forecasting

MDPI

Article

Look-Ahead Bias in Wind Speed Forecasting: A Strictly Causal Evaluation of VMD-Based Hybrid Ensembles

Ali Çetinkaya [1], Ercan Aykut [2,*] and Bahtiyar Dursun [1]

[1] Edunetsoft Educational Technology and Software, 34672 Istanbul, Türkiye; bahtiyar@edunetsoft.com (B.D.)
[2] Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, 34310 Istanbul, Türkiye
* Correspondence: eaykut@gelisim.edu.tr

Highlights

What are the main findings?

- Controlled ablation shows that causal rolling-window Variational Mode Decomposition (VMD) changes Extra Trees root mean square error (RMSE) from 1.1940 to 1.1902 m/s, whereas invalid full-series VMD produces an artificially optimistic RMSE of 0.7551 m/s.
- The proposed regime-dependent ensemble achieves the best single-test metrics (coefficient of determination, $R^2 = 0.7875$; RMSE = 1.1682 m/s; mean absolute error, MAE = 0.8877 m/s) but only marginally improves on global constrained combination.

What are the implications of the main findings?

- Full-series decomposition can create large apparent gains that disappear under forecast-origin-specific causal processing; controlled no-decomposition and causal-decomposition baselines are therefore essential.
- Persistence, equal-weight and global ensemble baselines, uncertainty intervals, corrected statistical tests, and purged walk-forward validation are necessary for credible comparisons.

Abstract

Decomposition-ensemble pipelines are widely used in short-term wind-speed forecasting, but global preprocessing can expose future information. This study quantifies that risk using 13,281 hourly observations from Ayvalık, with Bandırma and Balıkesir as spatial auxiliaries, to predict mean wind speed over the next three hours. Chronological splitting precedes all data-dependent operations; target windows crossing partition boundaries are purged; imputation is past-only; scaling is training-only; and Variational Mode Decomposition (VMD) is recomputed from a 168 h historical window. Eleven learners are compared with persistence, equal-weight and global constrained ensembles, deep out-of-fold residual correction, and regime-dependent fuzzy constrained weighting. On 1961 common test origins, the proposed ensemble achieved $R^2 = 0.7875$, RMSE = 1.1682 m/s, MAE = 0.8877 m/s, and 28.6% RMSE skill over persistence. Its improvement over the global constrained combination was marginal and not statistically significant after Holm correction. Controlled ablation showed nearly unchanged performance without decomposition ($R^2 = 0.7780$) and causal VMD (0.7794), while invalid full-series VMD reached 0.9112. Thus, large decomposition gains can primarily reflect look-ahead bias. The principal contribution is a reproducible audit framework for causal decomposition, ensemble calibration, and uncertainty-aware evaluation rather than a claim of universal model superiority.

Check for updates

Academic Editors: José-Santos López-Gutiérrez, Ted Soubdhan and Cong Feng

Received: 5 October 2026
Revised: 12 November 2026
Accepted: 23 November 2026
Published: 26 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: wind speed forecasting; look-ahead bias; causal forecasting; Variational Mode Decomposition

1. Introduction

Rising energy demand and the environmental burden of fossil fuels have accelerated the transition toward renewable generation. Wind energy combines low operating emissions with broad geographic applicability, but its stochastic and seasonally varying nature complicates grid integration, market scheduling, reserve allocation, and turbine operation. Accurate short-term wind speed forecasts are therefore important not only as predictive outputs, but also as inputs to operational decisions in power systems [1,2].

Wind forecasting methods include physical, statistical, machine-learning and hybrid families [3,4]. Recent work continues to combine signal decomposition with deep learners, fuzzy regimes and heterogeneous ensembles [5]. Yet forecast evaluation can be invalidated when smoothing, normalization or decomposition is estimated on the complete series before chronological partitioning [6]. Global EMD-family preprocessing has been shown to expose test information, and sliding or progressive decompositions have been proposed as remedies [7]. Wind-specific evidence likewise shows that overall decomposition can create future-data leakage and that rolling, stepwise and sliding-window strategies yield more realistic results [8].

This study therefore treats causality as a structural requirement. Every feature supplied at forecast origin t must be computable from observations available at or before t. Bidirectional recurrence is permitted only inside the fully observed 24 h input window; no timestamp after t enters either deep architecture. The evaluation asks how much of the apparent benefit of VMD and regime-dependent combination survives when decomposition provenance, forecast origins, learners and tuning budgets are held constant.

The research gap is not the absence of another VMD–deep-learning combination. Recent studies already identify decomposition-induced leakage [7,8] and show that locally decomposed inputs can retain forecasting and interpretability benefits [9]. What remains insufficiently documented is the magnitude of the bias in a multi-station wind-speed setting under a common prediction file, alongside no-VMD and deliberately leaky controls, persistence and ensemble baselines, purged three-hour targets, regime sensitivity, uncertainty intervals and full walk-forward retraining. This work addresses that gap and limits its claims to one target station and evaluation period.

2. Related Work

Physical methods feed atmospheric variables from numerical weather prediction (NWP) models into medium- and long-term forecasts. Computational cost is their main drawback. Coupling a Kalman filter to NWP output sharpened short-term accuracy [10]. Refining NWP parameterizations over the complex terrain of the northwestern United States cut winter mean absolute error by 20–30% [11]. Hybrid autoregressive integrated moving average (ARIMA)–neural-network models jointly represent linear and nonlinear dynamics [12]. Classical ARIMA models and Markov chains run cheaply yet represent nonlinear wind structure poorly, which pushed the field toward data-driven alternatives. Adaptive neuro-fuzzy inference systems (ANFISs) and neural networks outperformed competing data mining techniques on the same problem [2].

Turkish applications supply useful context. Geographic information system (GIS)-based wind speed maps of Balıkesir province quantified the regional resource [13]. Four recurrent neural network (RNN) architectures tested at Yalova surpassed classical time

series methods [14]. Internationally, a Cross-Industry Standard Process for Data Mining (CRISP-DM) comparison of artificial neural networks (ANNs), support vector regression (SVR), regression trees and random forest (RF) identified a well-tuned ANN as most accurate [15], while deep neural networks had earlier proved competitive for wind and solar alike [16]. Nine supervised learners, including a convolutional neural network (CNN), enriched by K-means clustering, formed an early clustering-assisted hybrid and remain the closest structural antecedent to the design examined below [17], with a comparable benchmarking exercise following shortly [18].

Field data show no architecture dominating. A Decision Tree Regressor ranked first among nine models on Denizli wind farm records ($R^2 = 0.998$) [19]. An ANN achieved the lowest MAE for microgrid operation [20], whereas bidirectional long short-term memory (BiLSTM) led a field of fourteen ($R^2 = 0.986$) [21]. A parallel thread pairs meta-heuristic optimization with deep learning: the iHow algorithm fused with a multi-scale attention network on French grid data [22], and a genetic algorithm (GA) ensemble improved minute-resolution forecasts across five Brazilian cities, where purely additive error-correction hybrids usually fell short of the single models they were meant to improve [23]. Tree-based gradient boosting combined with probabilistic weather ensembles delivered up to 53% MAE improvement over power-curve methods offshore [24]. Reviews converge on preprocessing and ensemble strategy as the decisive levers across 91 surveyed studies [4]; decomposition-ensemble frameworks emerge as the most promising direction in two recent surveys [3,25], and deep architectures receive dedicated treatment elsewhere [26].

Two studies are direct precursors of the weighting scheme adopted here. A Dynamic Neuro-Fuzzy Model embeds a small recurrent network in the consequent part of a Fuzzy C-Means rule base, matching or exceeding classical machine learning and deep learning while using far fewer parameters [27]. A later framework dispenses with rule definition altogether: Fuzzy C-Means softly clusters the input space, a separate Random Forest trains inside each cluster, and membership-weighted combination yields the prediction; across 13 real-world series, it competed successfully against classical statistics, fuzzy regression and a plain Random Forest [28]. Both support the per-regime non-negative least squares (NNLS) weighting used below. Supervisory control and data acquisition (SCADA)-based condition monitoring and digital twin integration situate forecasting inside the broader wind energy life cycle [29]. New model families keep arriving. Bayesian additive regression trees (BART), Gaussian process regression (GPR) and quantile regression forest (QRF) were compared on 21 years of daily records from 11 Australian stations placed regional QRF ahead for point prediction, although the reported R^2 values (0.15–0.35) sit far below those of an hourly, multi-station, feature-rich configuration [30]. An LSTM-CNN hybrid over Corsica merged ARPEGE/AROME output with surface observations and cut RMSE by 28–49% against raw NWP [31]. That design is the literature counterpart of the multi-station input used below. A classical-quantum hybrid network reached MAE = 0.1285 [32], and maximal overlap discrete wavelet transform (MODWT) denoising with a shape-aware loss inside a temporal convolutional network–bidirectional gated recurrent unit–attention (TCN–BiGRU–Attention) architecture addressed the multistep case [33]. Mode count and penalty factor bear directly on decomposition quality [34], so the VMD values held fixed here ($K = 4$, $\alpha = 2000$) merit later optimization. Table 1 lists the 19 studies methodologically closest to the design examined below: those employing signal decomposition, ensemble or hybrid modeling, fuzzy clustering, or multi-model comparison, and reporting concrete R^2/RMSE/MAE values. The final row summarizes the results obtained here.

Table 1. Summary of selected studies from the related literature.

Study [Ref.]	Method(s)	Application Domain	Performance Metric/Result	Dataset/Scale
[15]	ANN, SVR, Regression Tree, RF (CRISP-DM)	Wind energy generation forecasting	ANN most accurate; RF advantageous in training time	Wind turbine data
[14]	Four different Recurrent Neural Networks (RNN)	Yalova, 1 h ahead wind speed	RNN models outperform classical time series methods	Single station, Türkiye
[19]	DT, XGBoost, RF, KNN, etc. (9 models)	Denizli wind farm (real data)	Decision Tree Regressor $R^2 = 0.998$; RMSE = 0.151	1 year, minute resolution
[17]	K-NN, SVR, RF, XGB, Ridge, LASSO, DT, CNN (9 models) + K-means clustering	Wind power generation forecasting	Clustering-assisted tree/ensemble models most successful	Real wind turbine data
[22]	iHow optimization + Multi-Scale Attention Network (MSAN)	France, wind/solar forecasting	MSE = 1.1×10^{-6} (optimized; superior to HHO/GWO/PSO/JAYA)	French national grid data
[20]	RF, LSTM, ANN	Microgrid short-term forecasting	ANN lowest MAE (64.72 kW); RF sensitive to scaling errors	1 year of microgrid data
[21]	RF, DT, KNN, AdaBoost, XGB, CNN, RNN, TCN, LSTM, BiLSTM (14 models)	Tamil Nadu, 5 stations	BiLSTM most successful, $R^2 = 0.986$ (test MSE = 0.0093)	~12,900 records, 5 stations
[23]	GA-based ensemble (AR, ARMA, MLP, RBF, ESN, hybrid error correction)	Five Brazilian cities, minute-level wind speed	GA-based ensemble lowest MSE/MAE, highest R^2 (range 0.71–0.87)	Five cities, minute-resolution data (2019)
[35]	RNN, GRU, LSTM, BiLSTM, CNN (timestamp-based multivariate DL)	İzmir, wind speed and GHI forecasting	$R^2 = 0.9987$ for 3-day-ahead wind speed; RMSE = 0.067 m/s	15 min resolution, 3 years
[24]	Tree-based ML (Gradient Boosting) + probabilistic weather forecast ensemble	Belgian offshore wind farms	Up to 53% MAE improvement over the power curve	4 years of SCADA + NWP ensemble data
[27]	Dynamic Neuro-Fuzzy Model (FCM rule base + recurrent network consequent)	Wind power forecasting (single input)	Competitive accuracy against deep learning and classical ML models with far fewer parameters	Two real datasets of differing volume

Table 1. *Cont.*

Study [Ref.]	Method(s)	Application Domain	Performance Metric/Result	Dataset/Scale
[28]	Fuzzy C-Means clustering + regime-based Random Forest ensemble (membership-weighted combination)	13 real-world time series (finance/industry/environment)	Outperforms or matches classical statistical methods, fuzzy regression and plain Random Forest	13 different time series
[30]	BART, GPR, QRF (probabilistic, at-site vs. regional)	Australia, wind speed at 11 stations	Regional QRF lowest RMSE; at-site GPR $R^2 = 0.35$ (daily data, weak feature set)	21 years of daily data, 11 stations
[31]	NWP (ARPEGE/AROME) + station data hybrid LSTM-CNN	Corsica, 1–6 h ahead wind speed	28–49% RMSE improvement over raw NWP; a further 5–15% gain from fine-tuning	278 stations, 2016–2018 (MeteoNet)
[32]	Classical-quantum hybrid neural network (FNN + Quantum Neural Network)	Multi-step wind speed forecasting	Best model MAE = 0.1285, RMSE = 0.2550, MAPE = 8.20%	Publicly available wind speed data
[6]	Forecast-evaluation audit	General time-series forecasting	Identifies whole-series decomposition and scaling as leakage risks	Methodological review and experiments
[7]	Sliding/progressive EMD-family decomposition	Water-quality forecasting	Demonstrates information leakage and causal remedies	Controlled ablation
[8]	Rolling/stepwise/sliding EMD, VMD and DWT	Wind-speed forecasting	Shows global preprocessing can create spurious precision	Multiple wind datasets
[9]	Local model-agnostic decomposition	General time-series forecasting	Separates local causal from global leaky decomposition	Multiple datasets and models
Present work	Rolling causal VMD + 11 learners + persistence/equal/global/FCM ensembles	Ayvalık 3 h mean wind-speed forecasting	FCM: $R^2 = 0.7875$, RMSE = 1.1682; no VMD $R^2 = 0.7780$; invalid global VMD $R^2 = 0.9112$	13,281 h; 1961 common test origins

Directly related leakage studies change how decomposition results must be interpreted. Hewamalage et al. [6] identify whole-series decomposition and normalization as indirect leakage channels in forecast evaluation. Yu et al. [7] demonstrate the problem for EMD-family models and compare sliding and progressive remedies. Pang and Dong [8] focus specifically on wind-speed forecasting and show that global EMD/VMD preprocessing can create spurious precision, while causal rolling alternatives face boundary effects. Kreuzer et al. [9] distinguish global from local decomposition and restrict decomposition to each forecasting instance. These findings mean that high cross-study R^2 values cannot establish leakage in earlier work; leakage is a property of the protocol, not of the reported score. Accordingly, the present contribution is a controlled within-dataset comparison rather than an inference from cross-study accuracy differences.

The selected studies and their methodological context are summarized in Table 1.

3. Materials and Methods

The experimental design is organized around one central question: how do VMD-based learners and hybrid combination strategies behave when every preprocessing and evaluation step is constrained to information available at the forecast origin? Four research questions follow:

- How accurately do raw meteorological variables, lagged and rolling statistics, neighboring-station observations, and causally generated VMD components forecast mean wind speed over the next three hours?
- How much of the apparent contribution of VMD remains after look-ahead bias is removed?
- Do deep residual correction or regime-space fuzzy weighting improve on strong individual learners and a naive persistence benchmark across a single test period and multiple seasons?
- Which causal-design, statistical-testing, and reporting practices are necessary for credible comparison of decomposition–ensemble forecasting systems?

3.1. Computational Environment

All analyses were conducted in Google Colaboratory (https://colab.research.google.com/; accessed on 22 November 2026) using Python 3.13.16 (main, 1 October 2026, 15:40:24; GCC 13.3.0). The principal versioned libraries were PyTorch 2.11.0+cu130, TensorFlow 2.21.0, OpenCV 4.14.0, pandas 2.2.3, and NumPy 2.1.3. Additional software packages without archived version metadata were scikit-learn (https://scikit-learn.org/), XGBoost (https://xgboost.readthedocs.io/), LightGBM (https://lightgbm.readthedocs.io/), CatBoost (https://catboost.ai/), SciPy (https://scipy.org/), statsmodels (https://www.statsmodels.org/), scikit-fuzzy (https://github.com/scikit-fuzzy/scikit-fuzzy), and vmdpy (https://pypi.org/project/vmdpy/); all websites were accessed on 22 November 2026. Deep-learning models were trained using one NVIDIA L4 GPU (NVIDIA Corporation, Santa Clara, CA, USA). Random seeds and principal model hyperparameters are reported in the manuscript and the Supplementary Materials.

3.2. Dataset, Data Preprocessing, and Feature Engineering

The core record contains hourly wind speed and direction from Ayvalık. Bandırma and Balıkesir Meydan provide synchronous auxiliary observations. After timestamp validation and hourly reindexing, the common grid contains 13,281 h. Hours and months are represented by sine/cosine pairs. Wind direction, recorded as ten-degree sectors, is causally filled from past values and then transformed to sine/cosine components. Lagged and rolling features are shifted so that they use only observations available by the forecast origin.

- Sensor codes 99 and 999 are converted to missing values before causal completion. The statistical 0.999-quantile threshold is fitted on the training portion only and applied only there; validation and test observations remain unchanged.
- Missing values are filled without backward fill or two-sided interpolation. Forward fill is limited to 24 h; any remaining gap uses an expanding mean computed from observations available up to the current timestamp.
- The chronological 70/15/15 split precedes scaling and model fitting. Because the target uses t + 1 through t + 3, the last three candidate origins before the training and validation cutoffs are purged. Scaling is fitted on the training partition only.
- The raw file contains 40,367 station rows. The modeled hourly grid spans 1 January 2007 to 7 July 2008. After the 168 h VMD warm-up, 24 h lookback and three-hour boundary purge, the exact usable forecast-origin periods are 9 January 2007 22:00–25 January 2008 14:00 for training ($n = 9137$), 25 January 2008 18:00–16 April 2008 09:00 for validation/calibration ($n = 1960$), and 16 April 2008 13:00–7 July 2008 05:00 for testing ($n = 1961$). Each target window t + 1:t + 3 remains entirely inside its own partition.

Table 2 documents the executable pipeline, and Table 3 specifies every input, spatial auxiliary, target and temporal structure. Figure 1 illustrates the relation between the 24 h input window, forecast origin and mean target over the following three hours.

Table 2. Data preprocessing steps, VMD, and model parameters.

Stage/Parameter	Description/Value
Data source	Hourly wind speed/direction from Ayvalık, Bandırma and Balıkesir Meydan; 13,281 h common grid
Missing values	Codes 99/999 → NaN; forward fill limited to 24 h, then past-only expanding mean; no backward fill
Training outliers	0.999 quantile fitted on training only (15.0 m/s); eight training values replaced and causally refilled
Base features	17 variables: current/lagged speeds, shifted rolling statistics, cyclical hour/month, direction sine/cosine and spatial auxiliaries
VMD	Rolling historical window W = 168; K = 4, α = 2000, τ = 0, DC = 0, init = 1, tolerance = 10^{-7}; four last-point mode features
Total features	21 per hour (17 base + 4 VMD); 24 h sequence; 504 flattened inputs for tabular learners
Target	Mean of observed/imputed Ayvalık speed at t + 1, t + 2 and t + 3
Split and purge	Nominal 70/15/15; t + 1:t + 3 must remain inside its partition; usable origins 9137/1960/1961
Scaling	MinMaxScaler fitted on training only
CNN-GRU-Attention	Causal Conv1D(32,k = 3), MaxPool, GRU(64), Attention, GlobalAveragePool, Dense(16), output
Causal CNN + BiLSTM	Causal Conv1D(64,k = 3), BatchNorm, MaxPool, BiLSTM(64), Dropout 0.3, BiLSTM(32), Dense(32), output
Linear SVR	Linear kernel; C = 1.0; ε = 0.01; max_iter = 10,000
DE-XGBoost	Differential evolution; best1bin; maxiter = 10; popsize = 5; seed = 42
Residual network	Conv1D-GRU trained on chronological out-of-fold Extra Trees residuals
FCM ensemble	c = 3, m = 2; weights constrained to w ≥ 0 and Σw = 1; validation-only calibration
Uncertainty/testing	24 h moving-block bootstrap (2000); DM-HLN h = 3 with Holm correction

Table 3. Input variables, target output and temporal structure.

Variable	Code	Unit	Role	Temporal Structure
Ayvalik wind speed	rzghiz	m/s	Input and target source	t and lags 1, 3; rolling 6 h mean; rolling 24 h SD
Ayvalik wind direction	rzgyon	10-degree sector	Input	Sine/cosine at t
Bandirma wind speed	bandirma_hiz	m/s	Spatial auxiliary input	t and lag 1
Bandirma wind direction	bandirma_yon	10-degree sector	Spatial auxiliary input	Sine/cosine at t
Balikesir wind speed	balikesir_hiz	m/s	Spatial auxiliary input	t and lag 1
Calendar hour/month	saat, ay	index	Input	Cyclic sine/cosine
Causal VMD modes	vmd_imf_0–vmd_imf_3	scaled amplitude	Input	Last point from $[t - W + 1, t]$, $W = 168$
Future mean wind speed	target_3 h	m/s	Target	$\text{mean}(y[t + 1], y[t + 2], y[t + 3])$
Persistence	persist_3 h	m/s	Baseline output	$\text{mean}(y[t - 2], y[t - 1], y[t])$

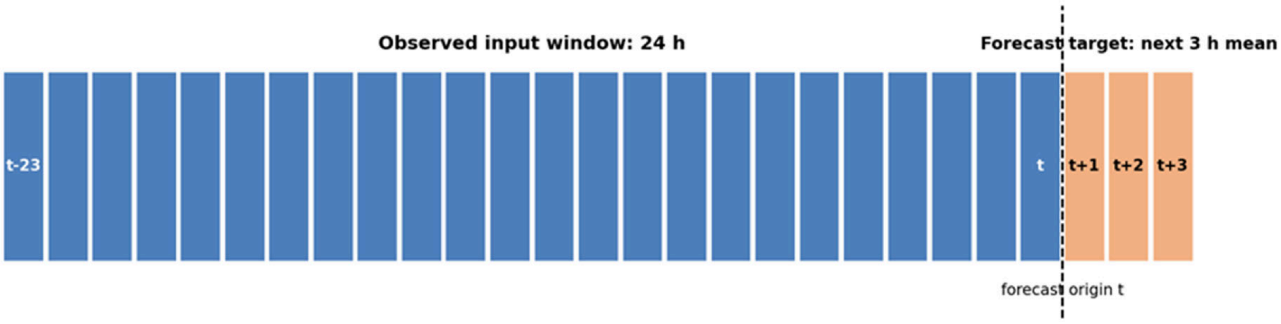


Figure 1. Sliding 24 h input window and three-hour mean forecast target.

3.3. Artificial Intelligence Model of the Developed System

Eleven base learners and two principal combination strategies are evaluated against persistence, equal-weight and global constrained least-squares ensembles. The convolutional neural network–gated recurrent unit with attention (CNN–GRU–Attention) architecture was selected as a compact convolutional-recurrent benchmark that combines local pattern extraction, gated memory and within-window attention. Causal padding ensures that each convolutional output uses only its current and earlier positions. Causal CNN + BiLSTM was retained as a deliberately stronger sequence benchmark: although the BiLSTM traverses the observed input window in both directions, every element of that window lies at or before forecast origin t, so the pipeline remains temporally causal. The model does not process validation/test targets or timestamps after t. DE-XGBoost uses differential evolution, not a genetic algorithm, to optimize learning rate, depth, subsampling and column sampling.

The residual-correction network uses chronological out-of-fold Extra Trees residuals rather than in-sample residuals. Regime-dependent combination applies Fuzzy C-Means (FCM) to three validation descriptors: wind speed at t, the past 6 h mean and the past 24 h standard deviation. For regime r, weights solve a constrained least-squares problem with $w_j \geq 0$ and $\Sigma_j w_j = 1$. Test forecasts are membership-weighted mixtures of the regime-

specific convex combinations. The same validation predictions calibrate equal-weight, global constrained and FCM ensembles; test data are never used to estimate weights.

The target, causal convolution, fuzzy partitioning, regime-specific constrained calibration and final membership-weighted forecast are defined as follows:

The three-hour mean target is

$$\bar{y}_t^{(3)} = \frac{1}{3}\sum_{h=1}^{3} y_{t+h} \tag{1}$$

For a causal convolutional layer, only current and earlier positions enter the output:

$$z_t^{(\ell)} = \sigma\left(b^{(\ell)} + \sum_{k=0}^{K-1} W_k^{(\ell)}\, z_{t-k}^{(\ell-1)}\right) \tag{2}$$

FCM estimates memberships and regime centroids by minimizing

$$J_m = \sum_{i=1}^{N}\sum_{r=1}^{c} u_{ir}^{m} \parallel x_i - v_r \parallel_2^2, \qquad u_{ir} = \left[\sum_{q=1}^{c}\left(\frac{\parallel x_i - v_r \parallel_2}{\parallel x_i - v_q \parallel_2}\right)^{\frac{2}{m-1}}\right]^{-1} \tag{3}$$

For each regime r, validation-only non-negative least squares (NNLS) weights solve

$$\mathbf{w}_r = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i\in\ell)} u_{ir}^{m}\left(y_i - \sum_{j=1}^{M} w_j \hat{y}_{ij}\right)^2, \qquad w_j \geq 0, \quad \sum_{j=1}^{M} w_j = \tag{4}$$

The final regime-dependent forecast is

$$\hat{y}_i^{\mathrm{FCM}} = \sum_{r=1}^{c} \tilde{u}_{ir} \sum_{j=1}^{M} w_{rj}\hat{y}_{ij}, \qquad \tilde{u}_{ir} = \frac{u_{ir}^{m}}{\sum_{q=1}^{c} u_{iq}^{m}} \tag{5}$$

The main configuration uses c = 3 and fuzzifier m = 2. The validation memberships contain 704, 769 and 487 dominant assignments. Sensitivity analysis over c ∈ {2,3,4} and m ∈ {1.5,2.0,2.5} yields test R^2 values from 0.7872 to 0.7889, indicating that the reported behavior is not driven by a single clustering setting. Because FCM labels are permutation-invariant, the revised table reports Regimes 0–2 rather than attaching unverified meteorological names to cluster indices.

The validation-calibrated regime-specific convex weights are reported in Table 4.

Table 4. Validation-calibrated convex weights by FCM regime (c = 3, m = 2).

Model	Regime 0 (n = 704)	Regime 1 (n = 769)	Regime 2 (n = 487)
Classical ANN	0.0697	0.1427	0.1219
CNN-GRU-Attention	0.0426	0.0353	0.2052
Causal CNN + BiLSTM	0.0406	0.0087	0.0000
Gradient Boosting	0.0000	0.0000	0.0000
Random Forest	0.1456	0.0000	0.0000
Extra Trees	0.1870	0.3000	0.4576
XGBoost	0.0000	0.0000	0.0000
DE-XGBoost	0.3421	0.0762	0.0000
LightGBM	0.0000	0.0038	0.0304
CatBoost	0.1724	0.3568	0.1849
Linear SVR	0.0000	0.0764	0.0000

3.4. System Flowchart

Figure 2 traces the complete evaluation workflow, including all eleven base learners, persistence, equal-weight and global constrained baselines, the FCM ensemble and out-of-fold residual correction. The split, causal VMD, sequence construction, calibration and test steps are shown in their execution order.

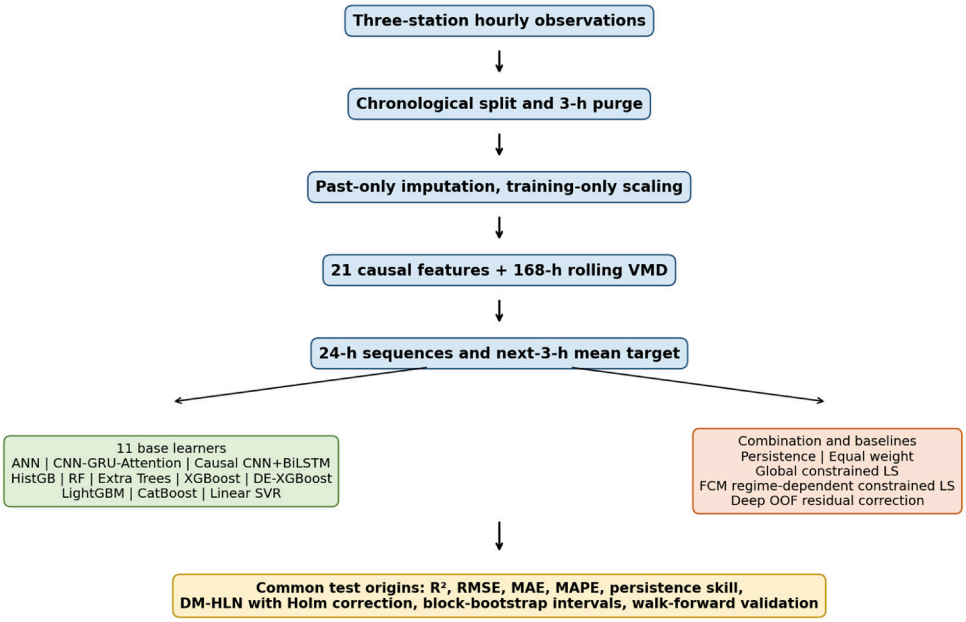


Figure 2. Strictly causal workflow and all evaluated model candidates.

The main setting ($c = 3$, $m = 2$) was retained a priori for continuity and interpretability; no alternative was selected from test performance. The narrow R^2 range across all nine settings supports parameter robustness but does not establish equivalence.

4. Experimental Design and Performance Metrics

The target is the arithmetic mean of wind speed at $t + 1$, $t + 2$ and $t + 3$, not the value exactly three hours ahead. All deep models use Adam with learning rate 0.001, batch size 64 and early stopping. The CNN–GRU–Attention architecture consists of causal Conv1D(32, kernel 3), max pooling, GRU(64), self-attention, global average pooling, Dense(16) and a scalar output. Causal CNN + BiLSTM uses causal Conv1D(64, kernel 3), batch normalization, max pooling, BiLSTM(64), dropout 0.3, BiLSTM(32), dropout 0.3, Dense(32) and a scalar output.

Performance is reported by R^2, RMSE, MAE, stabilized mean absolute percentage error (MAPE) and RMSE skill relative to persistence. All methods use the 1961 common test origins stored in one prediction file. Diebold–Mariano tests use squared-error loss, horizon $h = 3$, the Harvey–Leybourne–Newbold (DM–HLN) correction and Holm adjustment. Moving-block bootstrap intervals use 24 h blocks and 2000 repetitions. Non-significance is described only as failure to detect a difference, not as evidence of equivalence.

Table 5 reports the authoritative metrics recomputed from the unified prediction file. The proposed FCM ensemble achieved the highest R^2 (0.7875), lowest RMSE (1.1682 m/s), lowest MAE (0.8877 m/s) and lowest MAPE (24.04%), corresponding to 28.58% RMSE skill over persistence. Global constrained combination was nearly identical ($R^2 = 0.7871$; RMSE = 1.1692 m/s), and equal weighting remained competitive ($R^2 = 0.7844$).

Table 5. Authoritative test-set results from 1961 common forecast origins.

Model	R^2	RMSE (m/s)	MAE (m/s)	MAPE (%)	Skill	Interpretation
Proposed FCM ensemble	0.7875	1.1682	0.8877	24.04	0.286	Proposed regime ensemble
Global constrained least squares	0.7871	1.1692	0.8889	24.05	0.285	Common test origins
Equal weight	0.7844	1.1766	0.8937	24.21	0.281	Common test origins
CatBoost	0.7798	1.1890	0.9071	24.69	0.273	Common test origins
XGBoost	0.7796	1.1896	0.9040	25.03	0.273	Common test origins
Extra Trees	0.7794	1.1902	0.9080	25.02	0.272	Common test origins
Gradient Boosting	0.7773	1.1957	0.9062	25.13	0.269	Common test origins
DE-XGBoost	0.7757	1.2002	0.9146	25.14	0.266	Common test origins
LightGBM	0.7754	1.2008	0.9084	25.19	0.266	Common test origins
Random Forest	0.7706	1.2135	0.9156	25.47	0.258	Common test origins
Linear SVR	0.7622	1.2357	0.9552	27.13	0.244	Common test origins
VMD + Deep Residual Correc- tion (ET + CNN-GRU)	0.7600	1.2414	0.9533	25.77	0.241	Common test origins
Classical ANN	0.7454	1.2785	0.9779	26.65	0.218	Common test origins
CNN-GRU-Attention	0.7216	1.3370	1.0284	26.82	0.183	Common test origins
Causal CNN + BiLSTM	0.6331	1.5348	1.1851	32.20	0.062	Common test origins
Persistence	0.5834	1.6355	1.2635	36.44	0.000	Naive reference

The controlled no-decomposition, causal-VMD and full-series-VMD comparison is reported in Table 6.

Table 6. Controlled decomposition ablation with identical Extra Trees learner and forecast origins.

Preprocessing	Valid Forecasting Design	Train *n*	Test *n*	R^2	RMSE	MAE	MAPE_Percent
No decomposition	True	9137	1961	0.7780	1.1940	0.9077	24.9191
Causal rolling-window VMD	True	9137	1961	0.7794	1.1902	0.9080	25.0198
Non-causal full-series VMD (diagnostic only)	False	9137	1961	0.9112	0.7551	0.5893	16.7284

The equal-weight, global constrained and regime-dependent ensemble baselines are compared in Table 7.

The cluster-count and fuzzifier sensitivity results are summarized in Table 8.

The proposed ensemble significantly outperformed persistence and several individual models after Holm correction, but no significant difference was detected relative to XGBoost, Extra Trees or Gradient Boosting. It also did not differ significantly from equal weighting or global constrained combination. The result therefore supports competitive accuracy, not universal superiority or proof that regime dependence is necessary.

Table 7. Ensemble baselines under the same validation calibration protocol.

Ensemble	R^2	RMSE	MAE	MAPE_Percent
Equal weight	0.7844	1.1766	0.8937	24.2081
Global constrained least squares	0.7871	1.1692	0.8889	24.0549
FCM regime-dependent constrained least squares	0.7875	1.1682	0.8877	24.0382

Table 8. FCM cluster-count and fuzzifier sensitivity analysis.

Clusters	Fuzzifier_m	FPC	R^2	RMSE	MAE	MAPE_Percent
2.0000	1.5000	0.8712	0.7889	1.1643	0.8838	23.9659
2.0000	2.0000	0.7386	0.7879	1.1669	0.8862	23.9965
2.0000	2.5000	0.6505	0.7874	1.1685	0.8877	24.0155
3.0000	1.5000	0.7790	0.7877	1.1674	0.8871	24.0430
3.0000	2.0000	0.5882	0.7875	1.1682	0.8877	24.0382
3.0000	2.5000	0.4780	0.7872	1.1688	0.8883	24.0406
4.0000	1.5000	0.7764	0.7882	1.1662	0.8880	24.0927
4.0000	2.0000	0.5389	0.7883	1.1659	0.8868	24.0404
4.0000	2.5000	0.4019	0.7875	1.1680	0.8881	24.0523

Figure 3 compares the test-set R^2 values of all individual models, ensemble baselines and persistence.

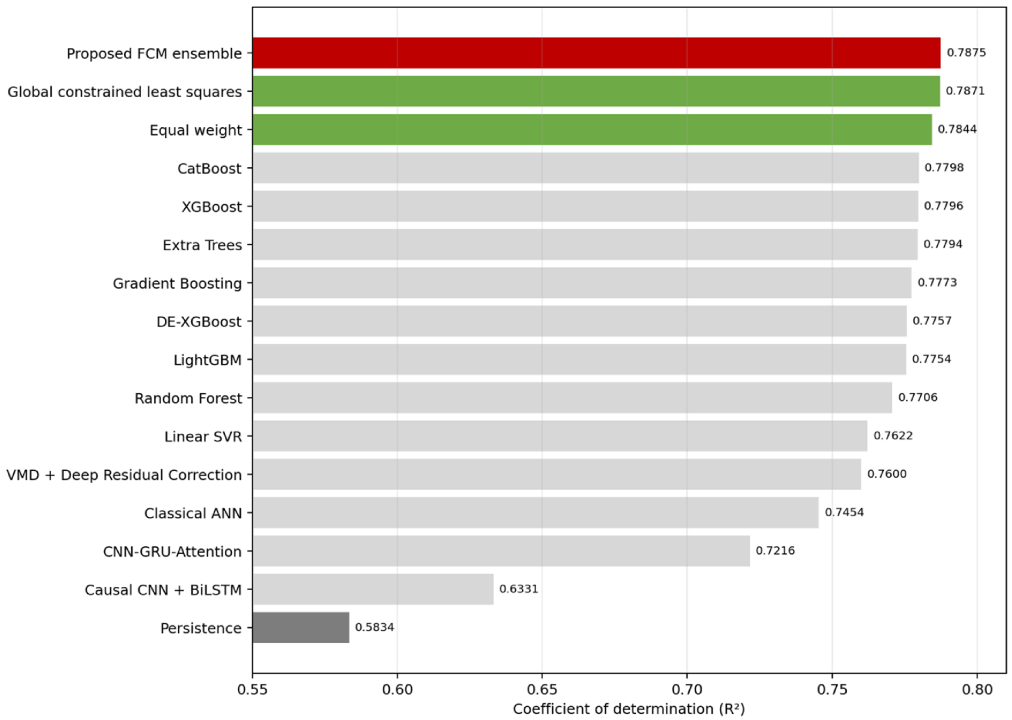


Figure 3. Test-set R^2 for all individual models, ensemble baselines and persistence.

Figure 4 shows the observed and proposed-ensemble wind-speed series over the complete test period and the peak-event window.

Across the full test period, the proposed ensemble achieved $R^2 = 0.7875$, RMSE = 1.1682 m/s and MAE = 0.8877 m/s. In the automatically selected two-week peak window ($n = 337$), R^2 increased to 0.8669 while RMSE and MAE were 1.0213 and 0.8016 m/s. Regression analysis shows systematic underprediction at high observed speeds across all leading models; mechanistic explanations are therefore treated as hypotheses rather than established causes.

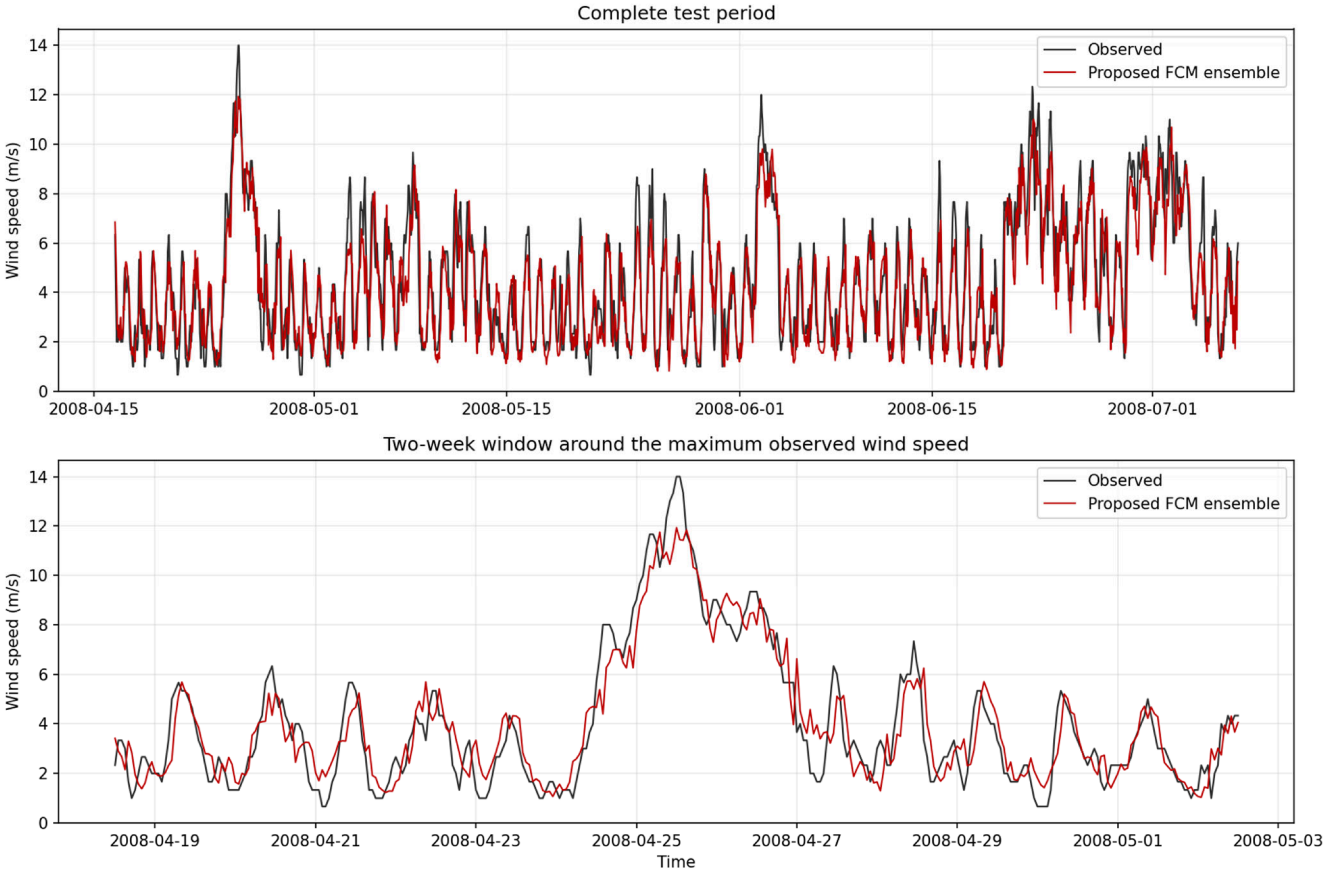


Figure 4. Observed and proposed-ensemble predictions over the full test period and the two-week window centered on the maximum observed wind speed.

4.1. Feature Importance Analysis

Permutation importance was computed twice, for Extra Trees and for Linear SVR. Two learners test whether a ranking merely reflects one inductive bias. Contributions from each time step of the flattened input vector were summed by parent feature.

Figure 5 presents the permutation-importance rankings obtained independently from Extra Trees and Linear SVR.

4.2. Four-Fold Walk-Forward Validation

Four purged walk-forward folds retrain all eleven learners and regime weights from scratch. Each validation and test block contains 867 usable origins after the three-hour boundary purge. These folds evaluate temporal stability, not literal full-season coverage; the revised text therefore avoids calling them four complete seasons.

Figure 6 summarizes the R^2 variation across the four purged walk-forward folds.

Across the four folds, CatBoost achieved the highest mean R^2 (0.7750 ± 0.0421), followed by the proposed FCM ensemble (0.7727 ± 0.0530) and Linear SVR (0.7707 ± 0.0502). The ensemble led Fold 3 at $R^2 = 0.8194$ but did not lead every fold. These results do not

support a claim of general seasonal superiority; they show comparable temporal stability with fold-dependent rankings.

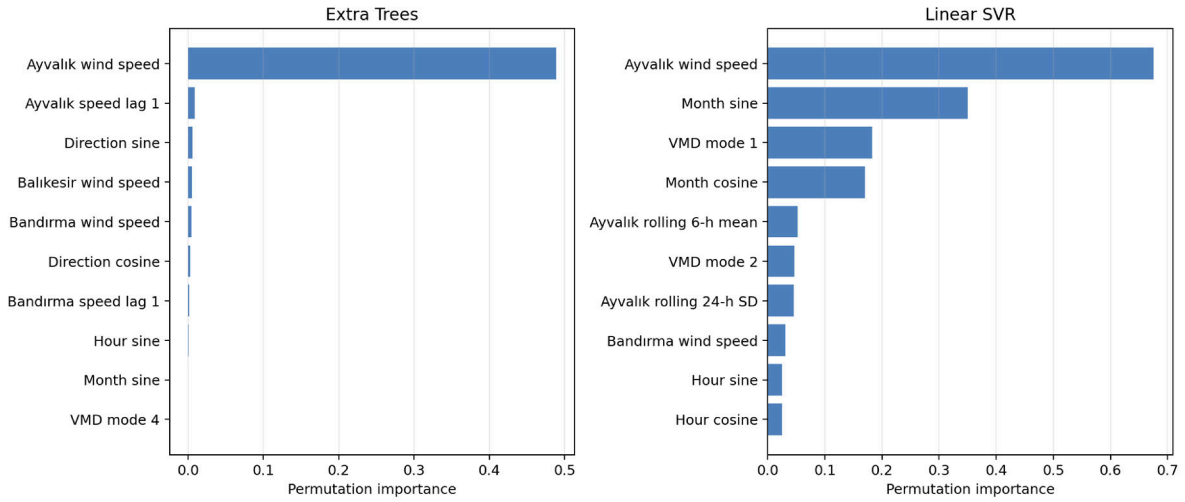


Figure 5. Permutation feature importance for Extra Trees and Linear SVR. Importance measures predictive reliance under permutation and is not interpreted as improvement attributable to VMD.

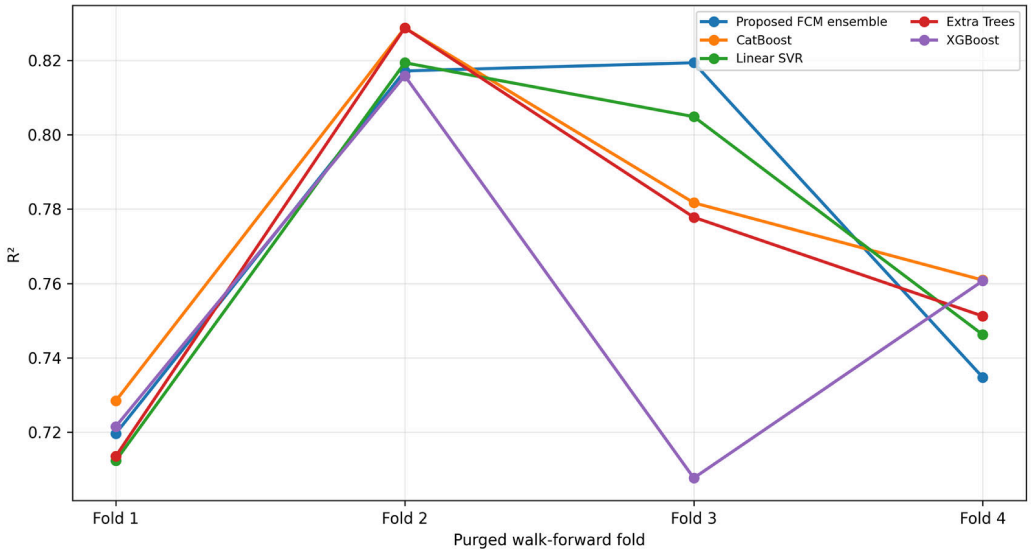


Figure 6. R^2 across four purged walk-forward folds for the proposed ensemble and selected strong individual learners.

Complete fold-level R^2/RMSE/MAE results for all eleven learners and the proposed ensemble are reported in Table 9.

Table 9. Complete purged walk-forward results; each fold cell reports R^2/RMSE/MAE.

Model	Fold 1	Fold 2	Fold 3	Fold 4	Mean R^2
Proposed FCM ensemble	0.7196/1.4894/1.1834	0.8172/1.3017/0.9822	0.8194/0.8290/0.6435	0.7347/1.2242/0.9163	0.7727
Classical ANN	0.2223/2.4807/1.9781	0.7341/1.5702/1.1666	0.6999/1.0686/0.8623	0.7010/1.2997/0.9846	0.5893
CNN-GRU-Attention	0.4431/2.0991/1.6623	0.7920/1.3887/1.0858	0.8204/0.8268/0.6113	0.7360/1.2213/0.9449	0.6979
Causal CNN + BiLSTM	0.5444/1.8986/1.5318	0.6030/1.9186/1.4441	0.7493/0.9768/0.6875	0.5038/1.6741/1.2325	0.6001
Gradient Boosting	0.7188/1.4915/1.1937	0.8063/1.3402/1.0092	0.7454/0.9844/0.7705	0.7526/1.1822/0.9019	0.7558

Table 9. *Cont.*

Model	Fold 1	Fold 2	Fold 3	Fold 4	Mean R^2
Random Forest	0.7089/1.5178/1.1995	0.8104/1.3260/0.9884	0.7768/0.9215/0.7720	0.7429/1.2050/0.9022	0.7598
Extra Trees	0.7135/1.5057/1.2169	0.8288/1.2597/0.9527	0.7778/0.9195/0.7710	0.7512/1.1854/0.9021	0.7678
XGBoost	0.7215/1.4843/1.1879	0.8159/1.3065/0.9894	0.7076/1.0549/0.8791	0.7607/1.1626/0.8829	0.7514
DE-XGBoost	0.6975/1.5471/1.2317	0.8234/1.2795/0.9622	0.7074/1.0552/0.8291	0.7498/1.1889/0.9078	0.7445
LightGBM	0.7250/1.4751/1.1785	0.8109/1.3239/0.9934	0.7217/1.0291/0.8133	0.7545/1.1776/0.8974	0.7530
CatBoost	0.7284/1.4659/1.1724	0.8288/1.2597/0.9486	0.7817/0.9115/0.7331	0.7609/1.1622/0.8883	0.7749
Linear SVR	0.7123/1.5087/1.1511	0.8194/1.2939/0.9859	0.8049/0.8617/0.6787	0.7462/1.1973/0.9195	0.7707

Moving-block bootstrap uncertainty intervals for the proposed ensemble, leading individual models and persistence are summarized in Table 10.

Table 10. Moving-block bootstrap 95% confidence intervals (24 h blocks; 2000 repetitions).

Model	R^2 95% CI	RMSE 95% CI (m/s)	MAE 95% CI (m/s)
Proposed FCM ensemble	0.7386–0.8232	1.0885–1.2421	0.8274–0.9452
CatBoost	0.7294–0.8162	1.1119–1.2612	0.8476–0.9650
XGBoost	0.7256–0.8176	1.1103–1.2677	0.8461–0.9621
Extra Trees	0.7269–0.8161	1.1135–1.2602	0.8472–0.9670
Persistence	0.4695–0.6667	1.5426–1.7155	1.1936–1.3257

4.3. Extreme-Event and Tail Error Analysis

Aggregate regression plots indicate increasing negative bias at the upper end of observed wind speed. The proposed ensemble improves average error but does not remove upper-tail underprediction. Because the experiment was not designed to isolate a physical mechanism, possible explanations—leaf-value averaging, squared-error shrinkage and convex blending—are presented as hypotheses for future controlled testing.

The upper-tail limitation matters for ramp warnings and extreme-event operation. Future work should evaluate quantile, distributional or tail-weighted objectives and report calibration and exceedance metrics. The present point forecasts should not be interpreted as a validated extreme-event warning system.

Figure 7 shows the training and validation loss curves for the three deep-learning architectures.

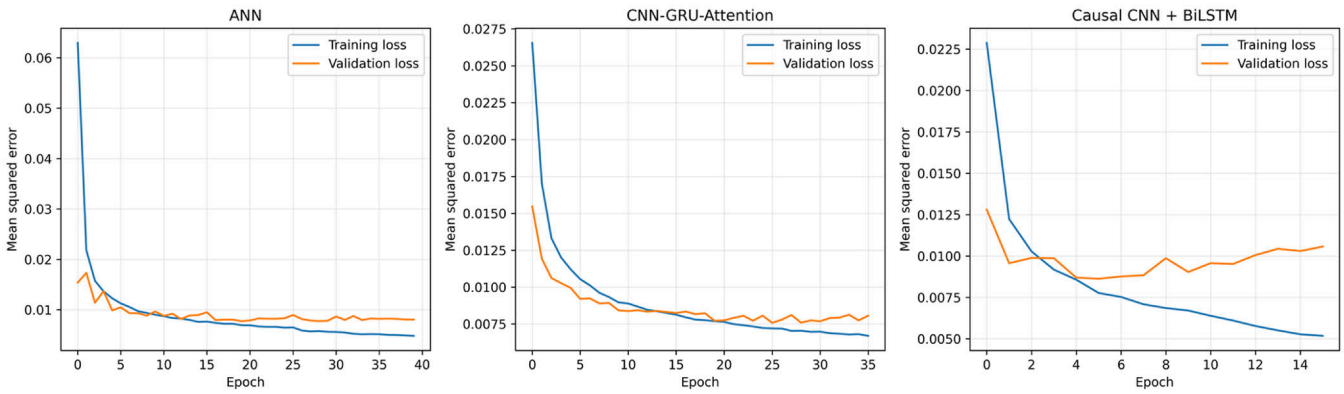


Figure 7. Training and validation loss curves for the three deep architectures.

Figure 8 presents the observed-versus-predicted regression plots for the leading individual models and the proposed ensemble.

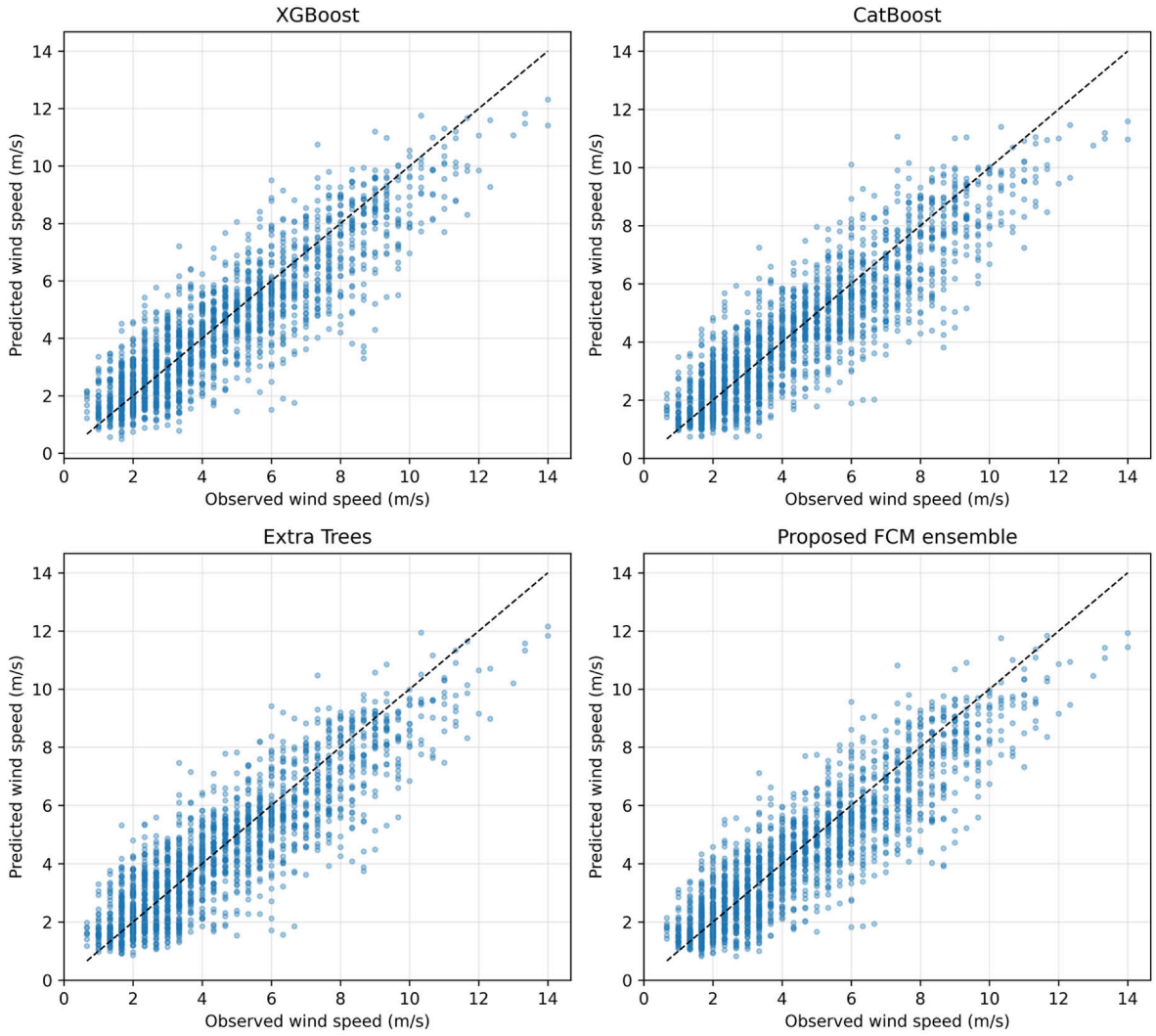


Figure 8. Observed-versus-predicted regression plots for leading individual models and the proposed ensemble. Blue points represent observed–predicted pairs, and the gray dashed line denotes the 1:1 reference line.

5. Discussion

The results show that model ranking depends on dataset, horizon, temporal regime and evaluation protocol. The strongest contribution is methodological: all reported values arise from a single prediction file, target windows are purged at cutoffs, and decomposition provenance is varied while learner and forecast origins remain fixed.

Controlled ablation (Table 6) changes Extra Trees R^2 from 0.7780 without decomposition to 0.7794 with causal rolling-window VMD; RMSE improves by only 0.0038 m/s, while MAE changes from 0.9077 to 0.9080 m/s. In contrast, deliberately invalid full-series VMD reaches $R^2 = 0.9112$ and RMSE = 0.7551 m/s. This large apparent gain is attributable to look-ahead because future observations influence the global modes. Permutation importance is reported separately and is not treated as a measure of predictive improvement.

Table 7 separates the benefit of combining predictors from the benefit of regime-dependent weights. Equal weighting reaches $R^2 = 0.7844$, global constrained least squares 0.7871 and FCM weighting 0.7875. The FCM-over-global RMSE difference is 0.0010 m/s and is not significant (Holm-adjusted $p = 1.000$). Regime conditioning is therefore a

marginal refinement in this dataset, whereas predictor combination supplies most of the ensemble gain.

6. Conclusions and Recommendations

This study shows that causal pipeline design is more consequential than added architectural complexity. On the common test set, the proposed ensemble achieved the best point metrics but only marginally exceeded global constrained combination. In walk-forward evaluation, CatBoost had the highest mean R^2 and the ensemble ranked second. Deep out-of-fold residual correction did not improve on the strongest tree learners.

The findings support four conclusions:

- Regime-dependent weighting can provide a small improvement, but equal-weight and global constrained baselines show that most of the gain arises from combining predictors rather than from fuzzy regimes alone.
- Added architectural complexity is not evidence of added predictive information; deep models and residual correction must be compared with persistence and strong tabular learners under the same forecast origins.
- Causal VMD provides almost no improvement over no decomposition, while global full-series VMD produces a large invalid gain. Decomposition-based studies should therefore report no-decomposition, causal and deliberately leaky diagnostic controls.
- Statistical non-significance does not establish equivalence. Uncertainty intervals, paired tests, common prediction files and walk-forward folds are required to distinguish detectable improvement from small or unstable ranking differences.

The conclusions are limited to one target station, one three-hour mean target and a 2007–2008 record. The auxiliary Balıkesir series contains substantial missingness, and upper-tail forecasts remain biased downward. Missing-data sensitivity using the same Extra Trees specification is provided in the Supplementary Materials, including exclusion of forecast origins with any imputed target hour. Future work should repeat the protocol on contemporary multi-site data, multiple horizons and probabilistic objectives.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/forecast8060116/s1, File S1, minimal reproducibility package containing the hourly meteorological dataset, executed analysis notebook, unified test predictions, fold-level results, model weights, statistical-test outputs, and figure-generation files.

Author Contributions: Conceptualization, A.Ç. and E.A.; methodology, A.Ç.; software, A.Ç.; validation, E.A. and B.D.; formal analysis, A.Ç.; investigation, A.Ç.; data curation, A.Ç.; writing—original draft preparation, A.Ç.; writing—review and editing, E.A. and B.D.; visualization, A.Ç.; supervision, E.A. and B.D.; project administration, E.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The hourly meteorological dataset, executed analysis notebook, unified test predictions, fold-level results, model weights, statistical-test outputs and figure-generation files are provided as Supplementary Materials with this revision. These files constitute the minimal reproducibility package used to generate every reported result.

Acknowledgments: The authors acknowledge the institutions that provided the meteorological observations used in this study. During the preparation and revision of this manuscript, the authors used OpenAI ChatGPT (GPT-5.6, accessed October 2026) for language editing, manuscript organization and assistance with code review. All computational analyses were executed and independently verified by the authors. The authors reviewed and take full responsibility for the accuracy, integrity and final content of the manuscript. Generative AI was not used to generate, fabricate, modify or replace the empirical data or numerical results.

Conflicts of Interest: Author Ali Çetinkaya and Bahtiyar Dursun are employed by the company Edunetsoft Educational Technology and Software. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Simankov, V.; Buchatskiy, P.; Teploukhov, S.; Onishchenko, S.; Kazak, A.; Chetyrbok, P. Review of estimating and predicting models of the wind energy amount. *Energies* **2023**, *16*, 5926. [CrossRef]
2. Colak, I.; Sagiroglu, S.; Yesilbudak, M. Data mining and wind power prediction: A literature review. *Renew. Energy* **2012**, *46*, 241–247. [CrossRef]
3. Liu, Z.; Guo, H.; Zhang, Y.; Zuo, Z. A comprehensive review of wind power prediction based on machine learning: Models, applications, and challenges. *Energies* **2025**, *18*, 350. [CrossRef]
4. Valdivia-Bautista, S.M.; Domínguez-Navarro, J.A.; Pérez-Cisneros, M.; Vega-Gómez, C.J.; Castillo-Téllez, B. Artificial intelligence in wind speed forecasting: A review. *Energies* **2023**, *16*, 2457. [CrossRef]
5. Arrassi, I.; Arkhouch, S.; El Myasse, I.; Dib, F.; Ben Meziane, K.; Benaya, N. A comprehensive review of AI-based wind power forecasting over multiple time horizons. *Process Integr. Optim. Sustain.* **2026**, *10*, 2933–2955. [CrossRef]
6. Hewamalage, H.; Ackermann, K.; Bergmeir, C. Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Min. Knowl. Discov.* **2023**, *37*, 788–832. [CrossRef] [PubMed]
7. Yu, X.; Li, J.; Jiang, X. Research on information leakage in time series prediction based on empirical mode decomposition. *Sci. Rep.* **2024**, *14*, 28362. [CrossRef] [PubMed]
8. Pang, J.; Dong, S. Stop unrealistic data preprocessing in wind speed forecasting: Approaches and discussions on preventing future data leakage. *Appl. Math. Model.* **2026**, *150*, 116376. [CrossRef]
9. Kreuzer, T.; Zdravkovic, J.; Papapetrou, P. Unpacking the trend: Decomposition as a catalyst to enhance time series forecasting models. *Data Min. Knowl. Discov.* **2025**, *39*, 54. [CrossRef]
10. Cassola, F.; Burlando, M. Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output. *Appl. Energy* **2012**, *99*, 154–166. [CrossRef]
11. Olson, J.B.; Kenyon, J.S.; Djalalova, I.; Bianco, L.; Turner, D.D.; Pichugina, Y.; Choukulkar, A.; Toy, M.D.; Brown, J.M.; Angevine, W.M.; et al. Improving wind energy forecasting through numerical weather prediction model development. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 2201–2220. [CrossRef]
12. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [CrossRef]
13. Yıldız, S.S. Preparation of wind speed maps for Balıkesir province and their assessment in terms of wind energy potential. *Geomatik* **2021**, *6*, 198–206. (In Turkish) [CrossRef]
14. Kiriş, Z.N.; Beyca, Ö.F.; Kosanoğlu, F. Recurrent neural network based wind speed forecasting models: An application in the Yalova region. *J. Intell. Syst. Theory Appl.* **2022**, *5*, 178–188. (In Turkish) [CrossRef]
15. Buturache, A.-N.; Stancu, S. Wind energy prediction using machine learning. *Low. Carbon Econ.* **2021**, *12*, 1–21. [CrossRef]
16. Díaz-Vico, D.; Torres-Barrán, A.; Omari, A.; Dorronsoro, J.R. Deep neural networks for wind and solar energy prediction. *Neural Process. Lett.* **2017**, *46*, 829–844. [CrossRef]
17. Yürek, Ö.E.; Birant, D.; Yürek, İ. Wind power generation prediction using machine learning algorithms. *Dokuz Eylul Univ. Fac. Eng. J. Sci. Eng.* **2021**, *23*, 107–119. [CrossRef]
18. Karaman, Ö.A. Prediction of wind power with machine learning models. *Appl. Sci.* **2023**, *13*, 11455. [CrossRef]
19. Hocu, A.F.; Turker, G.F. Wind turbine power prediction using machine learning models: A case study with real farm data. *Int. J. 3D Print. Technol. Digit. Ind.* **2025**, *9*, 395–404. [CrossRef]
20. Patel, V.R.; Cakar, H.S.; Jamil, M. Short-term solar and wind power forecasting using machine learning algorithms for microgrid operation. *Energies* **2026**, *19*, 550. [CrossRef]
21. Rakshit, S.; Sengupta, A.R. Comparative analysis of various machine learning and deep learning models for wind power forecasting in Tamil Nadu, India. *Innov. Energy Syst. Technol.* **2026**, *2*, 15–27. [CrossRef]

22. Radwan, M.; Ibrahim, A.; Abdelsalam, M.M.; Alhussan, A.A.; Mattar, E.A.; El-Kenawy, E.-S.M. Optimizing solar and wind forecasting with iHow optimization algorithm and multi-scale attention networks. *Sci. Rep.* **2026**, *16*, 8597. [CrossRef] [PubMed]
23. Barchi, T.M.; dos Santos, J.L.F.; Alves, T.A.; de Mattos Neto, P.S.G.; Stevan, S.L., Jr.; Corrêa, F.C.; Medina, M.C.C.; Siqueira, H.V.; de Oliveira, J.F.L. A genetic algorithm-based ensemble framework for wind speed forecasting. *Sci. Rep.* **2026**, *16*, 6847. [CrossRef] [PubMed]
24. Bruninx, M.; van Binsbergen, D.; Verstraeten, T.; Nowé, A.; Helsen, J. Probabilistic wind power forecasting with tree-based machine learning and weather ensembles. *Electr. Power Syst. Res.* **2027**, *264*, 113847. [CrossRef]
25. Song, Z.; Gu, Y.; Liu, H.; Zou, T.; Lin, Y.; Ye, K. Application of deep learning in wind, solar, and ocean energy: An analysis of prediction, optimization, and operation & maintenance. *Renew. Sustain. Energy Rev.* **2026**, *230*, 116663. [CrossRef]
26. Wu, Z.; Luo, G.; Yang, Z.; Guo, Y.; Li, K.; Xue, Y. A comprehensive review on deep learning approaches in wind forecasting applications. *CAAI Trans. Intell. Technol.* **2022**, *7*, 129–143. [CrossRef]
27. Kandilogiannakis, G.; Mastorocostas, P.; Voulodimos, A.; Hilas, C.; Varsamis, D. Wind power prediction using a dynamic neuro-fuzzy model. *Electronics* **2025**, *14*, 2326. [CrossRef]
28. Dalar, A.Z. A hybrid fuzzy–ensemble method for time series forecasting. *Sci. Rep.* **2026**, *16*, 20705. [CrossRef] [PubMed]
29. Jia, N.; Feng, J.; Zuo, Z.; Liu, Z.; Wang, T.; Cai, C.; Li, Q. AI-driven fault detection and O&M for wind turbine drivetrains: A review of SCADA, CMS and digital twin integration. *Energies* **2026**, *19*, 1370. [CrossRef]
30. Haddad, K.; Rahman, A. Probabilistic wind speed forecasting under at-site and regional frameworks: A comparative evaluation of BART, GPR, and QRF. *Climate* **2026**, *14*, 21. [CrossRef]
31. Baggio, R.; Pujol, K.; Pantillon, F.; Lambert, D.; Filippi, J.-B.; Muzy, J.-F. Local wind speed forecasting at short time horizons based on numerical weather prediction and observations from surrounding stations. *J. Geophys. Res. Mach. Learn. Comput.* **2025**, *2*, e2025JH000709. [CrossRef]
32. Kumar, A.; Singh, A.J.; Kumar, S. Wind speed forecasting using hybrid classical-quantum-infused model. *Electr. Eng.* **2026**, *108*, 96. [CrossRef]
33. Pang, J.; Dong, S. Shape-aware Seq2Seq model for accurate multistep wind speed forecasting. *J. Ocean Univ. China* **2026**, *25*, 55–73. [CrossRef]
34. Mao, M.; Zeng, K.; Tan, Z.; Zeng, Z.; Hu, Z.; Chen, X.; Qin, C. Adaptive VMD–K-SVD-based rolling bearing fault signal enhancement study. *Sensors* **2023**, *23*, 8629. [CrossRef] [PubMed]
35. Shafiullah, M.; Katranji, A.R.; Hassan, M.; Rahman, M.M.; Shezan, S.A. Advanced multivariate deep learning methodology for forecasting wind speed and solar irradiation. *Smart Cities* **2026**, *9*, 59. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.