https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

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

A Semi-Analytic Wind-State Predictability Benchmark for Single-Turbine Wind Power Forecasting

Blerant Ramadani[1,2] and Vangel Fustic[2]

[1]Faculty of Technical Sciences, Mother Teresa University, Rr. Mirce Acev 4, 1000 Skopje, North Macedonia
[2]Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University in Skopje, Rugjer Boshkovikj 18, PO Box 574, 1000 Skopje, North Macedonia

Correspondence: Blerant Ramadani (blerant.ramadani@unt.edu.mk)

Abstract. Wind-power forecasting is commonly evaluated in relative terms, so a reported error is difficult to interpret without a reference for the uncertainty implied by the available information. We estimate a model-conditional wind-state predictability benchmark for single-turbine forecasting: the Bayes risk under squared loss implied by a Gaussian autoregressive description of future wind speed, propagated through an empirical turbine power curve and combined with scatter about that curve. The benchmark is conditional on the stated wind-only information set and fitted wind-dynamics model; it is not a universal lower bound for forecasters supplied with additional variables. We evaluate it deterministically by Gauss–Hermite quadrature over the conditional distribution of future wind speed and empirical averaging over contiguous observed AR(3) wind states. Low-order Taylor approximations overestimate the curve-variation component at long horizons by 22–61 %, depending on site and smoothing window. A variance decomposition separates the horizon-dependent contribution of wind uncertainty from residual scatter about the curve. Across two operational UK wind farms and nine horizons from one to forty-eight hours, every tested wind-only method has a point-estimate RMSE above its corresponding benchmark at all eighteen site–horizon combinations; point-estimate gaps range from 0.9 to 4.5 percentage points. The framework provides an interpretable reference for assessing forecast difficulty and for identifying where richer meteorological inputs may add value.

1 Introduction

1.1 The missing reference point

Forecast accuracy in wind energy is routinely evaluated across various statistical and deep-learning architectures (Nielsen et al., 1998; Madsen et al., 2005; Hanifi et al., 2020; Wang et al., 2021), yet reported error metrics are rarely interpreted against an explicit predictability reference expressed in equivalent error units. Although recent multi-dataset benchmarks have improved the empirical comparison of neural forecasting models (Xu et al., 2025), an isolated reduction in root mean squared error (RMSE) does not establish how much reducible uncertainty remains under a given information set.

This absence of an explicit reference has direct operational consequences. When evaluating whether to deploy a more complex model architecture, an operator must determine whether forecast errors reflect structural deficiencies of the model or the intrinsic stochastic limits imposed by the available wind record. The distinction between these two sources of error

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

clarifies whether further gains require algorithmic refinements or access to richer external predictors, such as numerical weather prediction (NWP) inputs.

1.2 Relation to prior work on bounds

To situate this framework within the relevant literature, we distinguish it from existing bounds and predictability measures. For instance, Kolumbán et al. (2023) established a lower bound associated with residual scatter around a power curve; however, the present construction extends this by formally evaluating the horizon-dependent contribution induced by uncertainty in future wind speed. In parallel, probabilistic wind-power forecasting focuses on propagating wind uncertainty into predictive distributions (Pinson, 2013; Zhang and Wang, 2014), while broader uncertainty taxonomies differentiate between aleatoric and epistemic sources (Abdar et al., 2021; Kendall and Gal, 2017; Der Kiureghian and Ditlevsen, 2009). Furthermore, methods such as forecastable component analysis (Goerg, 2013), permutation entropy (Bandt and Pompe, 2002), and other intrinsic measures (Karimi-Arpanahi et al., 2023) quantify the structural properties of a time series. While those approaches characterise the intrinsic nature of the series itself, the proposed benchmark is explicitly conditional on a specified wind-history information set and a stated wind-dynamics model, yielding a reference directly in normalized-RMSE units.

Consequently, the primary methodological distinction lies not between deterministic and probabilistic forecasting, but rather between evaluating a full forecast distribution and establishing a model-conditional baseline for the squared error of its conditional mean. The latter represents the core focus of this study.

1.3 Scope and assumptions

We derive the benchmark for a forecaster conditioning exclusively on wind speed, modeled as a stationary Gaussian autoregressive process. The evaluation is conducted at single-turbine resolution using one calendar year of hourly SCADA data from two UK sites. Consequently, this represents a model-conditional reference rather than an absolute information-theoretic limit. The resulting numerical values are inherently site- and resolution-specific; applying this framework to different turbines, wind regimes, or temporal resolutions requires re-estimating both the wind-dynamics model and the empirical power curve.

1.4 Contributions

The primary contributions of this work are threefold:

1. We define a wind-state predictability benchmark for wind-only, single-turbine forecasting, decomposing it into residual scatter about the empirical power curve and horizon-dependent variation induced by future wind uncertainty.

2. We implement a deterministic semi-analytic evaluation using Gauss–Hermite quadrature over the conditional wind distribution, averaged across observed AR(3) states, and quantify the approximation error of local Taylor expansions.

3. We benchmark the achieved errors of established wind-only forecasting methods across nine horizons at two distinct sites, explicitly quantifying the margin between operational performance and the theoretical model-conditional reference.

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

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

2 Data and Preprocessing

We utilize one calendar year of operational supervisory control and data acquisition (SCADA) data from two onshore wind farms operated by Cubico Sustainable Investments: Kelmarsh (Plumley, 2022a) and Penmanshiel (Plumley, 2022b). The analysis focuses on one 2,050 kW turbine at each site (Kelmarsh WT1 and Penmanshiel WT01). The original ten-minute records are aggregated to hourly mean wind speed and active power, aligning with standard IEC power-performance evaluation guidelines (International Electrotechnical Commission, 2022).

To prevent data leakage, all quantities required to construct the predictability benchmark are estimated exclusively on a pre-test period (7,470 hours at Kelmarsh; 5,432 hours at Penmanshiel). This precedes a strict chronological test period of 1,000 hours per site, ensuring the theoretical bound is never exposed to the data used for forecast evaluation. We estimate the empirical mean power curve using 40 equal-width wind-speed bins from zero to the 99.5th percentile of the pre-test distribution. Bins with fewer than 20 observations are discarded, and the remaining valid bin means are linearly interpolated onto a 2,000-point grid. This setup yields a bin width of approximately $0.33\,\mathrm{m\,s^{-1}}$, balancing local curve resolution with the sample size needed to reliably calculate residual scatter.

A non-parametric estimation approach is chosen over a standard cubic idealisation to preserve the true observed dynamics of cut-in behaviour, the transition to rated power, and high-wind saturation (Lydia et al., 2014). Inputs outside the fitted wind-speed support are clipped to the nearest empirical endpoint. This preprocessing pipeline formally yields the power-curve and scatter estimates applied throughout the benchmark derivation.

3 Methodology

3.1 Definition

Let $m(v)$ denote the empirical mean power curve, V_{t+h} the wind speed at the target horizon, and P_{t+h} the generated active power. Let $\mathcal{F}_t = \sigma(\{V_s : s \leq t\})$ represent the information set generated exclusively by the observed wind-speed history; past power observations are deliberately excluded. We define the squared benchmark $B^2(h)$ as the model-conditional minimum mean squared error attainable by a forecaster restricted to this wind-only information set:

$$B^2(h) = \mathbb{E}[\mathrm{Var}(P_{t+h} \mid \mathcal{F}_t)] \tag{1}$$

Under squared loss, Eq. (1) corresponds to the Bayes risk of the conditional-mean predictor, given the specified information set and fitted wind model (Gneiting, 2011). Relative to $\mathcal{F}_t$, this quantity represents the irreducible residual predictive uncertainty. Recognizing that the boundary between aleatoric and epistemic uncertainty is fundamentally dependent on the available information (Hüllermeier and Waegeman, 2021), this benchmark must not be interpreted as an absolute, information-theoretic limit of the underlying physical wind process.

A common but suboptimal heuristic maps the conditional mean wind speed directly through the nonlinear power curve, evaluating $m(\mu_h)$ rather than calculating the true conditional expectation of power. Due to the nonlinearity of $m(\cdot)$, these two

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

Table 1. Autoregressive order selection on the pre-test period. Lower BIC is preferred. Evaluated on a common, gap-aware response set (7,430 rows at Kelmarsh; 5,382 at Penmanshiel).

Site	Pre-test hours	BIC(1)	BIC(2)	BIC(3)	BIC(4)	BIC(5)
Kelmarsh WT1	7470	18779.4	18749.5	**18748.7**	18757.6	18758.1
Penmanshiel WT01	5432	17672.9	17650.0	**17639.1**	17646.0	17647.9

predictors diverge. The plug-in approach incurs an excess mean squared error equal to the squared conditional bias:

$$\mathbb{E}[(P_{t+h} - m(\mu_h))^2 \mid \mathcal{F}_t] = \mathrm{Var}(P_{t+h} \mid \mathcal{F}_t) + [\mathbb{E}[m(V_{t+h}) \mid \mathcal{F}_t] - m(\mu_h)]^2 \tag{2}$$

where $\mu_h = \mathbb{E}[V_{t+h} \mid \mathcal{F}_t]$ is the expected future wind speed. In subsequent analyses, we report Eq. (1) as the strict reference benchmark, and treat the excess term in Eq. (2) as the quantifiable penalty of adopting the plug-in shortcut.

By definition, $\mathcal{F}_t$ isolates historical wind speed. Auxiliary meteorological variables—such as temperature, air density, and wind direction—are excluded as explicit predictors; their aggregate effects are captured implicitly via the residual scatter surrounding the empirical power curve. Consequently, a forecasting system leveraging a broader multivariate information set can legitimately achieve a conditional variance lower than $B^2(h)$. Equation (1) serves strictly as a lower bound for wind-only models. If a multivariate forecaster outperforms this benchmark, the margin of outperformance quantifies the added predictive value of the external meteorological data, rather than signaling a violation of the bound.

3.2 Wind dynamics

Building upon established autoregressive frameworks for wind-speed simulation (Brown et al., 1984), we estimate Gaussian autoregressive models (orders $p = 1$ through 5) via ordinary least squares over the pre-test period. Model order is subsequently selected using the Bayesian Information Criterion (BIC) (Schwarz, 1978).

We adopt the stationary Gaussian autoregressive family as a transparent, reproducible baseline. Because its h-step conditional distribution is available analytically, this formulation enables exact deterministic propagation through the empirical power curve. This choice represents a deliberate mathematical specification rather than a phenomenological claim that physical wind dynamics are strictly stationary, Gaussian, or homoscedastic.

To ensure a rigorous, unbiased information-criterion comparison despite missing timestamps in the SCADA record, all candidate models are evaluated strictly on a common subset of contiguous response times where all five requisite lags are available. Table 1 reports the resulting BIC values, confirming AR(3) as the optimal specification at both sites.

For an arbitrary-order autoregressive process, the h-step conditional wind-speed variance is analytically derived via the Wold representation:

$$\sigma_h^2 = \sigma^2 \sum_{k=0}^{h-1} \psi_k^2 \tag{3}$$

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

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

where σ^2 denotes the estimated innovation variance, and ψ_k represents the impulse-response coefficients obtained via standard recursion from the fitted autoregressive parameters.

3.3 Semi-analytic evaluation

Equation (1) necessitates an inner expectation over the conditional distribution of future wind speed, coupled with an outer expectation over the observed wind states at the forecast origin. We compute the inner expectation deterministically via 64-node Gauss–Hermite quadrature (Golub and Welsch, 1969), while the outer expectation is approximated by averaging over contiguous AR(3) state vectors extracted from the pre-test record.

This semi-analytic approach propagates the exact Gaussian AR(3) conditional distribution directly through the empirical power curve, circumventing the need for stochastic simulation. To validate its numerical accuracy, we verify that the quadrature integration matches an independent Monte Carlo estimation to within 0.07 percentage points.

3.4 Why local approximations fail

A common analytical shortcut approximates the variance via a local Taylor expansion of the empirical power curve around the conditional mean wind speed (Oehlert, 1992):

$$\text{First-order delta:} \quad \mathrm{Var}[m(V_{t+h}) \mid \mathcal{F}_t] \approx [m'(\mu_h)]^2 \sigma_h^2$$
$$\text{Second-order Taylor:} \quad \mathrm{Var}[m(V_{t+h}) \mid \mathcal{F}_t] \approx [m'(\mu_h)]^2 \sigma_h^2 + \frac{1}{2}[m''(\mu_h)]^2 \sigma_h^4 \tag{4}$$

Our empirical evaluation demonstrates that these local expansions severely overestimate the true curve variation at extended horizons. At $h = 48$ hours, the second-order Taylor approximation inflates the variance component by 22–30 % at Kelmarsh and by 45–61 % at Penmanshiel, depending on the chosen smoothing window. The underlying failure mechanism is strictly geometric: as the forecast horizon lengthens, the conditional wind distribution disperses into the saturated upper region of the power curve. In this region, the true physical gradient approaches zero, rendering local polynomial expansions fundamentally invalid.

3.5 Variance decomposition

Applying the law of total variance yields:

$$\mathrm{Var}(P_{t+h} \mid \mathcal{F}_t) = \underbrace{\mathbb{E}[s^2(V_{t+h}) \mid \mathcal{F}_t]}_{\text{Component (A): Scatter}} + \underbrace{\mathrm{Var}(m(V_{t+h}) \mid \mathcal{F}_t)}_{\text{Component (B): Curve variation}} \tag{5}$$

where $m(\cdot)$ denotes the empirical power curve and $s^2(\cdot)$ represents the residual variance around it. Component (A) captures this residual scatter, which remains fundamentally unresolved under a wind-only information set. To contextualize this limitation, a contemporaneous regression of residual power against temperature, lagged power, and temporal encodings yields marginal explanatory power ($R^2 = 0.025$ at Kelmarsh; $R^2 = 0.043$ at Penmanshiel). Conversely, Component (B) isolates the horizon-dependent variance induced directly by uncertainty in the future wind state.

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

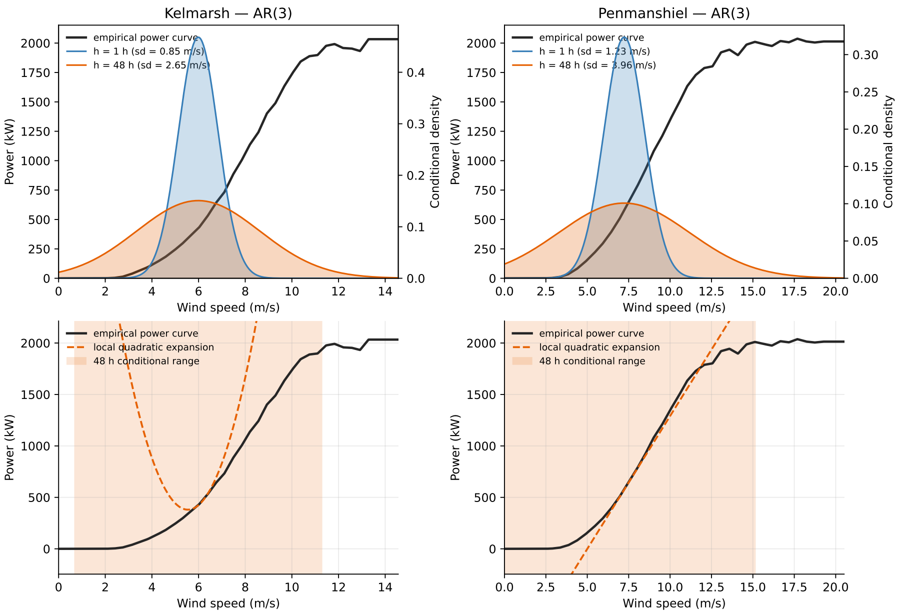

Figure 1. Geometry of the wind-state predictability benchmark. For a representative AR(3) wind state at each site, the empirical power curve is evaluated across the conditional distribution of future wind speed implied by the fitted model. This exact conditional expectation rigorously accounts for the nonlinearity and high-wind saturation of the power curve, which local Taylor approximations fail to capture.

3.6 Bootstrap uncertainty of the benchmark

To rigorously quantify finite-sample and estimator uncertainty, we implement an end-to-end moving-block bootstrap (Künsch, 1989) using 1,000 resamples with 48-hour block lengths. Across all 18 site–horizon combinations, the resulting 95 % confidence intervals exhibit half-widths ranging from 1.1 to 2.3 percentage points of rated capacity.

4 Results

4.1 The bound and its components

The AR(3)-derived wind-state benchmark exhibits a strictly monotonic increase with forecast horizon at both sites. As detailed in Table 2, the theoretical minimum RMSE rises from 10.7 % to 28.3 % of rated capacity at Kelmarsh, and from 11.6 % to 33.0 % at Penmanshiel, over the 1- to 48-hour range.

4.2 Cost of the plug-in forecaster

The excess error incurred by the plug-in heuristic (Eq. (2)) provides a quantifiable metric of its suboptimality. At Kelmarsh, this structural penalty is negligible at 1 hour (0.1 percentage points) but inflates to 1.4 points by hour 48. At Penmanshiel, the

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

Table 2. Exact wind-state predictability benchmark and RMSE-equivalent decomposition under AR(3) (% of rated capacity).

Site	h (h)	σ_h ($\mathrm{m\,s^{-1}}$)	Curve variation	Conditional scatter	Benchmark
Kelmarsh	1	0.85	9.2	5.4	10.7
	2	1.22	13.0	5.5	14.1
	3	1.45	15.4	5.5	16.4
	6	1.89	20.0	5.5	20.8
	12	2.31	24.3	5.6	25.0
	18	2.49	26.2	5.6	26.8
	24	2.57	27.0	5.6	27.6
	36	2.63	27.6	5.7	28.2
	48	2.65	27.7	5.8	28.3
Penmanshiel	1	1.23	10.8	4.2	11.6
	2	1.77	15.3	4.3	15.9
	3	2.10	18.1	4.3	18.6
	6	2.73	23.2	4.4	23.6
	12	3.37	28.2	4.6	28.5
	18	3.66	30.5	4.6	30.8
	24	3.81	31.6	4.7	32.0
	36	3.92	32.5	4.8	32.8
	48	3.96	32.6	5.1	33.0

penalty peaks at 1.3 points at 18 hours before slightly receding to 1.0 point at 48 hours, a dynamic driven by the conditional wind distribution increasingly sampling the flat, saturated region of the power curve.

4.3 The benchmark relative to achieved error

Across all eighteen evaluated site–horizon combinations, the best-performing operational wind-only model consistently registers a point-estimate RMSE above the theoretical AR(3) benchmark (Table 3).

5 Discussion

5.1 Where the constraint binds

At the shortest forecast horizons ($h \leq 3$), the operational models operate critically close to the derived benchmark. This tight convergence suggests that further algorithmic tuning or architectural changes to wind-only point forecasters will yield marginal returns, as performance is bounded by the intrinsic stochasticity of the wind record itself.

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

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

Table 3. Forecasting performance relative to the exact AR(3) model-conditional predictability benchmark (all values in normalized RMSE %).

Site	h (h)	Benchmark	Best achieved RMSE	Method	Gap (pp)
Kelmarsh	1	10.7	11.5	WPC	0.9
	2	14.1	15.9	MLP	1.7
	3	16.4	18.7	MLP	2.3
	6	20.8	23.7	MLP	2.9
	12	25.0	28.2	LR	3.2
	18	26.8	29.6	LR	2.9
	24	27.6	31.0	LR	3.4
	36	28.2	32.7	PatchTST	4.5
	48	28.3	32.7	PatchTST	4.3
Penmanshiel	1	11.6	13.0	GRU	1.4
	2	15.9	17.0	WPC	1.1
	3	18.6	19.6	GRU	1.1
	6	23.6	25.9	MLP	2.4
	12	28.5	31.8	LR	3.2
	18	30.8	33.6	MLP	2.7
	24	32.0	34.4	MLP	2.5
	36	32.8	36.0	LSTM	3.1
	48	33.0	35.8	MLP	2.7

WPC: wind persistence through power curve; LR: linear regression; MLP: multilayer perceptron; LSTM: long short-term memory; GRU: gated recurrent unit.

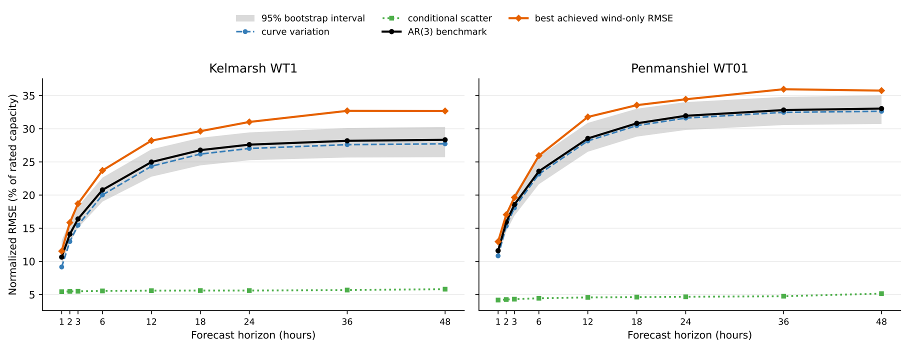

Figure 2. The conditional wind-state predictability benchmark evaluated against the best achieved wind-only error at each site. Shaded regions denote 95 % moving-block bootstrap confidence intervals.

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

5.2 Observed margins under a model-conditional reference

As the horizon extends, structural prediction gaps ranging from 2.0 to 4.5 percentage points emerge. This expanding margin confirms that beyond the intra-hour domain, substantial predictive value remains locked outside the univariate wind history. Integrating Numerical Weather Prediction (NWP) and richer meteorological inputs is the mathematically mandated path to breaking this bound (Sweeney et al., 2020; Landberg et al., 2003). These findings directly align with recent probabilistic frameworks that formally isolate weather-forecast uncertainty from power-conversion dynamics (Dantas and Browell, 2026).

5.3 Generality

Beyond wind energy, this methodological framework generalizes to any renewable generation setting where a stochastic meteorological input is mapped to power output via a nonlinear transfer function, such as solar photovoltaic generation.

6 Limitations

The present benchmark is explicitly constrained to wind-only conditioning at single-turbine resolution over a one-year period. Furthermore, while the stationary Gaussian autoregressive specification enables exact semi-analytic evaluation, it inherently abstracts away heavy-tailed distributions and non-stationary transient dynamics (Ladopoulou et al., 2026).

7 Conclusions

We derived a rigorous, model-conditional wind-state predictability benchmark for univariate single-turbine power forecasting. By evaluating the underlying expectations via Gauss–Hermite quadrature, we achieved deterministic precision within 0.07 percentage points of Monte Carlo simulations. Crucially, we demonstrated that standard local Taylor approximations fail at extended horizons, overestimating the curve-variation variance by 22–61 % at 48 hours due to high-wind saturation. Future extensions of this framework will incorporate NWP-augmented information sets, farm-level spatial aggregations, and evaluation under strictly proper scoring rules for full predictive distributions.

Code and data availability. The processed hourly data, analysis code, parameter estimates, and reproducibility scripts are available at https://github.com/blerantramadani/wind-state-predictability-benchmark and archived on Zenodo at https://doi.org/10.5281/zenodo.22249291.

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

Appendix A: Forecasting Evaluation Protocol

A1 Task configuration and contiguity

All baseline forecasts are generated using a direct multi-step forecasting protocol across horizons $h \in \{1, 2, 3, 6, 12, 18, 24, 36, 48\}$ hours. Input-target pairs are extracted exclusively from continuous hourly timestamp spans to prevent temporal leakage. Sequence-based architectures are uniformly conditioned on a 48-hour lookback window of historical wind speeds.

A2 Architectures and holdout partition

The evaluated neural architectures—MLP, LSTM, GRU, DLinear, and PatchTST—are optimized using the Adam optimizer (learning rate 0.001, MSE loss, batch size 32) and averaged over 10 random weight initializations. At each site, the final 1,000 chronological hours are strictly reserved as an independent holdout set for final evaluation.

Author contributions. BR: Conceptualization, methodology, software, formal analysis, data curation, visualization, writing – original draft. VF: Supervision, validation, writing – review and editing.

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

Disclaimer. During manuscript preparation, the authors used OpenAI ChatGPT and Codex for language editing, document formatting, and programming assistance. The authors reviewed and revised all resulting content and take full responsibility for the published work.

Acknowledgements. The authors acknowledge the open-access SCADA data provided by Cubico Sustainable Investments.

https://doi.org/10.5194/wes-2026-154
Preprint. Discussion started: 10 September 2026

WIND ENERGY SCIENCE DISCUSSIONS
Open Access

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Information Fusion, 76, 243–297, https://doi.org/10.1016/j.inffus.2021.05.008, 2021.

Bandt, C. and Pompe, B.: Permutation entropy: A natural complexity measure for time series, Physical Review Letters, 88, 174 102, https://doi.org/10.1103/PhysRevLett.88.174102, 2002.

Brown, B. G., Katz, R. W., and Murphy, A. H.: Time series models to simulate and forecast wind speed and wind power, Journal of Climate and Applied Meteorology, 23, 1184–1195, https://doi.org/10.1175/1520-0450(1984)023<1184:TSMTSA>2.0.CO;2, 1984.

Dantas, G. and Browell, J.: Seamless short- to mid-term probabilistic wind power forecasting, Wind Energy, 29, e70 079, https://doi.org/10.1002/we.70079, 2026.

Der Kiureghian, A. and Ditlevsen, O.: Aleatoric or epistemic? Does it matter?, Structural Safety, 31, 105–112, https://doi.org/10.1016/j.strusafe.2008.06.020, 2009.

Gneiting, T.: Making and evaluating point forecasts, Journal of the American Statistical Association, 106, 746–762, https://doi.org/10.1198/jasa.2011.r10138, 2011.

Goerg, G. M.: Forecastable component analysis, in: Proceedings of the 30th International Conference on Machine Learning, vol. 28, pp. 64–72, 2013.

Golub, G. H. and Welsch, J. H.: Calculation of Gauss quadrature rules, Mathematics of Computation, 23, 221–230, https://doi.org/10.1090/S0025-5718-69-99647-1, 1969.

Hanifi, S., Liu, X., Lin, Z., and Lotfian, S.: A critical review of wind power forecasting methods – past, present and future, Energies, 13, 3764, https://doi.org/10.3390/en13153764, 2020.

Hüllermeier, E. and Waegeman, W.: Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Machine Learning, 110, 457–506, https://doi.org/10.1007/s10994-021-05946-3, 2021.

International Electrotechnical Commission: Wind energy generation systems – Part 12-1: Power performance measurements of electricity producing wind turbines, Tech. Rep. IEC 61400-12-1:2022, IEC, 2022.

Karimi-Arpanahi, S., Pourmousavi, S. A., and Mahdavi, N.: Quantifying the predictability of renewable energy data for improving power systems decision-making, Patterns, 4, 100 708, https://doi.org/10.1016/j.patter.2023.100708, 2023.

Kendall, A. and Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision?, Advances in Neural Information Processing Systems, 30, 2017.

Kolumbán, S., Kapodistria, S., and Nooraee, N.: Short term wind turbine power output prediction, in: Performance Evaluation Methodologies and Tools: 15th EAI International Conference, Springer, 2023.

Künsch, H. R.: The jackknife and the bootstrap for general stationary observations, The Annals of Statistics, 17, 1217–1241, https://doi.org/10.1214/aos/1176347265, 1989.

Ladopoulou, D., Hong, D. M., and Dellaportas, P.: Probabilistic wind power modelling via heteroscedastic non-stationary Gaussian processes, Energy Reports, 15, 108 895, https://doi.org/10.1016/j.egyr.2025.12.032, 2026.

Landberg, L., Giebel, G., Nielsen, H. A., Nielsen, T. S., and Madsen, H.: Short-term prediction – an overview, Wind Energy, 6, 273–280, https://doi.org/10.1002/we.96, 2003.

Lydia, M., Suresh Kumar, S., Immanuel Selvakumar, A., and Edwin Prem Kumar, G.: A comprehensive review on wind turbine power curve modeling techniques, Renewable and Sustainable Energy Reviews, 30, 452–460, https://doi.org/10.1016/j.rser.2013.10.030, 2014.

Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. A., and Nielsen, T. S.: Standardizing the performance evaluation of short-term wind power prediction models, Wind Engineering, 29, 475–489, 2005.

Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L., and Giebel, G.: A new reference for wind power forecasting, Wind Energy, 1, 29–34, https://doi.org/10.1002/(SICI)1099-1824(199809)1:1<29::AID-WE6>3.0.CO;2-7, 1998.

Oehlert, G. W.: A note on the delta method, The American Statistician, 46, 27–29, https://doi.org/10.1080/00031305.1992.10475842, 1992.

Pinson, P.: Wind energy: Forecasting challenges for its operational management, Statistical Science, 28, 564–585, https://doi.org/10.1214/13-STS445, 2013.

Plumley, C.: Kelmarsh Wind Farm Data, https://doi.org/10.5281/zenodo.5841834, 2022a.

Plumley, C.: Penmanshiel Wind Farm Data, https://doi.org/10.5281/zenodo.5946808, 2022b.

Schwarz, G.: Estimating the dimension of a model, The Annals of Statistics, 6, 461–464, https://doi.org/10.1214/aos/1176344136, 1978.

Sweeney, C., Bessa, R. J., Browell, J., and Pinson, P.: The future of forecasting for renewable energy, WIREs Energy and Environment, 9, e365, https://doi.org/10.1002/wene.365, 2020.

Wang, Y., Zou, R., Liu, F., Zhang, L., and Liu, Q.: A review of wind speed and wind power forecasting with deep neural networks, Applied Energy, 304, 117 766, https://doi.org/10.1016/j.apenergy.2021.117766, 2021.

Xu, X., Cao, Q., Deng, R., Guo, Z., Chen, Y., and Yan, J.: A cross-dataset benchmark for neural network-based wind power forecasting, Renewable Energy, 254, 123 463, https://doi.org/10.1016/j.renene.2025.123463, 2025.

Zhang, J. and Wang, J.: A review on probabilistic forecasting of wind power generation, Renewable and Sustainable Energy Reviews, 32, 255–270, https://doi.org/10.1016/j.rser.2014.01.033, 2014.