

Supplementary information: 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;

Supplementary figures, tables and algorithms

These items are cited in the main manuscript. Numerical values and images are carried over from the supplied CMES source.

Table S1 Supplementary Acronyms used in this manuscript.

Acronym	Definition
4PLF	Four-Parameter Logistic Function
5PLF	Five-Parameter Logistic Function
ACT-1, ACT-2	Pseudonymized identifiers for the two wind farm sites studied in this work: ACT-1 (30 MW installed capacity) and ACT-2 (50 MW installed capacity)
BinPoly	Bin-Mean Polynomial Regression
DKC	Domain-Knowledge Correction
GP	Grid Power
IDP	Insufficient Data Pattern
ISA	International Standard Atmosphere
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
RMSE	Root Mean Square Error
SCADA	Supervisory Control and Data Acquisition
SmSpl	Smoothing Spline
SS@ α	α -Shooting Score
TP	Theoretical Power
WMAPE	Weighted Mean Absolute Percentage Error
WPC	Wind Power Curve
WTG	Wind Turbine Generator
XGB-SI	Gradient Boosting, single-input (wind speed only)
XGB-MI	Gradient Boosting, multi-input (wind speed, generator speed, blade pitch, direction, temperature)

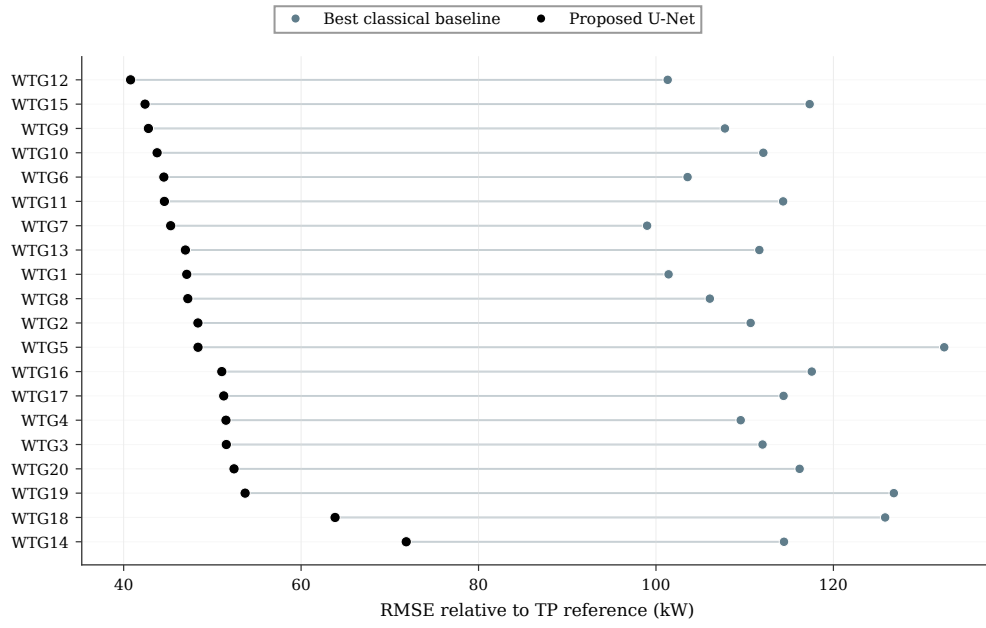


Fig. S1 Supplementary Per-turbine RMSE of the U-Net and best classical baseline under S3.

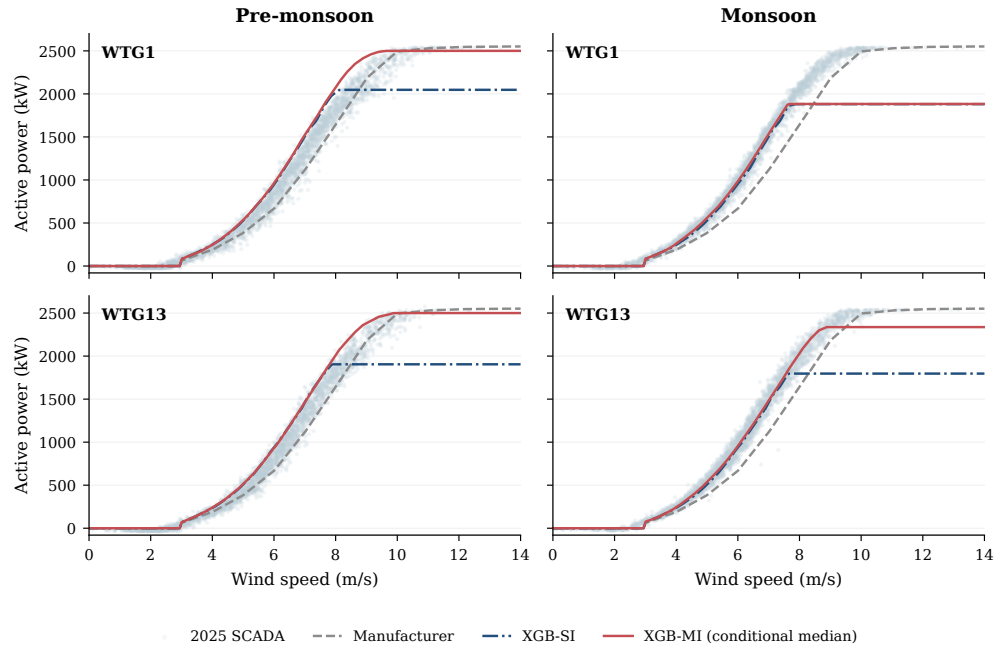


Fig. S2 Supplementary Representative 2025 behaviour for WTG1 and WTG13. XGB-MI is a development-set conditional median, not joined row-wise predictions.

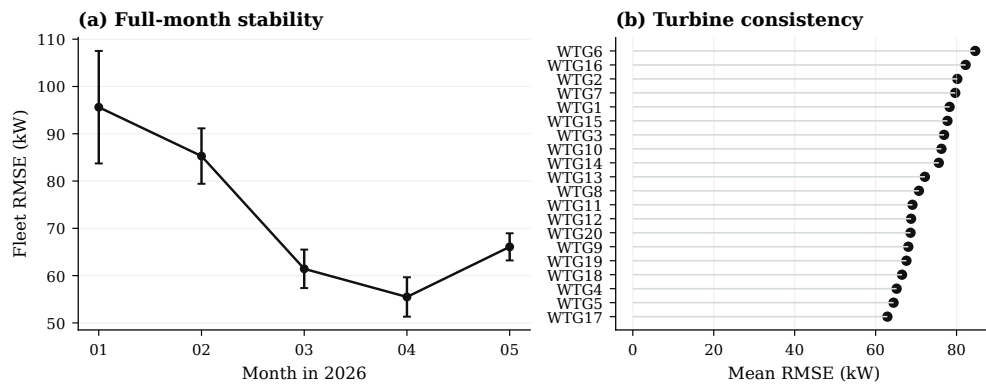


Fig. S3 Supplementary Raw-input ACT-2 consistency: monthly fleet RMSE and ranked turbine means.

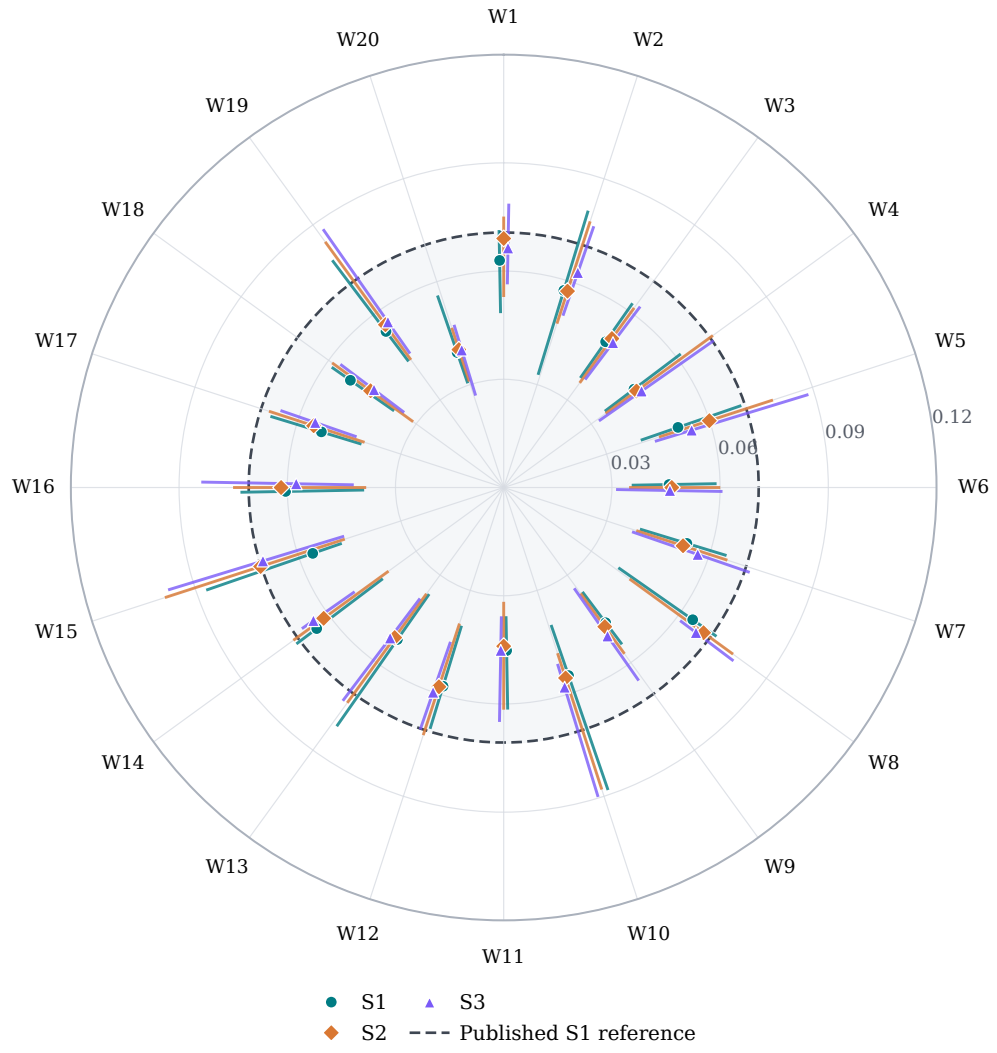


Fig. S4 Supplementary ACT-1 zero-shot transfer constellation. Marker radius is the median monthly normalized RMSE for one turbine, and radial whiskers give the interquartile range over 12 months. The dashed circle is the published DITU-Net S1 reference; smaller radius indicates lower error.

Table S2 Supplementary IDP RMSE (kW), mean $\pm$ standard deviation, six turbines.

Method	$q = 0.85$	$q = 0.90$	$q = 0.95$
4PLF	156.17 ± 7.70	146.18 ± 9.82	153.12 ± 13.17
5PLF	119.67 ± 9.95	165.80 ± 40.56	224.29 ± 41.91
BinPoly	305.23 ± 9.06	311.08 ± 12.21	326.67 ± 10.76
PCHIP	523.97 ± 46.60	402.42 ± 55.68	293.57 ± 22.50
SmSpl	524.74 ± 46.86	403.74 ± 56.90	294.35 ± 22.87
XGB-SI	414.19 ± 54.27	339.21 ± 46.35	268.00 ± 42.35
U-Net	38.62 ± 3.37	36.63 ± 3.15	35.97 ± 4.81

Bold: best value per column.

Table S3 Supplementary DKC ablation, normalized RMSE.

Evaluation	Enabled	Disabled	Δ
ACT-2/2026	0.02939	0.02939	$< 10^{-5}$
ACT-1/2025	0.04694	0.04695	10^{-5}

Table S4 Supplementary No-network ablation, normalized RMSE.

Curve source	ACT-2	ACT-1
U-Net $\rightarrow$ native reader	0.0294	0.0469
GP $Q_{0.90}$ envelope spline	0.0862	0.0329
Raw-GP spline	0.1987	0.1232

Table S5 Supplementary Readout substitution on the Yang et al. checkpoint.

Dataset	Native reader	Site reader
ACT-2/2025	0.2395	0.4471
ACT-2/2026	0.1755	0.3816
ACT-1/2025	0.2706	0.6335

Table S6 Supplementary Verified ACT-1 monthly curtailment-energy accounting for 2025.

Month	RMSE	Curt. frac.	$\widehat{EL}$ (MWh)	B (MWh)	Month	RMSE	Curt. frac.	$\widehat{EL}$ (MWh)	B (MWh)
Jan.	0.0600 $\pm$ 0.0183	0.050	988.9	1014.7	Jul.	0.0310 $\pm$ 0.0034	0.027	1464.6	1011.7
Feb.	0.0706 $\pm$ 0.0085	0.039	806.0	916.4	Aug.	0.0288 $\pm$ 0.0036	0.097	2314.0	1005.5
Mar.	0.0585 $\pm$ 0.0190	0.033	776.3	1013.4	Sep.	0.0355 $\pm$ 0.0093	0.163	2383.9	967.6
Apr.	0.0334 $\pm$ 0.0065	0.033	1318.9	958.3	Oct.	0.0546 $\pm$ 0.0090	0.252	2357.5	1014.3
May	0.0408 $\pm$ 0.0076	0.083	2103.6	1012.8	Nov.	0.0691 $\pm$ 0.0137	0.156	1305.8	981.9
Jun.	0.0319 $\pm$ 0.0059	0.104	2172.1	973.7	Dec.	0.0484 $\pm$ 0.0160	0.213	1602.3	1001.6

Total $\widehat{EL}$ = 19,593.8 MWh. RMSE is normalized; Curt. frac. is the fraction of records classified below capability; B is defined by ((see equation in main manuscript)).

Table S7 Supplementary Study site, data partitions, and experimental roles.

Partition	Records	WTGs	Role
2023–2024	1,491,030	20	XGBoost development
2025	874,415	20	XGBoost year holdout
2023–2025	5,800 images	20	U-Net development
Jan.–Jun. 2026	506,005	20	Frozen inference
ACT-1 2025	1,043,701	20	Zero-shot transfer

Table S8 Supplementary Physical and climatic comparison of the ACT-2 and ACT-1 sites.

Property	ACT-2	ACT-1
Location	Jhampir, Sindh	Jhampir, Sindh
Turbine model	Goldwind GW121/2500	1.5 MW class
Rated power per turbine	2.5 MW	1.5 MW
Number of turbines	20	20
Hub height	90 m	70 m
Cut-in / rated / cut-out	3 / 10 / 22 m/s	3 / – / 22 m/s
Mean site wind speed	5.70 m/s	5.13 m/s
Site air density	1.17 kg/m ³	N/A*
Curtailement fraction	17–19%	9.5% (range 2.7–25.2%)
Role	Train + evaluate	Zero-shot test only

*Nacelle temperature is not recorded in the ACT-1 SCADA export, so air density cannot be computed by the DKC mode rule in the main manuscript for this site.

Algorithm S1 Supplementary Site-Adapted Training and Checkpoint Selection.

1. **Input:** Rendered image pairs $\{(I_x^{(j)}, I_{wpc}^{(j)})\}_{j=1}^N$ with group labels $\{g^{(j)}\}$
 2. **Output:** Frozen deployment checkpoint θ^*
 3. **Split:** *GroupShuffleSplit* (fixed seed) holds out 15% of turbine-months as group-disjoint test partition $\mathcal{T}$; remaining 85% forms development set $\mathcal{D}$
 4. **Fold:** 5-fold *GroupKFold* on $\mathcal{D}$, grouped by turbine-month $\rightarrow$ 5 train/val partitions
 5. **For:** fold $k = 0$ to 4
 6. **For:** seed $s \in \{42, 43, 44\}$
 7. Initialise U-Net weights θ with seed s
 8. **For:** epoch = 1 to 30
 9. Compute $\mathcal{L}(\theta) = \|I_{\text{TP,ref}} - \text{Unet}_\theta(I_{\text{GP}})\|_2^2$ on training partition
 10. Update θ via Adam (lr = 10^{-3} , weight decay 10^{-5} , batch size 16, dropout 0.10)
 11. Evaluate $J_{cv}(\theta)$ on validation partition
 12. Save checkpoint if J_{cv} improves; apply plateau-based LR scheduler
 13. **End for.**
 14. **End for.**
 15. **End for.**
 16. $\theta^* \leftarrow \arg \min$ validation MSE across all $5 \times 3 = 15$ runs
 17. Evaluate θ^* once on held-out $\mathcal{T} \rightarrow$ unbiased $J_{test}(\theta^*)$
 18. Freeze θ^* ; no further weight update, fine-tuning, or hyperparameter adjustment
-

Algorithm S2 Supplementary Deployment Inference and Curve Extraction.

1. **Input:** SCADA record $(v_k, GP_k)_{k=1}^N$ for turbine i , month t ; frozen θ^* ; normalisation constants WS_{MAX} , P_{MAX}
 2. **Output:** Recovered capability curve $\hat{P}_i(v)$
 3. Normalise: $v_k^n \leftarrow v_k/WS_{MAX}$; $GP_k^n \leftarrow \max(0, GP_k)/P_{MAX}$
 4. Render (v^n, GP^n) as 256×256 RGB scatter image I_{GP} using the fixed marker-size rule
 5. Forward pass: $I_{wpc} \leftarrow \text{Unet}_{\theta^*}(I_{GP})$
 6. Convert I_{wpc} to grayscale; threshold dark pixels
 7. **Pixel mapping:** For each column c , compute intensity-weighted centroid of dark pixels $\rightarrow$ normalised scatter point (x_c, y_c)
 8. Discard columns with monotonicity jumps exceeding 0.10
 9. Fit monotone penalised B-spline to $\{(x_c, y_c)\}$ with isotonic pre-fit
 10. **DKC:** Evaluate mode by the DKC mode rule in the main manuscript; if truncated $\rightarrow$ saturated, complete curve to rated power; otherwise retain as-is
 11. De-normalise: $\hat{P}_i(v) \leftarrow f_{\text{spline}}(v/WS_{MAX}) \times P_{MAX}$
-

Table S9 Supplementary Profiled model complexity and deployment cost.

Method	Stored model	Size	Offline fitting	Inference
Manufacturer	lookup	< 1 kB	none	0.008 ms
Classical curves ^a	4–10 params/knots	< 1 kB	< 1 s	0.046–0.130 ms
XGB-SI/MI ^b	350 trees	~ 100 kB	< 5 s	~ 0.100 ms
U-Net	7.76M weights	29.6 MB	15 runs, ≤ 30 epochs	66.4 ms

^aBinPoly, 4PLF, 5PLF, PCHIP, and SmSpl. ^bA direct size and timing audit was recorded for XGB-SI; XGB-MI uses the same tree count with additional input columns.