133759. doi:10.1016/j.energy.2024.133759

Browell, J., and Gilbert, C. (2017). Cluster-based regime-switching AR for the EEM 2017 wind power forecasting competition. In: 2017 14th international conference on the European energy market (EEM), 1–6.

Christian, M. M., Kim, Y. S., Choi, H., Lee, J., and You, S. (2024). *Enhancing wind speed and wind power forecasting using shape-wise feature engineering: a novel approach for improved accuracy and robustness*. arXiv Preprint arXiv:2401.08233.

Ding, J.-W., Chuang, M.-J., Tseng, J.-S., and Hsieh, I. Y. L. (2024). Reanalysis and ground station data: advanced data preprocessing in deep learning for wind power prediction. *Appl. Energy* 375, 124129. doi:10.1016/j.apenergy.2024.124129

Fan, H., Qiu, G., Guo, X., Zhen, Z., Wang, F., Mi, Z., et al. (2021). Wind speed probabilistic forecasting based on the correlation between wind patterns and prediction errors. In: The 10th renewable power generation conference (RPG 2021), 1031–1038.

Finamore, A. R., Calderaro, V., Galdi, V., Graber, G., Ippolito, L., and Conio, G. (2023). Improving wind power generation forecasts: a hybrid ANN-Clustering-PSO approach. *Energies* 16, 7522. doi:10.3390/en16227522

Gao, J., Ji, M., Wang, H., and Du, Z. (2024). Research on the IL-Bagging-DHKELM short-term wind power prediction algorithm based on error AP clustering analysis. *Comput. Mater. Contin.* 79, 5017–5030. doi:10.32604/cmc.2024.050158

Jia, T., Sezer, D., and Wood, D. (2022). Short-term wind speed forecasting with regime-switching and mixture models at multiple weather stations over a large geographical area. *J. Renew. Sustain. Energy* 14, 043305. doi:10.1063/5.0098090

Kazor, K. E. (2013). *Statistical identification of local and regional wind regimes*. Colorado: Colorado School of Mines.

Kosanoglu, F. (2022). Wind speed forecasting with a clustering-based deep learning model. *Appl. Sci.* 12, 13031. doi:10.3390/app122413031

Li, Y., Pan, J., and Wang, J. (2025). A hybrid framework for offshore wind power forecasting: integrating CNN-BiGRU-XGBoost with advanced feature engineering and analysis. *Energies* 18, 5153. doi:10.3390/en18195153

Liu, C., Li, J., and Wang, H. (2023). Predicting wind turbine power output based on XGBoost. In: International conference on 6GN for future wireless networks. New York, NY: Springer, 315–330.

Lv, S.-X., and Wang, L. (2023). Multivariate wind speed forecasting based on multi-objective feature selection approach and hybrid deep learning model. *Energy* 263, 126100. doi:10.1016/j.energy.2022.126100

Miao, C., Li, H., Wang, X., and Li, H. (2021). Ultra-short-term prediction of wind power based on sample similarity analysis. *IEEE Access* 9, 72730–72742. doi:10.1109/access.2021.3080140

Monjoly, S., and Blonbou, R. (2012). "Configuration of a wind power forecasting model based on fuzzy c-means clustering," in *Power and Energy System and Application*. Las Vegas: IASTED International Conference, 12–14.

National Renewable Energy (2023). *Wind resource database (WRDB) and WIND toolkit longterm ensemble dataset (WTK-LED)*. Golden, CO: National Renewable Energy Laboratory (NREL).

Phan, Q. T., Wu, Y. K., and Phan, Q. D. (2021). A hybrid wind power forecasting model with XGBoost, data preprocessing considering different NWPs. *Appl. Sci.* 11, 1100. doi:10.3390/app11031100

Ru, Y., Zhao, Y., and Liao, H. (2024). Ultra-short-term wind power forecasting based on deep embedded clustering and LSTM. In: 2024 9th Asia conference on power and electrical engineering (ACPEE), 2722–2726.

Wadood, H., Haris, M., Hassan, A., Malik, M. O., Yousaf, H., and Ullah, K. (2024). Deep learning applications for wind energy forecasting in smart grids. In: 2024 international conference on engineering and emerging technologies (ICEET), 1–6.

Wang, Y., Wang, J., Cao, M., Li, W., Yuan, L., and Wang, N. (2022). Prediction method of wind farm power generation capacity based on feature clustering and correlation analysis. *Electr. Power Syst. Res.* 212, 108634. doi:10.1016/j.epsr.2022.108634

Xing, Z., and He, Y. (2023). Multi-modal multi-step wind power forecasting based on stacking deep learning model. *Renew. Energy* 215, 118991. doi:10.1016/j.renene.2023.118991

Xu, H., Zhen, Z., and Wang, F. (2022). NWP feature selection and GCN-based ultra-short-term wind farm cluster power forecasting method. In: 2022 IEEE industry applications society annual meeting (IAS), 1–22.

Yaqiong, L., Chaobo, D., Tongxun, W., Zhou, Z., Shengjun, Z., Linhai, C., et al. (2016). Separate wind power and ramp predictions based on meteorological variables and clustering method. In: 2016 IEEE 6th international conference on power systems (ICPS), 1–6.

Zhang, Y., Li, Y., and Zhang, G. (2020). Short-term wind power forecasting approach based on Seq2Seq model using NWP data. *Energy* 213, 118371. doi:10.1016/j.energy.2020.118371

Zhao, X., Liu, H. P., Jin, H., Cao, S., and Tang, G. (2024). An improved hybrid model for wind power forecasting through fusion of deep learning and adaptive online learning. *Comput. Electr. Eng.* 120, 109768. doi:10.1016/j.compeleceng.2024.109768