

Inaccurate Wind Power Forecasting During Extreme Weather Events Using Normalized Machine Learning

Amisha Nandansingh Negi*, Dr. Manoj Singh**

Department of Computer Science, S.I.E.S College of Arts, Science and Commerce (Empowered Autonomous) Mumbai, India

Email: negi2000amisha@gmail.com

Department of Computer Science, S.I.E.S College of Arts, Science and Commerce (Empowered Autonomous) Mumbai, India

Email: manojks@sies.edu.in

Abstract:

Wind-power prediction is reasonably accurate in normal weather but degrades during high winds, gusts and rapid changing conditions, and a model fitted to one turbine does not directly transfer to turbines of other capacities. This paper presents an extreme-weather-aware, capacity-normalized framework that converts 10-minute SCADA and meteorological data into turbine power with Random Forest regressor. The model predicts power as a fraction of rated capacity, uses features such as air density, density-corrected wind speed, gust factor, rapid wind changes and one-hour variability, labels statistically extreme conditions with thresholds learned from training data, and gives extreme observations five times the training weight. The framework is demonstrated on a dataset of 11,520 records from site (coastal, flat inland, hilly and plateau) with 2.0–5.0 MW turbines, using a per-site chronological 80/20 split. On the unseen test period the model reached an NMAE of 2.01%, an NRMSE of 2.55% and an R^2 of 0.992, with an aggregate energy error of -0.06% (414.42 MWh predicted versus 414.68 MWh actual). Errors rose under extreme conditions (NMAE 2.25% versus 1.97% in normal conditions), confirming that extreme periods should be reported separately. Sensitivity test showed that the extreme weighting did not reduce extreme-condition error on this dataset.

Keywords — wind power forecasting, extreme weather, Random Forest, SCADA, normalized power, multisite modeling

I. INTRODUCTION

A wind turbine turns the kinetic energy carried by moving air into electricity, yet its output responds to the weather in a strongly nonlinear way. Below the cut-in wind speed a turbine produces little or no power; between cut-in and rated speed output rises steeply; near rated speed it is limited by the controller; and at very high wind speed the turbine may derate or shut down for protection. Output also depends on gusts, direction, turbulence, pressure, temperature and air density, and on site-specific factors such as terrain, wakes, curtailment and maintenance.

Extreme conditions are especially difficult to predict. Extreme observations are rare in historical records, wind can change rapidly within a few intervals, gusts and turbulence create unstable inflow, and turbines may derate or stop. A model trained mostly on normal weather can therefore show excellent overall scores while making larger errors in the periods that matter most. In addition, many models are trained and

evaluated for a single turbine or site and predict power directly in kilowatts, so they learn the scale of that turbine and do not transfer to machines of different rated capacity. Most evaluations also report one overall error, which can hide weaker performance in high-wind periods [1].

This paper addresses both gaps with a simple, reproducible framework. Its contributions are:

- a single model for several turbine capacities, obtained by predicting normalized power and rescaling by rated capacity;
- atmospheric and terrain-related feature engineering, including air density and density-corrected wind speed;
- statistical extreme-condition labels derived only from training data, with greater training weight for extreme rows;
- chronological unseen testing with overall, normal and extreme metrics reported separately, plus site-level energy comparison.

Section II reviews related work, Section III describes the data and method, Section IV presents results, Section V discusses them with the limitations, and Section VI concludes.

II. RELATED WORK

Wind-power forecasts are evaluated with a range of error measures; Piotrowski et al. [1] surveyed these *measures* across a large body of published studies, examined how the reported errors are distributed, and favor normalized measures such as NMAE and NRMSE when turbines and farms of different sizes are compared. Transfer between locations is a recognized difficulty: Sajol et al. [2] report that shifts in geographic and climatic data weaken model robustness and use deep domain adaptation to carry a power model from one location to another, while Chokhachian et al. [3] *adjust* a power curve learned at an operating farm so that it can be applied at a different farm. Random Forests [4] are a common choice for nonlinear regression on tabular data; in the scikit-learn implementation [5], [6], many regression trees are grown on bootstrap samples and their outputs are averaged, which raises accuracy and limits overfitting. The present work does not propose a new learner. Its focus is a complete, transparent pipeline that combines capacity normalization, training-only extreme labelling, extreme weighting and condition-wise evaluation.

III. METHODOLOGY

Fig. 1 summarizes the framework. Each stage is described below.

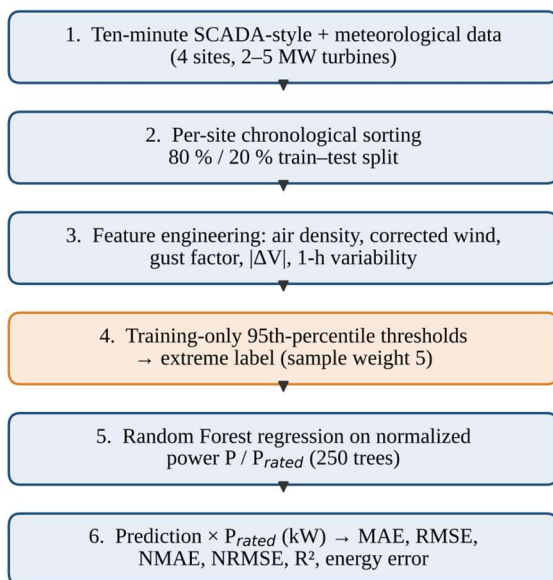


Fig. 1 Proposed extreme-weather-aware, capacity-normalized forecasting framework.

A. Dataset

The study uses an original dataset constructed for controlled multisite model development. It follows a ten-minute SCADA-style schema and contains four site profiles with different terrain, elevation, hub height, surface roughness and rated capacity (Table I). Each site has 2,880 records covering 1–20 January 2025, giving 11,520 observations in total. Recorded fields are timestamp, site and turbine identifiers, terrain, rated power, active power, mean wind speed, maximum (gust) wind speed, wind direction, temperature, pressure, humidity, turbulence intensity and air density.

Wind speed was generated as a diurnal and weekly cycle plus a first-order autoregressive fluctuation and six storm-like pulses per site. Power follows a cubic power curve between a cut-in speed of 3 m/s and a rated speed of 13 m/s on density-corrected wind, with cut-out at 25 m/s, gust derating, a terrain-dependent loss factor and measurement noise. The dataset is not an independently measured operational SCADA collection. It provides controlled variation in capacity, terrain, altitude and extreme-wind patterns for testing the workflow, but it lacks real faults, curtailment, missing records and maintenance. Fig. 2 shows the resulting normalized power curve for the test period.

TABLE I
SITE PROFILES IN THE MULTISITE DATASET

Site	Terrain	Rated (kW)	Elev. (m)	Hub (m)	Rough h. z_0	Mean wind (m/s)	Mean ρ (kg/m ³)
Coastal_A	Coastal	2,000	25	90	0.03	9.46	1.185
Inland_B	Flat inland	3,000	180	100	0.12	8.04	1.167
Hilly_C	Hilly	4,200	620	110	0.28	8.90	1.119
Plateau_D	Plateau	5,000	1,050	120	0.08	10.03	1.073

a. Mean wind speed and air density are computed over all 2,880 records of each site.

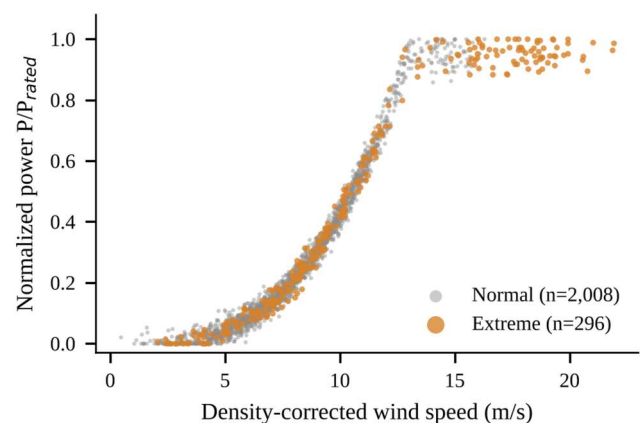


Fig. 2 Normalized power versus density-corrected wind speed in the unseen test period (all sites), separated into normal and statistically extreme observations.

B. Data Preparation and Chronological Split

Records were sorted by site and timestamp, and the essential fields (timestamp, rated power, wind speed and, for training, actual power) were validated. Every row carries its own rated power, so turbines of different capacities share one model. Change and rolling features were computed within each site so that differences never cross turbine boundaries. For each site, the earliest 80% of records (1–16 January) formed the training set and the latest 20% (17–20 January) the unseen test set, giving 9,216 training and 2,304 test rows. A random split was avoided because, in time-ordered data, it would let later observations leak into training and make the scores look better than they really are.

C. Feature Engineering

The target is normalized power, which expresses output as a fraction of rated capacity:

$$P_{\text{norm}} = P_{\text{actual}} / P_{\text{rated}} \quad (1)$$

A value of 0.5 therefore means half of rated output for both a 2 MW and a 5 MW turbine. Predictions are converted back to kilowatts by multiplying by P_{rated} . Air density is approximated with the dry-air ideal-gas relation, and wind speed is corrected to a reference density $\rho_0 = 1.225 \text{ kg/m}^3$ because denser air carries more kinetic energy at the same speed:

$$\rho = p / (R_d T) \quad (2)$$

$$V_c = V (\rho / \rho_0)^{1/3} \quad (3)$$

where p is pressure in pascals ($100 \times \text{hPa}$), T is temperature in kelvin ($^{\circ}\text{C} + 273.15$) and $R_d = 287.05 \text{ J/(kg} \cdot \text{K)}$. Gust behaviour and rapid change are described by the gust factor, the absolute wind change between consecutive records and the one-hour (six-record) rolling standard deviation:

$$G = V_{\text{max}} / \max(V, 0.1) \quad (4)$$

$$|\Delta V_t| = |V_t - V_{t-1}| \quad (5)$$

$$\sigma_{V,t} = \text{std}(V_{t-5}, \dots, V_t) \quad (6)$$

The absolute pressure change $|\Delta p|$ is computed in the same way. Wind direction and time of day are circular, so they are encoded with sine and cosine components so that 0° and 360° (or 23:50 and 00:00) are adjacent. Elevation, hub height, surface roughness and an integer terrain code describe the site. In total the model uses 18 input features; actual test power is never an input.

D. Extreme-Condition Labelling

An observation is labelled extreme when any of four variables reaches the 95th percentile Q_{95} of the training data:

$$E_t = 1 \text{ if } \max_k (x_{k,t} / Q_{95}(x_k)) \geq 1, \text{ else } 0 \quad (7)$$

where $x_k \in \{V, V_{\text{max}}, |\Delta V|, \sigma_V\}$, so a row is extreme if any one variable reaches its threshold.

Thresholds (Table II) are computed from training data only, because calculating them on the test set would let unseen observations influence evaluation. The label marks statistically unusual wind conditions and is not an official weather-agency storm declaration. It flagged 1,295 of 9,216 training rows (14.1%) and 296 of 2,304 test rows (12.8%).

TABLE II
TRAINING-ONLY EXTREME THRESHOLDS (95TH PERCENTILE)

Variable	Threshold
Mean wind speed V	16.49 m/s
Maximum (gust) wind V_{max}	21.06 m/s
Absolute wind change $ \Delta V $	1.48 m/s
One-hour variability σ_V	1.34 m/s

E. Random Forest Model and Extreme Weighting

A Random Forest regressor [4], [5] was chosen because it captures nonlinear effects and interactions between features, works without strict feature scaling, accepts sample weights, reports feature importance, and is easier to build and explain than a deep network. The forest uses 250 trees, a maximum depth of 16, at least three samples per leaf and a fixed random seed of 42. Because extreme rows are less common, they receive a sample weight of 5 against 1 for normal rows, so their errors carry more influence during training. Predictions are clipped to $[0, 1]$ before rescaling to kilowatts.

F. Evaluation Metrics

With actual power y_i , prediction $\hat{y}_i$ and n test records, the metrics are:

$$\text{MAE} = (1/n) \sum |y_i - \hat{y}_i| \quad (8)$$

$$\text{RMSE} = \sqrt{(1/n) \sum (y_i - \hat{y}_i)^2} \quad (9)$$

$$\text{NMAE} = (100/n) \sum |y_i - \hat{y}_i| / P_{\text{rated},i} \quad (10)$$

$$R^2 = 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2 \quad (11)$$

NRMSE is the normalized form of (9), computed on per-turbine normalized power, and is reported in percent. Energy is the sum of ten-minute power divided by six, and energy error is (predicted – actual)/actual. Because forecasting is a regression problem, a single “percentage accuracy” such as $100 - \text{NMAE}$ is not reported [1].

IV. RESULTS

A. Overall and Condition-Wise Accuracy

Table III gives the unseen-test results. Over all 2,304 test records the model predicted 414.42 MWh against

414.68 MWh actual, an energy difference of -0.26 MWh (-0.06%). Overall MAE was 71.00 kW, NMAE 2.01%, NRMSE 2.55% and R^2 0.9923. Fig. 3 shows that the predicted series follows the measured output closely at Hilly_C, including ramps and the plateau at rated power during the extreme period on 19 January.

TABLE III
UNSEEN-TEST ACCURACY BY CONDITION (ALL SITES)

Condition	Rows	MAE (kW)	RMSE (kW)	NMAE (%)	NRMSE (%)	R^2
Overall	2,304	71.00	94.85	2.01	2.55	0.9923
Normal	2,008	68.83	91.74	1.97	2.50	0.9907
Extreme	296	85.77	113.76	2.25	2.84	0.9945

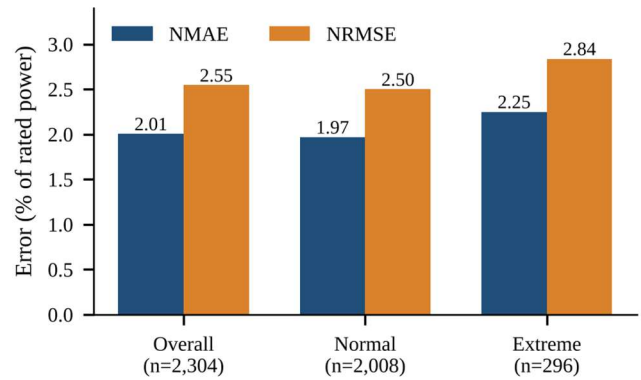


Fig. 4 NMAE and NRMSE on the unseen test set for all, normal and extreme observations.

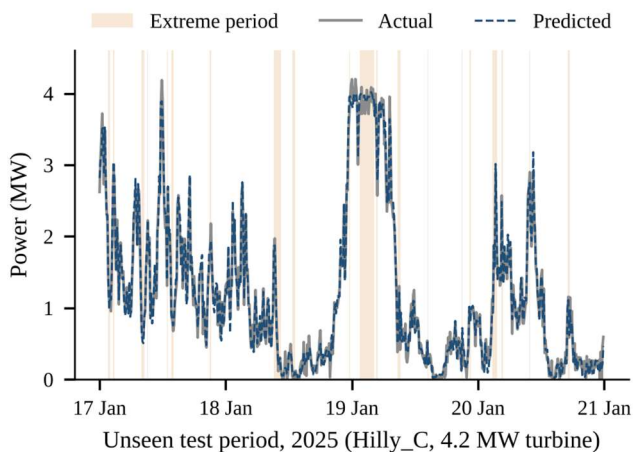


Fig. 3 Actual and predicted power for Hilly_C (4.2 MW) over the unseen test period; shaded intervals are training-threshold extreme conditions.

Every error measure is higher under extreme conditions (Fig. 4): MAE rises from 68.83 to 85.77 kW, NMAE from 1.97% to 2.25% and NRMSE from 2.50% to 2.84%. NRMSE is above NMAE in every group because squaring gives large individual errors more weight in RMSE. An overall figure alone would hide this increase, which supports reporting normal and extreme results separately.

B. Site-Wise Results

Table IV and Fig. 5 compare actual and predicted energy at each site. Site energy errors lie between -0.74% (Coastal_A) and $+0.40\%$ (Hilly_C), and site NMAE between 1.91% and 2.14%. One model, trained on normalized power, therefore rescaled correctly to 2.0, 3.0, 4.2 and 5.0 MW turbines. Extreme-condition NMAE was highest at Coastal_A (2.46%); at Inland_B it was slightly lower than the normal-condition value (1.83% against 1.94%), based on only 51 extreme rows. Agreement in total energy does not mean every ten-minute prediction is accurate, which is why MAE, RMSE and Fig. 3 are also reported.

TABLE IV
ENERGY AND ERROR BY SITE (UNSEEN TEST PERIOD)

Site	Actual (MWh)	Predicted (MWh)	Energy error (%)	NMAE (%)	Extreme NMAE (%)
Coastal_A	70.39	69.87	-0.74	2.05	2.46
Inland_B	75.63	75.67	+0.05	1.93	1.83
Hilly_C	117.19	117.66	+0.40	2.14	2.25
Plateau_D	151.48	151.22	-0.17	1.91	2.39
All sites	414.68	414.42	-0.06	2.01	2.25

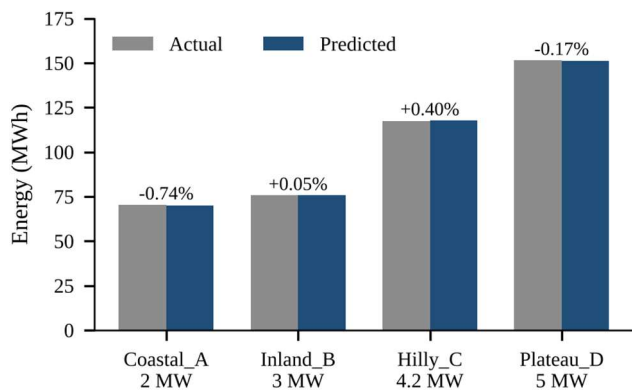


Fig. 5 Actual versus predicted energy at each site in the unseen test period; labels give the energy error.

C. Baseline Comparison and Weighting Sensitivity

Table V compares the proposed model with a mean-power baseline, which predicts the training mean of normalized power, and with Random Forests trained using extreme weights of 1, 2 and 10. All forests share the other settings. The Random Forest reduced overall NMAE from 23.21% to about 2%, so weather and SCADA inputs predict power far better than the baseline. However, raising the extreme weight did not lower extreme-condition error: the unweighted model ($w = 1$) gave an extreme NMAE of 2.24%, the proposed $w = 5$ gave 2.25% and $w = 10$ gave 2.27%. The differences are small, but on this dataset the weighting did not deliver the intended benefit.

TABLE V
BASELINE AND EXTREME-WEIGHT SENSITIVITY (UNSEEN TEST SET)

Model	Overall NMAE (%)	Normal NMAE (%)	Extreme NMAE (%)	Extreme MAE (kW)	Overall R^2
Mean-power baseline	23.21	22.00	31.46	1191.1	0.072
RF, $w = 1$	1.99	1.96	2.24	85.3	0.992
RF, $w = 2$	1.99	1.95	2.24	85.4	0.992
RF, $w = 5$ (proposed)	2.01	1.97	2.25	85.8	0.992
RF, $w = 10$	2.02	1.99	2.27	86.8	0.992

b. w is the sample weight of extreme training rows; normal rows have weight 1.

D. Feature Importance

Density-corrected wind speed dominates the impurity-based importance (0.993), followed by maximum (gust) wind speed (0.003); all other features are below 0.001 (Fig. 6). This is expected because the power was generated mainly from corrected wind speed, and it should not be read as evidence that pressure, humidity or turbulence are unimportant in real turbines.

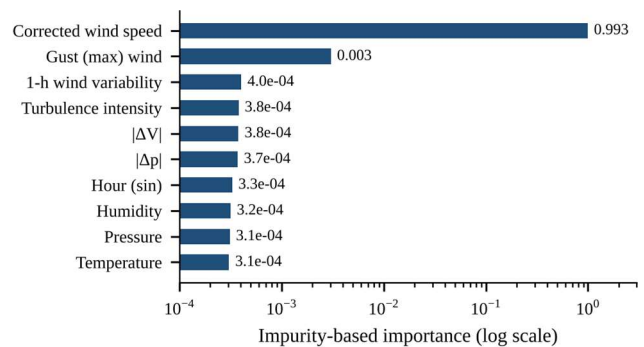


Fig. 6 Ten most important Random Forest features (impurity-based importance, logarithmic scale).

V. DISCUSSION

The results support the first hypothesis: weather and SCADA inputs predict power with far lower error than a mean-power baseline. The second hypothesis, that extreme weighting lowers extreme-condition error, was not supported here. One likely reason is visible in Fig. 2: many extreme observations lie on the rated-power plateau above 13 m/s, where output is easy to predict, so extra weight mainly reinforces an already well-learned region. Extreme-condition R^2 (0.9945) is higher than normal-condition R^2 even though its MAE is larger, because R^2 measures explained variation and extreme periods span a wider power range. R^2 must therefore be read together with MAE and RMSE. The third hypothesis, on transfer to unseen sites, was not tested, because chronological testing on known sites is not the same as testing a completely new site.

The main limitations are:

- ten-minute records can hide short gusts, and turbulence intensity here is a simplified variable rather than one computed from high-frequency samples;
- the dry-air density equation ignores moist-air effects, and real power is also affected by curtailment, faults, icing, maintenance and wakes that depend on farm layout;
- normalization and atmospheric features reduce, but cannot remove, terrain effects;
- the Random Forest does not learn long temporal sequences and may not extrapolate beyond the training range;
- the 95th-percentile threshold and the weight of 5 are modelling choices, and the extreme label is not an official storm warning;
- measured weather was used as input, so this is weather-to-power estimation; true future

forecasting needs weather-forecast inputs, whose errors add to the power error.

VI. CONCLUSION AND FUTURE SCOPE

This paper presented a complete extreme-weather-aware, capacity-normalized and multisite wind-power framework. One Random Forest, trained on normalized power with atmospheric, gust and variability features, predicted output for 2.0–5.0 MW turbines at four site profiles with an overall NMAE of 2.01%, an NRMSE of 2.55%, an R^2 of 0.992 and an aggregate energy error of –0.06%. Errors were consistently higher in statistically extreme conditions (NMAE 2.25% against 1.97%), which confirms the value of condition-wise reporting. Extreme sample weighting did not improve extreme-condition accuracy on this dataset. The results demonstrate a working, leakage-free pipeline, not universal real-world performance.

Future work will validate the framework on independently sourced operational SCADA, such as the open EDP onshore wind-farm dataset [7] and others catalogued in [8]; apply leave-one-site-out validation and local calibration for unseen sites [2], [3]; add official weather-event labels, curtailment and availability fields, and turbine-layout and wake information; compare against gradient boosting, LSTM and transformer models; use numerical weather predictions for true multi-horizon forecasting; and report prediction intervals to quantify uncertainty.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science, S.I.E.S College of Arts, Science and Commerce (Empowered Autonomous), Mumbai, for providing the facilities and academic support needed to carry out this work. They also acknowledge the developers of the open-source Python libraries, including scikit-learn, that were used to build and evaluate the models.

REFERENCES

- [1] P. Piotrowski, I. Rutyna, D. Baczyński, and M. Kopyt, "Evaluation metrics for wind power forecasts: A comprehensive review and statistical analysis of errors," *Energies*, vol. 15, no. 24, Art. no. 9657, Dec. 2022, doi: 10.3390/en15249657.
- [2] M. S. I. Sajol, M. S. Islam, A. S. M. J. Hasan, M. S. Rahman, and J. Yusuf, "Wind power prediction across different locations using deep domain adaptive learning," arXiv:2405.11188, 2024.
- [3] A. Chokhachian, V. R. Joseph, and Y. Ding, "Domain-adapted power curve for cross-farm applications," arXiv:2607.19744, Jul. 2026.
- [4] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [5] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [6] scikit-learn developers, "RandomForestRegressor," scikit-learn documentation. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
- [7] "Description of the SCADA dataset of the EDP onshore wind farm in Portugal," Mendeley Data. [Online]. Available: <https://data.mendeley.com/datasets/zjxjnp3xs>
- [8] "OpenWindSCADA: A curated list of open wind turbine data sets and corresponding code," GitHub repository. [Online]. Available: <https://github.com/sltzgs/OpenWindSCADA>