

Site-adapted U-Net and classical benchmarking for wind turbine power curve modeling under SCADA data quality variation

Syed Muhammad Hamza Tauqeer

`stauqeer.msee24mcs@student.nust.edu.pk`

National University of Sciences and Technology

Shibli Nisar

National University of Sciences and Technology

Armoghan Khan

National University of Sciences and Technology

Ammad Ali Khan

National University of Sciences and Technology

Article

Keywords:

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11205206/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Site-adapted U-Net and classical benchmarking for wind turbine power curve modeling under SCADA data quality variation

Syed Muhammad Hamza Tauqeer^{1*}, Shibli Nisar¹,
Armoghan Khan¹, Ammad Jadoon¹

¹Department of Electrical Engineering, Military College of Signals,
National University of Sciences and Technology, Islamabad, 44000,
Pakistan.

*Corresponding author(s). E-mail(s):
stauqeer.msee24mcs@student.nust.edu.pk;

Abstract

Wind turbine power curves support performance monitoring, but operational records mix normal production with curtailment, faults and sparse high-wind data. We compared classical curves, conditional gradient boosting and an image-based U-Net using supervisory control and data acquisition records from two turbine fleets in Pakistan. The U-Net was trained on measured-power scatter with controller-derived capability targets; frozen inference used only wind speed and measured power. On a 2025 holdout, multi-input boosting reduced power-estimation root mean square error from 143.1 to 81.9 kW compared with wind-speed-only boosting. On 100 unseen 2026 turbine-months, the U-Net achieved normalized error 0.0286 and $R^2 = 0.9946$ against controller capability. The unchanged checkpoint achieved error 0.0544 and $R^2 = 0.9578$ on 240 turbine-months from a lower-rated fleet after input-only axis calibration. In a paired pipeline comparison, error was 83.0% and 79.8% lower than a released image-based method on the two field holdouts, respectively. These results support capability screening across two ratings within one wind corridor; controller-derived targets and limited geographic coverage constrain broader claims.

Keywords: Wind turbine power curve, SCADA, U-Net, curtailment, model transfer, gradient boosting

1 Introduction

The wind power curve (WPC) relates wind speed at hub height to the electrical power produced by a wind turbine. It supports power forecasting, performance monitoring, maintenance planning, and curtailment assessment. A manufacturer curve is obtained under prescribed atmospheric and operating conditions and therefore may not reproduce the behaviour of a turbine exposed to site-specific air density, wake interaction, grid limitation, control actions, and measurement uncertainty.

SCADA systems provide the measurements required to construct an operational WPC, but the resulting cloud combines normal generation, curtailed states, faults, and sparse high-wind observations. Parametric logistic curves offer compact analytical descriptions but constrain the transition shape. Interpolation, smoothing splines, and machine-learning regressors are more flexible, although their result can depend strongly on preprocessing and turbine-specific fitting [1–3].

Recent approaches span outlier filtering, multivariable regression, and image-based recovery; their distinct targets and assumptions are reviewed below. Two complementary learning formulations are considered here. A multi-input regressor uses measured control and environmental state to estimate contemporaneous power, whereas an image-to-curve model recovers a one-dimensional capability relation from an irregular SCADA cloud. The DITU-Net method in [4] showed that a U-Net trained on scatter images can recover WPCs without a separate outlier-removal stage. However, synthetic target families may not represent prolonged site curtailment, and a multi-input estimator must not be presented as a unique function of wind speed.

This study compares classical curves and a conditional XGBoost estimator with a site-adapted image model on a year holdout and a second turbine rating. It tests raw and filtered SCADA inputs, a published checkpoint on identical images, missing high-wind observations, and capability-based energy accounting. The image model uses controller theoretical power (TP) to form historical training targets but receives only wind speed and measured grid power (GP) at inference. Its claims concern capability recovery within the studied wind corridor, rather than future-power forecasting.

2 Related research

Power-curve studies address different targets: a wind-speed-only capability relation, a conditional prediction using operating variables, and a future-power forecast. Their errors cannot be compared directly when the reference, horizon or data partition differs. A broad review of power-curve applications and modelling methods provides this context [5]; a separate multi-time-scale forecasting review clarifies the forecasting task [6].

For SCADA curve construction, clustering-based outlier removal [7], hybrid physical and data-driven cleaning [8], isolation-forest and density-based cleaning [1], and two-stage outlier detection with subsequent power prediction [2] all address contamination before or during fitting. Marti-Puig et al. [9] examine how power-curve estimation reliability depends on the modelling and data treatment. Yaw-adjusted local regression [10] and stacked normal-behaviour regressors [11] show why departures from a fitted curve require care when interpreted as faults. These approaches motivate reporting

raw and GP-only-filtered inputs separately rather than treating one cleaning choice as ground truth.

Flexible multivariable fits capture information unavailable to a one-dimensional curve. Qiao et al. [3] represent control segments and short-time dependence, Wang et al. [12] develop a deep data-driven curve estimator, and Yin et al. [13] combine multivariable gradient boosting with optimization. Risk-regularized curve learning explicitly addresses uncertain and anomalous SCADA observations [14]. For resource assessment, Haddad [15] combines wind-speed distributions and turbine curve assessment; the distinct fixed-wing airborne system of Joshi et al. [16] illustrates why power-curve transfer claims should specify the device and operating regime. These works support the inclusion of both flexible baselines and an explicit cross-rating boundary.

The closest image-based predecessor is the scatter-to-curve DITU-Net of Yang et al. [4]; Wu et al. [17] extend vision-based modelling to other turbine performance relations. We therefore evaluate the released DITU-Net pipeline on the same rendered inputs, record curve-reader failures, and distinguish site-adapted training targets from post-processing effects. A common rendered image does not by itself isolate the learned weights because each pipeline retains its native curve reader.

Short-term forecasting uses different supervision and is relevant to downstream use of a curve, not a direct accuracy baseline. The theoretical-curve and temporal-fusion approach of Michalakopoulos et al. [18], coastal-farm machine learning of Gu et al. [19], adaptive wind-speed calibration of Wang et al. [20], ensemble integration of Rajaperumal and Columbus [21], extreme-weather combination forecasts of Li et al. [22], and day-ahead probabilistic ramps of Cui et al. [23] target future power or its uncertainty. Their forecasting metrics cannot establish that an inferred monthly capability curve predicts future grid output. Dynamic assessment against reserve scarcity [24] further motivates an operator-facing, explicitly qualified interpretation of capability shortfall.

The remaining question is whether an image model trained against a controller-derived site capability signal can recover a stable curve from raw measured-power scatter on a later year and another turbine rating. The present study tests that question alongside conditional power prediction, classical fitting, input-quality sensitivity, and an auditable external-checkpoint control.

3 Results

3.1 Classical and Image-Model Benchmark

Fig. 1 compares fleet-average RMSE and MAPE under the three scenarios. TP denotes the grid-power-to-controller-capability discrepancy and is not a fitted competitor. The U-Net is the lowest-error fitted model in each scenario. Its small variation across S1–S3 indicates reduced dependence on a particular cleaning rule. Full per-scenario RMSE, MAE, MAPE, and WMAPE for every classical baseline are given in Supplementary Table S2 (truncation) and Table 2 (primary deployment); the α -shooting scores at the 0.05/0.10/0.15 tolerance bands follow the same ordering established here and are omitted for space.

Supplementary Fig. S1 compares the U-Net with the lowest-RMSE classical baseline for each turbine. The U-Net remains on the lower-error side for all 20 machines,

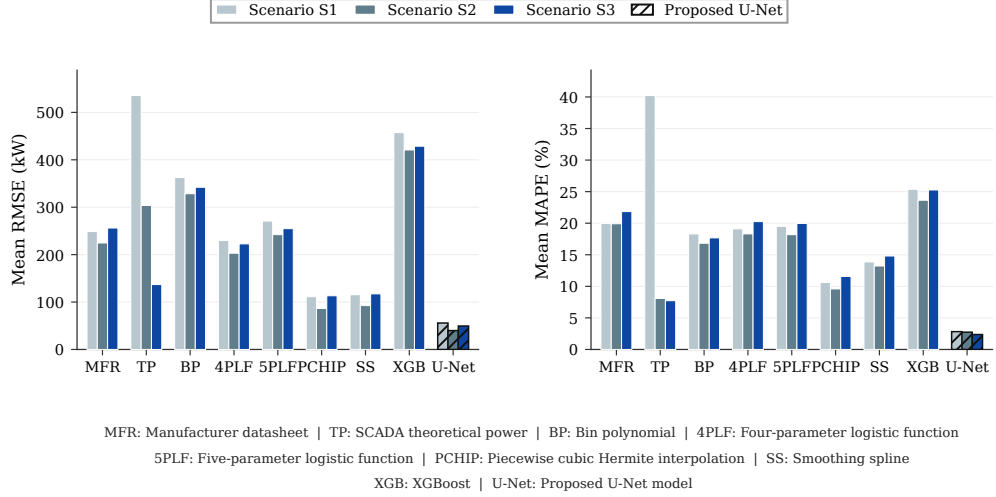


Fig. 1 Fleet-average RMSE and MAPE under S1-S3. Hatched bars denote the U-Net; TP is a reference discrepancy.

showing that the fleet improvement is not produced by a small favourable subset. Variation in magnitude is consistent with turbine-specific curtailment, wake exposure, and sampling density near rated power.

3.2 Temporal and Seasonal XGBoost Performance

Across the complete 2025 holdout, XGB-MI obtains RMSE 81.9 kW, MAE 54.7 kW, MAPE 7.57%, WMAPE 8.08%, and $R^2 = 0.978$. XGB-SI obtains 143.1 kW, 95.0 kW, 15.31%, 13.95%, and 0.941. Adding operating state reduces overall RMSE by 42.8%; relative to the manufacturer curve (157.6 kW), the reduction is 48.0%.

Fig. 2 resolves this aggregate improvement temporally. Its monthly panel shows that XGB-MI remains below XGB-SI throughout the holdout rather than benefiting from one isolated month, while the paired seasonal panel shows the direction and magnitude of the reduction after averaging within each turbine. The larger separation during the monsoon period indicates that auxiliary operating-state variables are most useful when one wind speed corresponds to several controller states; it is not interpreted as proof of a specific atmospheric mechanism.

The largest monthly difference occurs in August (190.5 versus 66.8 kW). Table 1 shows improvement in every season and all 20 paired turbines. Monsoon gives the largest reduction, 89.6 kW or 56.3%, with a 95% bootstrap interval of 81.7–97.9 kW and $p < 0.001$.

The result is consistent with an operating-state explanation but does not establish a unique meteorological cause. At fixed wind speed, generator speed and blade angle distinguish control and curtailment states; directional encoding represents wake-sector changes; and temperature partly represents seasonal conditions. Attribution

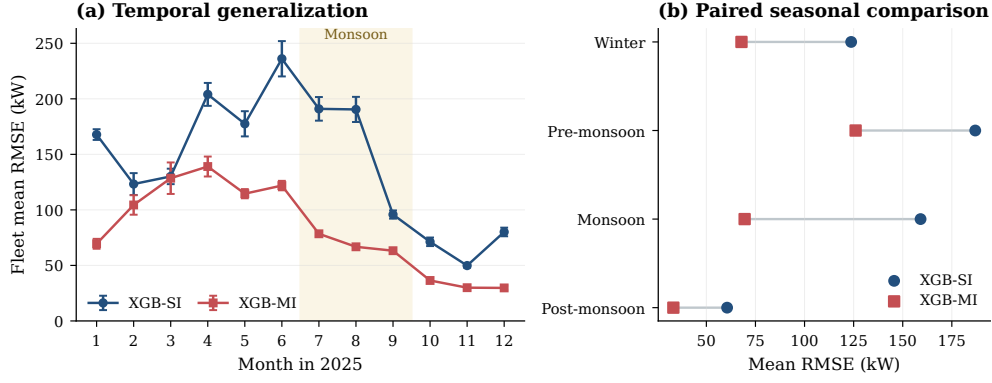


Fig. 2 Complete 2025 holdout: monthly fleet RMSE and turbine-paired seasonal RMSE.

Table 1 Paired seasonal RMSE on the 2025 holdout.

Season	SI (kW)	MI (kW)	Gain (kW)	Gain (%)
Winter	123.7	67.9	55.9	45.1
Pre-monsoon	186.9	126.0	60.9	32.6
Monsoon	159.1	69.5	89.6	56.3
Post-monsoon	60.6	33.2	27.4	45.2

20 turbines per season; paired $p < 0.001$.

to turbulence or atmospheric stability would require measurements absent from this export.

Supplementary Fig. S2 provides the corresponding curve-level interpretation for WTG1 and WTG13. XGB-SI supplies one output for a wind speed and therefore averages distinct control states, whereas XGB-MI predicts observed states separately and follows the central generating branch more closely when the scatter contains curtailed or limited-power bands. The displayed XGB-MI line is a binned conditional median used only for visualization; all reported errors use the original row-wise predictions. Consequently, the figure illustrates model behaviour without redefining the multi-input estimator as a unique wind-speed-only WPC.

3.3 Fitted site power shortfall

The 4PLF scale parameter of (6), the maximum power the model predicts as a fraction of rated power, is estimated per turbine on the 2023–2024 development record introduced in Section 5.1.2. Across the twenty-turbine ACT-2 fleet it ranges from 0.812 to 0.960, with fleet mean 0.877 ± 0.046 . The fitted upper level is 12.3% below the rated-power reference on average. This difference is descriptive, not an air-density attribution: the assumed site-to-standard density ratio is 1.17/1.225 (approximately 0.955), whereas wakes, controller behaviour, curtailment and data selection may also

128 alter the fitted level. The manufacturer curve therefore remains a reference rather than
129 a directly measured counterfactual.

130 3.4 Frozen ACT-2 Inference Under Data-Quality Variation

131 The U-Net is evaluated on 100 complete turbine-months per scenario. Panel A of
132 Table 2 shows that raw S1 input produces the lowest normalized RMSE, 0.02864, cor-
133 responding to 72.8 kW; S2 and S3 correspond to 74.9 and 74.3 kW. Their normalized
134 RMSE differs from S1 by only 2.9% and 2.1%. Stronger filtering therefore does not
135 improve primary accuracy.

136 Fig. 3 gives the shape-level evidence behind Panel A of Table 2. For both WTG1
137 and WTG13, the recovered curve follows the independent TP stars through the partial-
138 load ramp and rated-power plateau despite substantial low-power and curtailed GP
139 scatter. The close agreement across S1–S3 is consistent with the small scenario-level
140 RMSE variation in the master table. Historical classical curves provide visual context;
141 only the frozen U-Net response changes with 2026 input cleaning. This distinction
142 avoids presenting historical baselines as newly refitted 2026 models.

143 Supplementary Fig. S3 checks whether the primary result is stable across time
144 and machines. S1 normalized RMSE is 0.0376 ± 0.0107 in January, 0.0336 ± 0.0053
145 in February, 0.0242 ± 0.0037 in March, 0.0218 ± 0.0037 in April, and 0.0260 ± 0.0026
146 in May, corresponding to 95.6, 85.3, 61.4, 55.5, and 66.1 kW. The reduction from
147 January to April and modest May increase show temporal variation in scatter quality
148 and TP coverage. The ranked turbine panel remains compact and contains no isolated
149 high-error cluster, indicating that the fleet mean is not dominated by a few favourable
150 machines.

151 3.5 Zero-Shot Cross-Rating Transfer to ACT-1

152 Panel A of Table 2 reports the transfer result of the unchanged ACT-2 checkpoint
153 on the complete ACT-1 year. Raw S1 input gives the lowest RMSE (0.05440), MAE
154 (0.03741), MAPE (10.46%), and WMAPE (12.13%), with $R^2 = 0.9578$. S2 retains
155 94.54% of physically admissible records and increases RMSE by 3.16%; S3 retains
156 92.05% and increases RMSE by 3.82%. The corresponding SS@0.05 values are 73.82%,
157 74.78%, and 75.04% for S1–S3. The result therefore reproduces, on a different rating
158 class, the ACT-2 observation that statistical input cleaning is not required for the
159 trained scatter-to-capability mapping.

160 Supplementary Fig. S4 resolves the fleet result by turbine and scenario. Every S1
161 turbine has a median monthly RMSE below the published normalized S1 reference
162 of 0.0696; the S1 turbine medians range from 0.0390 to 0.0639. The fleet mean is
163 0.0731 in January, reaches its annual maximum of 0.0786 in February, and falls to
164 0.0288 in August. These changes should be interpreted as temporal variation in input
165 distribution and reference-curve coverage rather than checkpoint adaptation, because
166 the weights remain frozen throughout the year.

Table 2 Consolidated frozen-deployment and cross-checkpoint accuracy.

Evaluation	Pipeline/scenario	N	RMSE	MAE	MAPE (%)	WMAPE (%)	R^2
<i>A. Primary frozen deployment against the TP capability reference</i>							
ACT-2/2026	Site-adapted/S1	100	0.02864	0.01957	4.30	4.00	0.9946
	Site-adapted/S2	100	0.02947	0.02011	4.31	4.10	0.9941
	Site-adapted/S3	100	0.02923	0.01976	4.29	4.02	0.9943
ACT-1/2025	Site-adapted/S1	240	0.05440	0.03741	10.46	12.13	0.9578
	Site-adapted/S2	240	0.05612	0.03852	10.92	12.57	0.9533
	Site-adapted/S3	240	0.05648	0.03908	11.47	12.74	0.9528
<i>B. Shared-renderer checkpoint control on identical scoring cases</i>							
Yang sample ^a	Yang et al.	–	0.14882	0.08762	30.62	27.97	0.7683
	Site-adapted	–	0.14765	0.08487	29.92	27.09	0.7719
ACT-2/2025 ^b	Yang et al.	187	0.28204	0.19729	81.30	41.44	0.1491
	Site-adapted	187	0.02819	0.02010	5.82	4.54	0.9937
ACT-2/2026 ^c	Yang et al.	86	0.17656	0.11939	44.47	22.59	0.5789
	Site-adapted	86	0.03007	0.02017	4.19	4.00	0.9943
ACT-1/2025 ^c	Yang et al.	198	0.25444	0.17737	60.21	51.38	0.0774
	Site-adapted	198	0.05128	0.03564	9.89	11.21	0.9631

Metrics are normalized in both panels. Panel A reports macro-averaged turbine-month results; Panel B includes only paired successful cases. ^aOne common 80/20 split scored against held-out actual power. ^bDevelopment-domain diagnostic. ^cIndependent TP capability reference. Bold values indicate lower error or higher R^2 within each comparison.

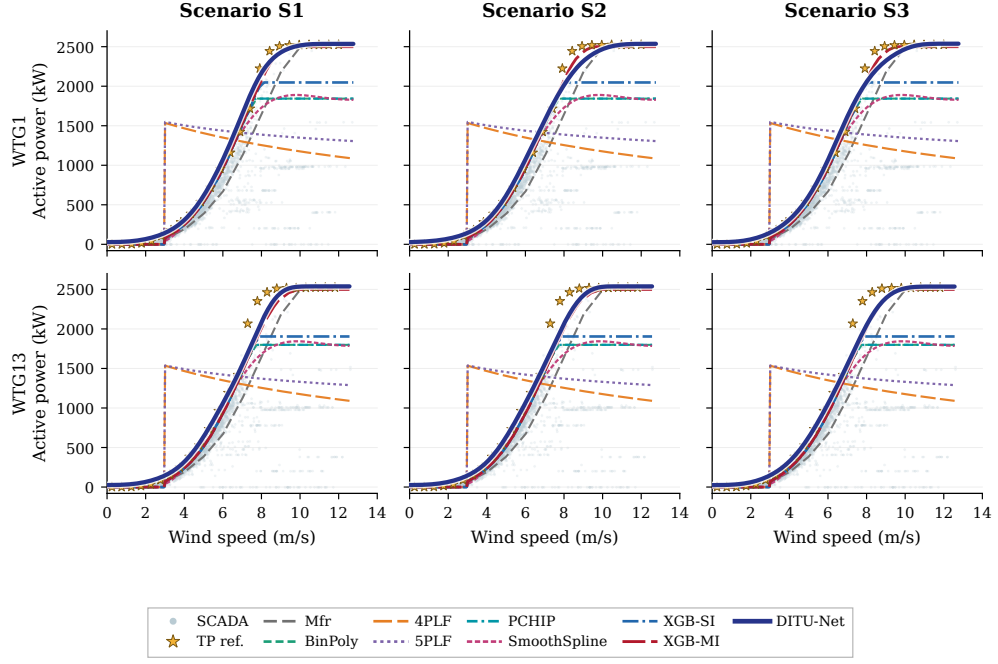


Fig. 3 WTG1 and WTG13 under S1–S3. TP stars are the scoring reference and are not supplied to the frozen U-Net.

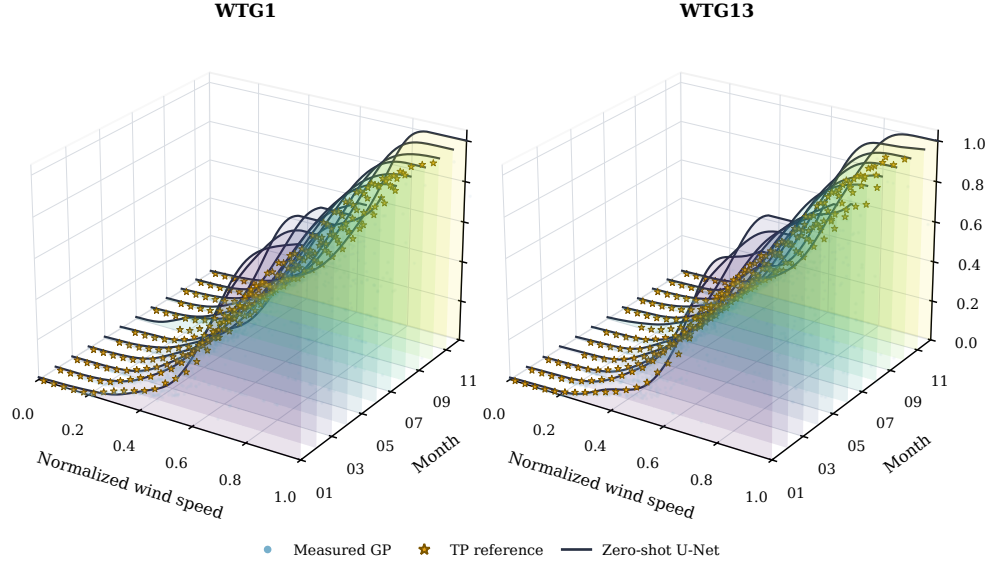


Fig. 4 Temporal capability landscapes for WTG1 and WTG13 during zero-shot ACT-1 deployment. Faint points are measured GP, stars are the independent TP scoring reference, and dark curves are predictions of the frozen ACT-2 U-Net.

The complementary wind-speed-month-power landscape in Fig. 4 resolves this temporal variation for WTG1 and WTG13. The transferred curves preserve the transition and rated regions while monthly GP scatter and TP-reference coverage change; no month-specific weight update is performed.

3.6 Insufficient-Data and Component Audits

Supplementary Table S2 tests whether recovery depends on dense high-wind data. In the insufficient-data pattern (IDP), observations above $q \in \{0.85, 0.90, 0.95\}$ are removed for six representative turbines. Classical models are refitted, whereas the frozen U-Net receives the matching truncated scatter. Its RMSE remains 35.97–38.62 kW and is lowest at every quantile; splines exceed 520 kW at $q = 0.85$ because their high-wind continuation is weakly constrained.

Disabling DKC changes RMSE by at most 10^{-5} , and 100% of ACT-2 and 96.7% of ACT-1 cases already reach rated power (DKC mode shares: reached-rated/genuine-plateau/truncated-to-saturation of 100/0/0% on ACT-2 and 96.7/2.5/0.8% on ACT-1); accuracy is therefore not caused by post-hoc saturation, as summarised in Supplementary Table S3.

A direct raw-GP spline has higher error than the full pipeline on both fleets (Supplementary Table S4). The $Q_{0.90}$ envelope helps under heavy ACT-1 curtailment but is less consistent across domains, while substituting the site-adapted reader into the Yang et al. checkpoint worsens every tested case (Supplementary Table S5). The network output and native reader operate as one pipeline, supporting the end-to-end interpretation of Table 2.

The audits retain their verified fixed-axis configuration; their absolute errors diagnose components and do not replace Table 2, Panel A.

3.7 Curtailment Energy-Loss Accounting

The recovered ACT-1 capability curves also support an operator-facing estimate of energy withheld below capability. For turbine i and month t ,

$$\widehat{EL}_{i,t} = \sum_{k=1}^{N_{i,t}} \max \left[0, \hat{P}_i(v_{i,k}) - P_{GP,i,k} \right] \Delta t, \quad (1)$$

where $\Delta t = 1/6$ h. A conservative RMSE-propagation quantity is

$$B_{i,t} = \text{RMSE}_S P_{\max,i} N_{i,t} \Delta t. \quad (2)$$

Supplementary Table S6 gives the monthly fleet aggregation. Estimated curtailed energy totals 19,593.8 MWh and ranges from 776.3 MWh in March to 2,383.9 MWh in September. During May–October it generally exceeds B by about a factor of two, while the curtailed-record fraction reaches 25.2% in October. This is capability-minus-delivered energy under the SCADA reference and ten-minute sampling, not a settlement or independently metered loss.

3.8 Bidirectional Cross-Checkpoint Verification

This checkpoint-generalization control tests whether the released Yang et al. checkpoint retains accuracy outside its training distribution; it does not test whether the two U-Net architectures differ in representational capacity. The sample-domain direction is therefore reported first as a fairness control, and the ACT-domain direction is then interpreted only as evidence about end-to-end transfer under distribution shift.

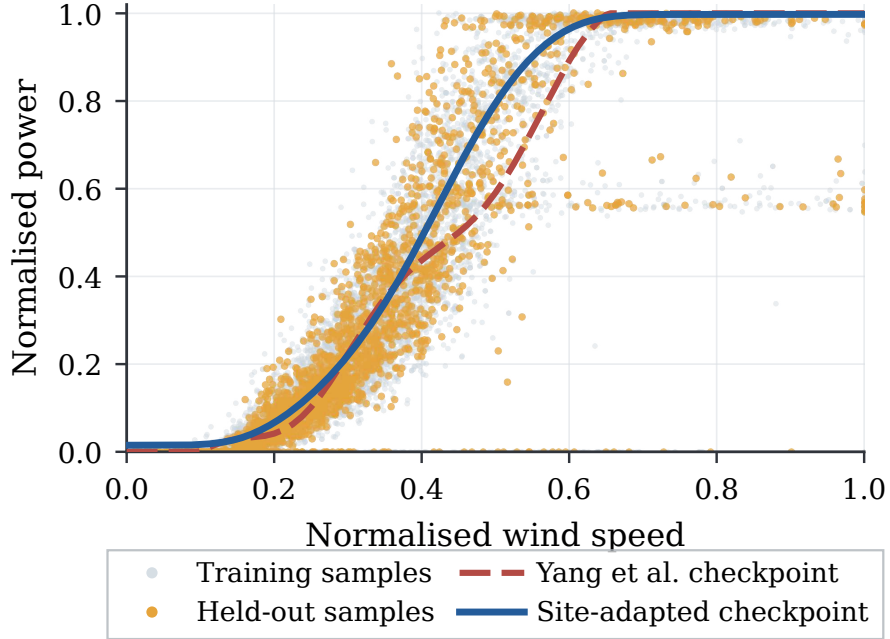


Fig. 5 Checkpoint comparison on the sample data reported by Yang et al. [4, 25]. Both pipelines receive the same training-scatter image and are evaluated against the identical held-out subset.

Fig. 5 gives the direct sample-domain comparison. Because the two networks receive the same rendered training scatter and are scored at the same held-out points, the result is not affected by split selection. The site-adapted pipeline obtains RMSE 0.1476, compared with 0.1488 for the Yang et al. pipeline; its MAE decreases from 0.0876 to 0.0849 and R^2 increases from 0.7683 to 0.7719. Thus, adaptation to ACT data does not produce a material loss on the external sample used by the published implementation.

Panel B of Table 2 extends the control to paired successful ACT-domain cases. On temporally held-out ACT-2/2026, the site-adapted pipeline reduces RMSE from 0.1766 to 0.0301; on zero-shot ACT-1/2025, it reduces RMSE from 0.2544 to 0.0513. The corresponding reductions are 83.0% and 79.8%, while MAPE decreases from 44.47% to 4.19% and from 60.21% to 9.89%, respectively. The ACT-2/2025 row is included only as a development-domain diagnostic and must not be interpreted as an independent holdout. The Yang et al. native reader returns valid curves for 187 of 240 ACT-2/2025,

86 of 100 complete ACT-2/2026, and 198 of 240 ACT-1 turbine-months; the site-adapted reader returns valid curves for all corresponding cases. The small differences between the paired site-adapted values and the primary values in Panel A arise because the control re-renders both checkpoints using one common image rule and uses the paired subset. Panel A remains the authoritative deployment result.

Because the Yang sample and the three ACT datasets are separate evaluation domains rather than points on a continuous site axis, Table 2 remains the appropriate inferential comparison. Fig. 5 provides the directly interpretable curve-level fairness control, whereas the ACT conclusions are drawn from paired successful cases and reader-success audits rather than from visual interpolation between domains.

3.9 Architectural and Methodological Differences

Three design differences may contribute to the accuracy gap documented in Table 2, Panel B, and in Supplementary Table S5. First, the *training target*: the Yang et al. checkpoint trains on synthetic double-exponential curves [4], whereas the site-adapted network trains on real ACT-2 scatter against a TP-derived capability target (Section 5.2.2), so curtailment is part of the training distribution rather than an artifact encountered only at inference. Second, *pixel extraction*: the published pipeline reads the single darkest pixel per column [4], so one curtailment-shifted pixel can displace the extracted point; the site-adapted reader instead takes an intensity-weighted centroid over all dark pixels in a column (Section 5.2.4), damping the influence of individual outlier pixels. Third, *curve fitting*: the published pipeline fits an unconstrained polynomial and enforces monotonicity only post hoc via its own domain-knowledge correction, whose sensitivity to persistent curtailment is reflected in the S1-to-S3 scenario swings for the Yang et al. checkpoint in Table 2, Panel B; the site-adapted pipeline instead uses an isotonic pre-fit followed by a monotone penalized B-spline with monotonicity enforced as a hard constraint rather than corrected after the fact.

Section 3.6 shows that readout substitution alone is insufficient: Supplementary Table S5 shows that substituting the site-adapted reader into the Yang et al. checkpoint’s own output *worsens* every tested case rather than improving it, so the published network’s image output is not shaped correctly for Jhampir scatter regardless of which reader transcribes it. The accuracy gap in Table 2, Panel B is consistent with an important contribution from the learned mapping and training distribution, but the present control does not quantify the separate causal contributions of training and readout.

4 Discussion

The estimators serve different purposes. XGB-MI is accurate when generator speed, pitch, direction, and temperature are available. It supports residual monitoring and state assessment but is not a standalone WPC when future auxiliary variables are unknown. The image model receives wind speed and GP scatter for a turbine-month and returns a monotone capability curve, supporting post-event curtailment assessment and fleet comparison.

S1 avoids offline anomaly removal and slightly outperforms S2/S3, reducing processing and the risk of discarding legitimate curtailed states. TP remains a controller-generated capability reference rather than an independent IEC measurement; error therefore quantifies agreement with controller capability, not absolute aerodynamic efficiency.

The ACT-1 result extends this operational interpretation across rated power. A common image coordinate system permits the ACT-2-trained network to process a 1.5 MW fleet without weight adaptation, while explicit nameplate scaling preserves physical units. This should be described as zero-shot model transfer with input-only axis calibration, not as fully calibration-free deployment.

The bidirectional control narrows the interpretation further. Comparable performance on the Yang sample indicates that site adaptation did not erase the generic scatter-to-curve capability represented in the released checkpoint. The much larger ACT-domain separation instead shows the value of matching the learned image prior and curve reader to the operating regime, consistent with the mechanistic account in Section 3.9. Because each pipeline retains its native read-out and the published reader fails on a subset of ACT turbine-months, the experiment supports an end-to-end robustness conclusion; it does not attribute the entire difference to U-Net weights alone.

Operationally, the recovered capability curve can support turbine-to-fleet comparison, post-event curtailment assessment, and screening of persistent underperformance before a more detailed engineering investigation. The 1.3 s fleet inference time leaves substantial margin within the ten-minute reporting cadence, so the image model can be deployed as an offline or supervisory analytics layer without entering the turbine control loop.

The verified audits strengthen this interpretation: stable truncated-data error shows that a dense rated region is unnecessary, and DKC is not the source of accuracy. Energy accounting converts the capability curve into an operational screening quantity, but TP and model RMSE do not replace settlement-grade metering.

4.1 Limitations

The full-year field results concern one wind corridor and two rating classes. Evaluation on the open-source Yang et al. sample [4, 25] supplies an external distribution check, but that compact sample is not a third operational fleet and does not establish geographical or controller transfer beyond Jhimpir. ACT-1 transfer is transductive because target-year wind speed fixes each input axis. TP inherits controller assumptions and is not an IEC reference; likewise, standardized-pressure correction does not measure humidity, turbulence, or stability. The 4PLF upper-level shortfall in Section 3.3 does not isolate an air-density effect. XGB-MI is concurrent conditional estimation, not a wind-speed-only forecast. Sample-domain errors use held-out actual power, whereas ACT errors use TP, so their absolute values are not comparable. The checkpoints retain native readers and are compared on paired successful cases; the result is an end-to-end pipeline effect, not a pure weight ablation. Energy loss is a ten-minute, model-based capability shortfall rather than a settlement estimate. Plateau and aggregated-energy uncertainty remain to be modelled probabilistically.

The train-once, GP-only deployment demonstrated here is therefore established across the 2.5 MW and 1.5 MW fleets within the shared Jhimpir meteorological and grid-operational regime. It should be interpreted as cross-rating generalization within that regime, not as unconditional cross-site, cross-climate, or controller-independent deployability. Validation in additional wind corridors, turbine-control philosophies, and curtailment regimes remains necessary.

4.2 Conclusions and outlook

This paper consolidated classical WPC benchmarking, conditional gradient boosting, and site-adapted image-to-curve recovery on 20 turbines. On the complete 2025 hold-out, XGB-MI reduced fleet RMSE from 143.1 to 81.9 kW relative to XGB-SI. The largest seasonal reduction occurred during monsoon, from 159.1 to 69.5 kW, and was observed across all paired turbines. A fleet-level 4PLF analysis found a fitted upper-level shortfall of 12.3% relative to rated power; the available data cannot assign that shortfall to air density alone.

On 100 unseen 2026 ACT-2 turbine-months, the frozen U-Net recovered controller capability from raw GP scatter with normalized RMSE 0.0286, MAPE 4.30%, and $R^2 = 0.9946$. The unchanged checkpoint was then transferred to 240 turbine-months from a 1.5 MW ACT-1 fleet, where S1 achieved normalized RMSE 0.0544 and $R^2 = 0.9578$. GP-only cleaning did not improve primary RMSE at either rating. In the shared-renderer cross-checkpoint control, the site-adapted pipeline matched the published pipeline on its sample domain while reducing paired-case RMSE by 83.0% on ACT-2/2026 and 79.8% on ACT-1/2025, a gap traced to differences in training target, pixel extraction, and curve fitting rather than network capacity alone. Auxiliary state therefore improves contemporaneous power estimation, whereas image recovery provides a transferable one-dimensional capability curve with reduced dependence on hand-designed filtering. Its 66.4 ms turbine-month latency is compatible with supervisory fleet analytics.

These results support use of the frozen model as a fleet-analytics layer without per-turbine refitting or offline anomaly removal, at least within the shared operational and meteorological regime examined here; they do not imply universal transfer across wind corridors or controller families.

Complementary audits retain 35.97–38.62 kW U-Net RMSE under 0.85–0.95 wind-speed truncation. DKC changes normalized RMSE by at most 10^{-5} , and raw-GP splines are 6.8 and 2.6 times worse on ACT-2 and ACT-1. The complete ACT-1 year yields 19,593.8 MWh of estimated capability shortfall, extending curve recovery to curtailment screening.

Future work will evaluate both formulations in an independent wind regime, isolate network and read-out contributions through a crossed ablation, reduce checkpoint size for edge deployment through quantization and pruning, and quantify curve uncertainty against met-mast or lidar observations.

346 5 Methods

347 5.1 Wind power curve and data arrangement

348 5.1.1 Wind Power Curve

349 The aerodynamic power available to a rotor is

$$P_w(v) = \frac{1}{2} \rho A v^3, \quad (3)$$

350 where v is wind speed, ρ is air density, and A is rotor swept area. Electrical output is
351 lower because of aerodynamic, mechanical, and electrical losses. The operating WPC
352 contains a near-zero region below cut-in speed, a nonlinear partial-load region, and a
353 rated-power plateau before cut-out.

354 Measured SCADA data differ from this ideal relation because normal operation,
355 faults, and curtailment coexist. Controller theoretical power (TP) is used to construct
356 a capability reference. In each of 25 wind-speed bins B_i ,

$$y_i^{\text{ref}} = Q_{0.90}(P_{\text{TP}} \mid v \in B_i), \quad (4)$$

357 so that dense low-power curtailed observations do not depress capability.

358 5.1.2 Study Site and Temporal Partitions

359 The ACT-2 site is a 50 MW wind farm in the Jhimpir wind corridor of Sindh, Pakistan.
360 It comprises 20 Goldwind GW121/2500 turbines rated at 2.5 MW. The ten-minute
361 SCADA export contains wind speed, grid active power (GP), TP, generator speed,
362 blade angle, temperature, wind direction, power-limit state, timestamp, and turbine
363 identifier.

364 Supplementary Table S7 makes the experimental roles explicit. The 2023–2024
365 records are used only to develop XGBoost, the complete 2025 year is its temporal
366 holdout, and the 2023–2025 image set is used only for U-Net development. The 2026
367 ACT-2 records and the ACT-1 fleet are therefore inference-only partitions. This sep-
368 aration prevents test-period observations from affecting model fitting or checkpoint
369 selection and distinguishes temporal generalization from cross-rating transfer.

370 The ACT-2 site has mean wind speed 5.70 m/s and site air density 1.17 kg/m³
371 against the 1.225 kg/m³ ISA reference; NEPRA-directed grid curtailment removes
372 approximately 17–19% of raw records per turbine at ACT-2. The ACT-1 fleet, eval-
373 uated only in Section 3 for zero-shot transfer, operates at a lower mean wind speed
374 of 5.13 m/s with a monthly curtailment fraction averaging 9.5% (range 2.7–25.2%
375 across the twelve months of 2025, Supplementary Table S6). Nacelle temperature is
376 not recorded in the ACT-1 SCADA export, so air density cannot be computed via (5)
377 for that site; the ACT-1 evaluation in Section 3.5 is accordingly restricted to the
378 wind-speed and grid-power channels shared with ACT-2.

379 The XGBoost experiment uses 2023–2024 for development and the complete 2025
380 year for testing. Identical 2025 timestamps are used for its single- and multi-input
381 variants. The U-Net uses grouped turbine-month image pairs from 2023–2025; no

382 observation from 2026 is used for training, checkpoint selection, hyperparameter
 383 selection, normalization, or cleaning decisions. January–May 2026 are complete farm
 384 months. Incomplete June records remain outside primary averages.

385 Cross-rating transfer is evaluated on the ACT-1 record, comprising 1,043,701 ten-
 386 minute observations from twenty 1.5 MW turbines over the complete 2025 calendar
 387 year. ACT-1 contributes no training image, validation sample, checkpoint decision, or
 388 weight update. It is used only after the ACT-2 checkpoint has been frozen.

389 5.1.3 Data-Quality Scenarios

390 Three 2026 scenarios evaluate sensitivity to deployable preprocessing. S1 retains
 391 physically admissible GP scatter without statistical removal. S2 adds a robust median-
 392 absolute-deviation band in 25 wind-speed bins. S3 adds a binwise interquartile-range
 393 filter. S2 and S3 use only (v, P_{GP}) available at deployment; TP is reserved for scoring.

394 For contextual classical comparisons, corrected wind speed is

$$\rho = \frac{P_0}{R_a T}, \quad v_c = v \left(\frac{\rho}{\rho_0} \right)^{1/3}, \quad (5)$$

395 where $P_0 = 101325$ Pa, $R_a = 287.05$ J/(kg·K), and $\rho_0 = 1.225$ kg/m³. Because site
 396 pressure and humidity are unavailable, this is a standardized-pressure correction rather
 397 than direct atmospheric-density measurement.

398 5.2 Model specifications

399 The classical and gradient-boosting benchmark is not incidental context but the evi-
 400 dentiary basis for the paper’s central claim: image-based capability recovery must be
 401 compared with the methods used in current WPC practice before its accuracy and
 402 deployment value can be interpreted. A common, leakage-controlled evaluation against
 403 manufacturer curves, parametric logistic models, non-parametric interpolants, splines,
 404 and gradient boosting establishes the operating envelope of established approaches
 405 surveyed in Section 2. The U-Net result can therefore be assessed as an improvement
 406 over relevant alternatives rather than as a standalone accuracy value.

407 5.2.1 Classical Curves and Gradient Boosting

408 The benchmark includes the manufacturer curve, bin-mean polynomial regression
 409 (BinPoly), four- and five-parameter logistic functions (4PLF and 5PLF), piecewise
 410 cubic Hermite interpolation (PCHIP), a smoothing spline (SmSpl), and XGBoost.
 411 PCHIP and SmSpl are fitted independently to isotonic-corrected wind-speed-bin
 412 statistics.

413 Writing $x = v/v_{co}$, the logistic curves are

$$P_{4\text{PLF}}(x) = P_f + \frac{P_s - P_f}{1 + |c/x|^b}, \quad (6)$$

414

$$P_{\text{5PLF}}(x) = P_f + \frac{P_s - P_f}{(1 + |c/x|^b)^g}, \quad (7)$$

415 where P_f and P_s are lower and upper levels, b and c control the transition, and g
 416 controls asymmetry. Parameters are estimated by differential evolution minimising
 417 mean-squared error.

418 XGB-SI receives corrected wind speed and its second- and third-order terms. XGB-
 419 MI additionally receives blade angle, generator speed, temperature, sine/cosine wind-
 420 direction encoding, and wind-speed interactions with generator speed and blade angle.
 421 Missing values are replaced using development-set medians only. The fixed estimator
 422 uses 350 trees, depth 5, learning rate 0.045, row and column fractions 0.85, minimum
 423 child weight 5, and L_1/L_2 regularization 0.05/1.5.

424 XGB-MI estimates $\hat{P}(v, \mathbf{z})$ rather than a unique $P(v)$. Its displayed response is
 425 obtained from development-set predictions by taking a median in 0.25 m/s wind-speed
 426 bins and applying isotonic interpolation. Pointwise metrics use unmodified row-wise
 427 predictions.

428 5.2.2 Site-Adapted U-Net

429 The proposed model learns

$$f_\theta : I(v, P_{\text{GP}}) \longrightarrow I(v, P_{\text{TP,ref}}), \quad (8)$$

430 where both channels are 256×256 images. The objective is

$$\mathcal{L}(\theta) = \|I_{\text{TP,ref}} - f_\theta(I_{\text{GP}})\|_2^2. \quad (9)$$

431 This formulation deliberately uses an offline-training/inference information asymme-
 432 try. TP is logged by the turbine controller in the historical SCADA archive and can
 433 therefore supervise capability-target construction during model development, but it is
 434 not an independently metered quantity and may be unavailable, delayed, or unreliable
 435 when the frozen model is deployed. The U-Net consequently learns a GP-scatter-
 436 to-TP-derived-capability mapping during development, whereas every deployment
 437 experiment in Sections 5.3 and 3 supplies only wind speed and GP. No TP value
 438 is observed, required, or used for cleaning, normalization, or prediction after check-
 439 point freezing; TP is introduced only after inference to form the independent scoring
 440 reference in (4).

441 Real turbine-month pairs form the training core. Curtailment-shaped, stacked, and
 442 sparse outlier patterns are injected only into the input image while the TP-derived
 443 target remains unchanged. All augmentations of one turbine-month retain the same
 444 group label, preventing related samples from entering both training and validation
 445 partitions.

446 Wind speed and power are normalized by fixed development-period 99.5th-
 447 percentile constants. A 15% group-disjoint image test set is held out, followed by
 448 five-fold GroupKFold validation with three seeds per fold. The network uses Adam,
 449 learning rate 10^{-3} , weight decay 10^{-5} , batch size 16, dropout 0.10, and early stopping

450 within 30 epochs. Mean validation MSE is 0.01136 ± 0.00068 and held-out image-
 451 test MSE is 0.01033. The selected checkpoint is frozen before 2026 inference. This
 452 procedure is formalised in Supplementary Algorithm S1.

453 5.2.3 Architectural Rationale

454 The encoder-decoder is selected for the structure of WPC recovery rather than only
 455 by analogy with generic image restoration [26]. The input contains a global S-shaped
 456 capability trajectory together with local features at cut-in, the knee, and the rated-
 457 power transition. Successive encoder levels enlarge the receptive field and represent the
 458 global trajectory, whereas skip connections return high-resolution spatial information
 459 to the decoder for local boundary placement. The same multi-scale representation
 460 supports separation of a dense capability locus from structured curtailment bands and
 461 sparse outliers without a hand-designed test-period target filter.

462 Normalization to a common $[0, 1]^2$ image coordinate makes the learned mapping
 463 shape-based rather than tied to one absolute power rating. This is the mechanism
 464 evaluated by the frozen 2.5-to-1.5 MW transfer in Section 3. The shared-renderer con-
 465 trol later compares complete pipelines because the two checkpoints retain their native
 466 curve readers; accordingly, it supports end-to-end robustness but is not presented as
 467 a pure weight ablation.

468 Fig. 6 summarizes the complete inference path. The rendered SCADA scatter
 469 enters the contracting encoder, progressively compressed features capture the global
 470 curve geometry, and skip connections return spatial detail to the decoder. The final
 471 1×1 convolution produces a single WPC image, which is converted to a monotone
 472 numerical curve by the reader described in Section 5.2.4. The figure also clarifies that
 473 TP appears in the development target but is not an inference input.

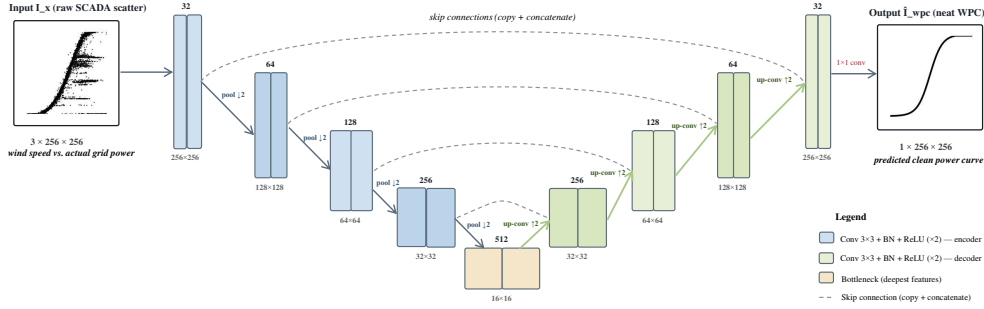


Fig. 6 Architecture and data flow of the site-adapted U-Net.

474 5.2.4 Curve Extraction

475 The raster is converted to a numerical curve using an intensity-weighted dark-pixel
 476 centroid, median filtering, isotonic pre-fitting, and a monotone penalized B-spline.
 477 Disjoint fragments after the principal upper plateau are rejected. A smooth Hermite

478 completion (domain-knowledge correction, DKC) is activated only when the retained
 479 upper branch remains positively sloped below its terminal level,

$$\text{mode} = \begin{cases} \text{reached_rated}, & y_{\text{top}} \geq 0.95, \\ \text{truncated} \rightarrow \text{saturated}, & y_{\text{top}} < 0.95 \text{ and } s_{\text{top}} > \tau, \\ \text{genuine_plateau}, & \text{otherwise,} \end{cases} \quad (10)$$

480 with $\tau = \max(0.5, 0.25 \cdot s_{\text{max}})$. No 2026 TP is used during extraction or normaliza-
 481 tion; DKC firing frequency and its effect on accuracy are quantified in Section 3.6.
 482 The complete deployment procedure, applied identically to every turbine-month in
 483 Sections 3, is formalised in Supplementary Algorithm S2.

484 5.2.5 Computational Cost

485 Supplementary Table S9 reports the stored model burden and measured deployment
 486 latency. The U-Net contains 7,763,041 trainable parameters and occupies 29.6 MB. Its
 487 66.4 ms turbine-month inference time is higher than the sub-0.13 ms classical timings,
 488 but a 20-turbine fleet requires only approximately 1.3 s, well below the ten-minute
 489 SCADA reporting interval. Training is a one-time offline workload of five folds, three
 490 seeds, and at most 30 epochs; it is not repeated for a new turbine-month. A hardware-
 491 specific total training wall time was not retained, so the reproducible workload is
 492 reported instead of an inferred duration.

493 The latency comparison is operationally asymmetric: each classical curve requires
 494 turbine-specific fitting to historical data for a new deployment cycle, whereas one
 495 frozen U-Net checkpoint is reused across turbine-months, cleaning scenarios, years,
 496 and the second fleet without refitting. In S1, its 55.76 kW fleet RMSE is 49.9% lower
 497 than the 111.29 kW error of the best classical baseline in Fig. 1; the additional latency
 498 therefore purchases a substantial fleet-wide accuracy gain while removing repeated
 499 turbine-specific fitting. The 66.4 ms value was measured in the Kaggle inference envi-
 500 ronment used for the reported experiments; because the exact processor/accelerator
 501 allocation of that timing cell was not retained, it is treated as an environment-
 502 specific deployment reference rather than a hardware-normalized benchmark. Since
 503 turbine-month images are processed independently, sequential cost is expected to scale
 504 approximately linearly with fleet size and can additionally be reduced by parallel
 505 batching. Thus, the relevant deployment tradeoff is accuracy and engineering effort
 506 per reporting cycle rather than raw single-curve latency alone.

507 5.3 Experimental protocol

508 Each classical curve is fitted per turbine on its historical development record. The U-
 509 Net is trained once and applied to each 2026 turbine-month without weight updating.
 510 For S1–S3, a fresh image-model curve is recovered from the corresponding GP scatter
 511 while the checkpoint and development-period normalization remain fixed.

512 For ACT-1 zero-shot deployment, the 1.5 MW nameplate rating defines the power-
 513 axis scale, $P_{\text{max}} = 1.005 \times 1500$ kW. The wind-speed scale is fixed per turbine as

1.005 times the 99.5th percentile of the available wind-speed input. This is an input-only, transductive axis calibration: it uses no TP value but observes the target-year wind-speed distribution. The ACT-1 TP channel is introduced only after prediction to form (4). S2 and S3 retain the same GP-only median-absolute-deviation and interquartile-range filters used in corrected ACT-2 inference. The resulting experiment contains 12 months $\times$ 20 turbines = 240 turbine-months per scenario and zero ACT-1 gradient steps.

Accordingly, the transfer is zero-shot with respect to model learning but is not fully calibration-free: no gradient step, GP target, or TP value influences the checkpoint, while the unlabeled target-year wind-speed distribution fixes the input axis. Stating this design choice in the protocol distinguishes deliberate input-only transductive calibration from target-assisted adaptation; its deployment scope is discussed further in Section 4.1.

Let y_j and $\hat{y}_j$ denote reference and estimated power. The metrics are

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2}, \quad (11)$$

$$\text{MAE} = \frac{1}{N} \sum_{j=1}^N |y_j - \hat{y}_j|, \quad (12)$$

$$\text{MAPE} = \frac{100}{N_+} \sum_{j: y_j > 0.05 P_r} \left| \frac{y_j - \hat{y}_j}{y_j} \right|. \quad (13)$$

WMAPE and R^2 are also reported. Winter is December–February, pre-monsoon March–June, monsoon July–September, and post-monsoon October–November. Metrics are first calculated by turbine-month and then averaged by turbine. Paired XGB differences use a two-sided Wilcoxon signed-rank test and a 5,000-resample turbine bootstrap interval.

5.3.1 Bidirectional Checkpoint Control

The checkpoint control separates transfer direction from the primary deployment experiments. The checkpoint released by Yang et al. [4, 25] and the site-adapted checkpoint receive the same 256×256 scatter image, rendered with the marker-size rule of [4]. Each network is followed by its native curve reader, so the comparison concerns the complete published and site-adapted pipelines rather than network weights alone.

Two directions are evaluated. First, both pipelines receive an identical 80/20 split of the sample data released with [4]; the training subset forms the image and the held-out measured-power subset forms the scoring target. Second, both pipelines are applied to ACT-2/2025, ACT-2/2026, and ACT-1/2025 under raw S1 input, with TP introduced only after prediction through (4). ACT-2/2025 is a development-domain diagnostic for the site-adapted checkpoint, whereas ACT-2/2026 and ACT-1/2025 retain their temporal-holdout and cross-fleet roles. Cross-checkpoint means are calculated only on turbine-months for which both native readers return a valid curve. Reader failures are retained in the audit rather than silently assigned an error.

550 Data availability

551 The ACT-1 and ACT-2 SCADA data analysed in this study are not publicly avail-
552 able because of commercial confidentiality restrictions. Requests for access may be
553 directed to the corresponding author at stauqeer.msee24mcs@student.nust.edu.pk
554 and are subject to approval by the data provider. The publicly available DITU-
555 Net sample data and code used for the checkpoint comparison are available at
556 <https://github.com/IkeYang/DITU-net>.

557 References

- 558 [1] Mehmood, Z. & Wang, Z. Hybrid iforest-dbscan for anomaly detection and wind
559 power curve modelling. *Expert Systems With Applications* **289**, 128381 (2025).
- 560 [2] Mazen, F. M. A., Shaker, Y. & Abul Seoud, R. A. A two-stage model for
561 outlier detection and power prediction of wind turbine using SCADA dataset.
562 *International Journal of Energy Research* **2025**, 2527561 (2025).
- 563 [3] Qiao, Y., Han, S., Zhang, Y., Liu, Y. & Yan, J. A multivariable wind turbine
564 power curve modeling method considering segment control differences and short-
565 time self-dependence. *Renewable Energy* **222**, 119894 (2024).
- 566 [4] Yang, L., Wang, L. & Zhang, Z. Generative wind power curve modeling
567 via machine vision: A deep convolutional network method with data-synthesis-
568 informed-training. *IEEE Transactions on Power Systems* **38**, 1111–1124 (2023).
- 569 [5] Bilendo, F. *et al.* Applications and modeling techniques of wind turbine power
570 curve for wind farms—a review. *Energies* **16**, 180 (2023).
- 571 [6] Li, F., Wang, H., Wang, D., Liu, D. & Sun, K. A review of wind power prediction
572 methods based on multi-time scales. *Energies* **18**, 1713 (2025).
- 573 [7] Paik, C., Chung, Y. & Kim, Y. J. Power curve modeling of wind turbines through
574 clustering-based outlier elimination. *Applied System Innovation* **6**, 41 (2023).
- 575 [8] Yao, Q. *et al.* Power curve modeling for wind turbine using hybrid-driven outlier
576 detection method. *Journal of Modern Power Systems and Clean Energy* **11**,
577 1115–1125 (2023).
- 578 [9] Marti-Puig, P., Hernández, J. Á., Solé-Casals, J. & Serra-Serra, M. Enhancing
579 reliability in wind turbine power curve estimation. *Applied Sciences* **14**, 2479
580 (2024).
- 581 [10] Nasery, P. & Aziz Ezzat, A. Yaw-adjusted wind power curve modeling: A local
582 regression approach. *Renewable Energy* **202**, 1368–1376 (2023).

- [11] Bilendo, F., Badihi, H., Lu, N., Cambron, P. & Jiang, B. A normal behavior model based on power curve and stacked regressions for condition monitoring of wind turbines. *IEEE Transactions on Instrumentation and Measurement* **71**, 3520013 (2022).
- [12] Wang, Y. *et al.* A novel data-driven deep learning approach for wind turbine power curve modeling. *Energy* **270**, 126908 (2023).
- [13] Yin, W. *et al.* Advanced power curve modeling for wind turbines: A multivariable approach with SGBRT and grey wolf optimization. *Energy Conversion and Management* **332**, 119680 (2025).
- [14] Zhang, Z., Ouyang, T. & Shen, X. Risk-regularization optimization for learning wind power curve. *IEEE Transactions on Emerging Topics in Computational Intelligence* (2025).
- [15] Haddad, K. An integrated goodness-of-fit and vine copula framework for wind-speed distribution selection and turbine power-curve assessment in new south wales and southern east queensland. *Atmosphere* **16**, 1068 (2025).
- [16] Joshi, R., Schmehl, R. & Kruijff, M. Power curve modelling and scaling of fixed-wing ground-generation airborne wind energy systems. *Wind Energy Science* **9**, 2195–2215 (2024).
- [17] Wu, C., Yang, L. & Zhang, Z. Autonomous wind turbine performance curve modeling based on SCADA data: A vision intelligence powered method. *Energy and AI* **22**, 100645 (2025).
- [18] Michalakopoulos, V., Zakynthinos, A., Sarmas, E., Marinakis, V. & Askounis, D. Hybrid short-term wind power forecasting model using theoretical power curves and temporal fusion transformers. *Renewable Energy* **256**, 124008 (2026).
- [19] Gu, G. *et al.* Wind power forecasting based on a machine learning model: considering a coastal wind farm in zhejiang as an example. *International Journal of Green Energy* **21**, 2551–2558 (2024).
- [20] Wang, D. *et al.* Enhancing wind power forecasting accuracy through lstm with adaptive wind speed calibration (c-lstm). *Scientific Reports* **15**, 5352 (2025).
- [21] Rajaperumal, T. A. & Columbus, C. C. Enhanced wind power forecasting using machine learning, deep learning models and ensemble integration. *Scientific Reports* **15**, 20572 (2025).
- [22] Li, W., Yang, T., Yang, J. & Peng, L. Predicting ultra-short-term wind power combinations under extreme weather conditions. *IEEE Access* **13** (2025).

- 617 [23] Cui, Y., Cheng, C., Zhu, Y., Meng, D. & Chen, Z. A novel two-step framework
618 for day-ahead wind power probabilistic forecasting considering uncertainties and
619 ramp events. *International Journal of Electrical Power and Energy Systems* **169**,
620 110780 (2025).
- 621 [24] Cao, M. & Yu, J. Dynamic setting method of assessment indicators for power
622 curves of renewable energy sources considering scarcity of reserve resources.
623 *Journal of Modern Power Systems and Clean Energy* **12**, 101–114 (2024).
- 624 [25] Yang, L., Wang, L. & Zhang, Z. DITU-net: Official code repository. GitHub
625 repository (2023). <https://github.com/IkeYang/DITU-net>. Accessed: Jul. 3,
626 2026.
- 627 [26] Ronneberger, O., Fischer, P. & Brox, T. in *U-Net: Convolutional networks*
628 *for biomedical image segmentation* (eds Navab, N., Hornegger, J., Wells, W. M.
629 & Frangi, A. F.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Vol. 9351 of *Lecture Notes in Computer Science* 234–241
630 (Springer, Cham, 2015).
631

632 **Supplementary information.** Supplementary figures, tables and algorithms are
633 supplied as a separate PDF.

634 **Acknowledgements.** The authors thank Ali Hassan, HydroChina International
635 Engineering Co., Ltd., for facilitating authorized access to the ACT-1 and ACT-2
636 SCADA datasets.

637 **Funding.** The authors received no specific funding for this study.

638 **Author contributions.** S.M.H.T. and S.N. conceived the study. S.M.H.T. devel-
639 oped the methodology, software, data curation, analyses, validation and visualizations
640 and drafted the manuscript. S.N. and A.K. contributed to validation. A.J. provided
641 resources and, with S.N., supervised the work. S.N., A.K. and A.J. reviewed and edited
642 the manuscript. All authors approved the final manuscript.

643 **Competing interests.** The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementary.pdf](#)