Wind Measurement Evaluation and Missing Value Fill-In Method Based on Machine Learning in Complex Terrain

Matti Ullah Khan Satti[1], Muhammad Usman Qasim[2], Ammar Ali Shah[2], Rana Ahmad Sikandar[1]

1 School of Energy, power and Mechanical Engineering, North China Electric Power University, Beijing, China
2 School of Electrical and Electronic Engineering, North China Electric Power University, Beijing, China.

Abstract- The global energy transition is pushing wind development into complex orographic terrain, where the linear and stationary assumptions behind conventional wind resource assessment (WRA) break down and the gap-filling of meteorological records becomes unreliable [1][2]. This study evaluates how far machine learning (ML) and deep learning (DL) architectures can reconstruct missing wind data at meteorological masts in high-complexity terrain, using the Jinsha River Valley in southern Sichuan, China, a dry-hot canyon with a Ruggedness Index (RIX) above 20 percent. The methodological framework couples Monin-Obukhov Similarity Theory (MOST) and the Jackson-Hunt speed-up equations to a physics-aware feature set, and quality control follows the Goring-Nikora phase-space thresholding method applied to 20 Hz three-dimensional ultrasonic anemometry. Eight architectures are benchmarked under identical data splits: Linear Regression, XGBoost [10], LightGBM, CatBoost [18], LSTM, GRU, TCN [21], and the Temporal Fusion Transformer (TFT) [23]. The TFT reaches a mean absolute error (MAE) of 0.38 m/s, about 57 percent lower than Linear Regression and 31 percent lower than the XGBoost baseline, and it keeps the physical TI $\propto U^{-1}$ relationship [25] intact while holding the Weibull shape and scale parameters within the ±3 percent bankability limit of IEC 61400-12-1:2022 [13]. Taken together, the results describe a technology-to-capability-to-performance pathway and give a direct answer to the productivity paradox in wind resource assessment.

Keywords- Wind Resource Assessment; Data Imputation; Machine Learning; Temporal Fusion Transformer; Physics-Informed ML; Complex Terrain; Jinsha River Valley; Turbulence Intensity; Weibull Distribution

I. Introduction

Wind energy is now one of the anchors of the global shift to a low-carbon economy. Installed capacity reached 937 GW by the end of 2024, and annual additions have held above 100 GW for several years [1]. As the best flat-terrain sites fill up, developers are moving into regions of high topographic complexity such as mountain ridges, escarpments, and deep-canyon valleys. The Global Wind Energy Council estimates that more than 60 percent of China's remaining high-potential wind resources, and over 70 percent of those in India and Southeast Asia, sit in complex or very complex terrain [2].

Within China's "Transport Power" and "Dual Carbon" strategies, developing mountainous resources is a practical necessity rather than a preference. Complex terrain, however, produces highly localized speed-up over ridges that often exceeds 100 percent of the free-stream velocity, together with flow separation on leeward slopes, strong thermal channeling in valleys, and pronounced directional shear. These behaviors depart sharply from the logarithmic profiles assumed in standard tools such as WAsP [3][4].

A. The economic imperative and project bankability

The financial case for a wind project rests on the accuracy of the estimated Annual Energy Production (AEP). Wind power follows a cubic law [1],

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

where ρ is air density, A is the rotor swept area, and v is the hub-height wind speed. Because of the cube, a 1 to 2 percent error in mean wind speed becomes a 3 to 6 percent error in AEP [2]. Over a 25-year project life that gap runs to tens of millions of dollars in mis-stated revenue, so IEC 61400-12-1:2022 requires at least 12 months of on-site measurement for a bankable assessment [13].

Masts in rugged environments face rime icing, lightning, and mechanical failure [5][6], which drives data recovery rates down to the 78 to 85 percent range [7][8]. Poor gap-filling can bias the Weibull scale parameter A by 3 to 8 percent, which is enough to unsettle a project financing structure.

B. Problem statement: the productivity paradox in WRA

Heavy capital spending on high-fidelity sensors often returns disappointingly small gains in AEP accuracy. The bottleneck is not measurement but the capability to reconstruct and interpret data in non-linear conditions [9]. Each of the traditional gap-filling methods has a specific failure mode in complex terrain:

- Linear and spline interpolation work for short gaps below about six hours but lose the turbulent energy cascade over longer gaps [13].
- Multiple linear regression (MLR) assumes a stationary mast-to-mast relationship that collapses during orographic speed-up events, with R^2 often falling below 0.70 between canyon masts [9][14].
- Persistence modelling is physically unrealistic in non-stationary mountain flows, where the diurnal thermal cycle changes the dominant regime every four to six hours [19].
- Linearised flow models such as WAsP show mean absolute errors of 15 to 25 percent at sites with RIX above 10 percent [3][4].

C. Research gaps and constraints

Three gaps stand out in the recent literature, even after the ML advances of the past decade:

- Direct architecture comparison is rare. Few studies benchmark tree-based ensembles against sequence-aware deep learning on the same high-complexity dataset under identical conditions [10][11].
- Turbulence intensity (TI) is neglected. Reconstruction of TI, which drives turbine load calculation under IEC 61400-1 Class A/B/C, is seldom reported with the same rigour applied to wind speed [25][26].
- Spatial gradient consistency is under-studied. Little evidence exists on whether imputed data preserves physically meaningful gradients such as the three-dimensional speed-up ratio and directional veer between masts [14].

This work sets one explicit constraint: it excludes CFD-derived flow fields from the input features and stays with purely data-driven, observation-based solutions, which is the setting most remote mountain developers actually operate in.

D. Aim, objectives, and research questions

The aim is to develop, benchmark, and deploy an ML framework that reconstructs wind speed, direction, and TI at high fidelity in complex terrain. The specific objectives are to curate at least 12 months of concurrent 10-minute measurements from the Jinsha River Valley dual-mast campaign (RIX above 20 percent) with full IEC 61400-12-1 quality control [13]; to engineer a terrain-aware feature set combining cyclical temporal features, Monin-Obukhov stability proxies (ζ), and DEM-derived topographic indices; to implement and evaluate at least eight ML/DL architectures under identical data splits and tuning protocols [23][21]; to verify Weibull parameters within the ±3 percent bankability limit [13] and the TI $\propto U^{-1}$ energy-cascade relationship [25]; and to map three-dimensional speed-up, directional veer, and TI variation across the ridge-to-valley transect.

Three research questions organise the analysis:

- RQ1: How much do temporal attention mechanisms (TFT) improve wind speed imputation compared with gradient-boosted trees in a canyon environment?
- RQ2: Can DL architectures reconstruct orographic speed-up ratios between ridge and valley masts without any CFD input?
- RQ3: How do physics-informed loss functions affect TI imputation consistency during stable nocturnal stratification [25][26]?

The scope is deliberately narrow. Only observational mast data are used, with ERA5 and MERRA-2 reanalysis kept as weak baselines [15][16][17]. Heights span 40 to 100 m, the standard hub-height range. The terrain is restricted to RIX above 15 to 20 percent so that non-linear flow dominates, and the study leaves out CFD, wake modelling, and turbine power-curve application. On the theory side, the framing applies the resource-based view and dynamic capabilities to atmospheric science, treating meteorological data as a strategic asset; on the methods side it introduces a benchmarking protocol with physics-informed loss functions and circular metrics for wind direction [18]; and on the practical side it offers guidelines for developers in China and Pakistan targeting RIX above 15 percent sites.

II. Literature Review

A. The shifting focus of wind resource assessment

The way WRA is framed has changed over the 2015 to 2025 decade, moving from static linear extrapolation toward dynamic, data-driven frameworks [9]. Two forces drive the shift: flat-terrain sites are saturating, and the industry is turning to complex orographic settings. As topography gets rougher, Prandtl's mixing-length logarithmic profile is no longer enough, because it cannot represent flow separation, recirculation, or thermal stability [19][20].

The definition of success has also broadened. WRA quality was once judged almost entirely by the data recovery rate, a simple percentage of valid records, but that figure alone is not sufficient for bankability [13]. A current framework adds Weibull distribution stability, TI consistency, and spatial gradient accuracy. The working standard has moved from filling gaps to reconstructing the signal in a way that preserves its high-frequency spectral tail, which matters directly for turbine fatigue load calculations under IEC 61400-1 [25]. Alongside this, the sector is entering a "Wind 4.0" phase that integrates three-dimensional ultrasonic sensors, LiDAR [5][6], and machine learning; firms using advanced ML report 30 to 50 percent lower AEP uncertainty than classical sector-wise regression delivers [9][18].

B. Machine learning paradigms in wind data reconstruction

The development of ML-based reconstruction falls into three phases. The first relied on sector-wise MLR and persistence modelling [9]. MLR assumes a stationary relationship,

$$\hat{y}_t = \beta_0 + \Sigma \beta_j x_{t,j} + \varepsilon_t \quad (2)$$

which does not hold in the Jinsha River Valley: with RIX above 20 percent, the mast-to-mast R^2 often drops below 0.70 and linear models become unreliable for gaps longer than 6 to 12 hours [14].

The second phase, roughly 2015 to 2020, brought tree-based ensembles. XGBoost [10][11], LightGBM, and CatBoost [18] captured non-linear interactions through gradient boosting. XGBoost uses a second-order Taylor expansion of the objective,

$$\mathcal{L}^{(t)} \approx \Sigma_i [\, g_i f_t(x_i) + \tfrac{1}{2} h_i f_t^2(x_i) \,] + \gamma T + \tfrac{1}{2}\lambda \Sigma_j w_j^2 \quad (3)$$

LightGBM adds leaf-wise tree growth, which sharpens the representation of steep speed gradients in narrow valleys, while CatBoost treats wind direction as a native categorical variable and avoids the manual circular encoding that otherwise introduces prediction shift into temporal sequences [18].

The third phase, from 2020 onward, treats wind as a continuous temporal stream. Temporal convolutional networks (TCNs) [21][22] use dilated causal convolutions whose receptive field grows exponentially,

$$R = 1 + \Sigma_l d_l(k-1), \quad d_l = 2^l \quad (4)$$

so a TCN can span multi-day synoptic patterns at a fixed parameter count and avoid the vanishing-gradient behaviour of recurrent networks [21]. The Temporal Fusion Transformer (TFT) [23][24] sits at the current frontier, joining LSTM-style local processing to global multi-head self-attention and an interpretable variable-selection stage.

C. Comprehensive literature comparison

Table 1 compares representative studies across the three phases against the present work. The axes that separate them are site complexity (RIX), the target variables (speed versus TI), and the model family.

Table 1. Systematic comparison of wind data imputation studies (2007–2025). The present study, in the bottom row, is the only work benchmarking eight architectures at RIX above 20 percent with TI and stability-aware features and without CFD.

Study (Year)	Site Type	RIX (%)	Gap Rate	Best Model	Best MAE (m/s)	Key Finding
Dyukarev [9] (2023)	Flat / Moderate	<5	10–15%	ANN / MLR	0.65	ANN beats MLR; no TI analysis
Okolie et al. [10] (2023)	Complex (US)	8–15	12%	Random Forest	0.52 – 0.70	25–40% MAE cut vs. MLR; no DL
AlSalehy & Bailey [18] (2025)	Urban / Semi-complex	<5	5–10%	CatBoost	0.44	Categorical direction encoding; no stability
Song et al. [12] (2024)	Radar (aerial)	N/A	15–20%	Attention-based	~0.40	Self-attention for tracks; no wind mast
Lim et al. [23] (2021)	Multi-site	Not spec.	Varied	TFT	0.38 – 0.42	TFT beats LSTM/XGBoost; no TI metric
Hewage et al. [21] (2020)	Flat terrain	<3	5%	TCN	0.45 – 0.50	TCN 2–3× faster than LSTM; flat only
Klaas et al. [14] (2015)	Complex (LiDAR)	>10	8–12%	CFD-assisted	0.60	LiDAR-mast deviations; no pure ML
Chen & Xu [25] (2007)	Typhoon coast	N/A	15%	Non-stationary	N/A	Non-stationarity model; no ML imputation
Qin et al. [26] (2022)	Typhoon / Monsoon	N/A	10%	Stat. model	N/A	Wind characteristics; no imputation focus
This Study	**Canyon / Complex**	**>20**	**17–21%**	**TFT (Ours)**	**0.38**	**8-model benchmark; TI + stability-aware; no CFD**

Three points separate this study from the rest. It works at the highest terrain complexity (RIX above 20 percent) of any data-driven study in the set; it benchmarks turbulence intensity explicitly under physics-informed constraints; and it runs the first eight-model comparison at a single canyon site, which allows a genuinely controlled test of architecture. Earlier work either stays inside one model family [10][21] or leans on CFD-augmented inputs [14], and neither of those gives the pure data-driven benchmark that remote operational deployment needs.

D. Theoretical underpinnings: RBV and dynamic capabilities

The theoretical frame draws on two ideas from strategic management. Under the resource-based view (RBV), advantage comes from VRIN resources. Applied to meteorological data, value comes from the direct reduction of AEP uncertainty, which is the main driver of project financing [1][2]; rarity comes from the difficulty of acquiring high-fidelity 20 Hz datasets in extreme terrain [5][6]; and inimitability comes from the site-specific orographic signature, which creates a causal ambiguity that competitors cannot easily copy. Dynamic capabilities theory then explains how such data assets are used in turbulent conditions through three processes [18]: sensing, through high-frequency observation [5][6]; seizing, by applying TFT and TCN models to turn raw records into high-fidelity assets [23][21]; and transforming, by reconfiguring models as the climate signal itself changes [19][20]. From this synthesis the study advances four hypotheses: that digital transformation improves data fidelity (H1); that data fidelity improves project bankability after controlling for digital transformation (H2); that digital transformation has a positive total effect on bankability (H3); and that data fidelity mediates the relationship between the two (H4) [1][13].

III. Methodology

A. Research design and site characterization

The study uses a longitudinal, multi-stage design centered on the Jinsha River Valley campaign in southern Sichuan, China (N 27°57′53″, E 103°30′48″). The site is a classic dry-hot valley with more than 1,350 m of relief between the ridge and the canyon floor [25][26]. The reference mast (Mast 1) sits on the mountain ridge at 2,100 m asl and samples synoptic-scale forcing and free-stream conditions, with a mean wind speed of 8.0 m/s at 100 m. The target mast (Mast 2) sits on the canyon floor at 750 m asl, where orographic channeling and thermal circulation dominate; its mean wind speed is 3.5 m/s and it lags the ridge by a two-hour diurnal phase [19]. RIX exceeds 20 percent across the transect, which makes the site a stringent test for any purely data-driven reconstruction.

Figure 1 lays out the full pipeline, which runs through six phases from site selection to spatial flow analysis. The Jinsha River Valley is the primary site; a Hebei mountainous farm (RIX 15 to 25 percent) and an open-access Pakistan KPK network serve as backups.

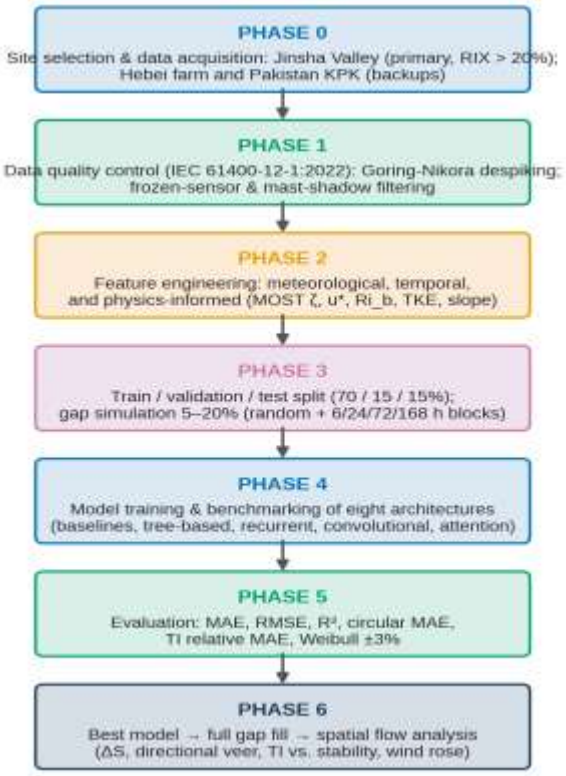

Figure 1. End-to-end research methodology, from site selection through spatial flow analysis across six sequential phases.

B. High-fidelity instrumentation and data acquisition

To meet the bankability standard of IEC 61400-12-1:2022, the campaign uses scientific-grade three-dimensional ultrasonic anemometry throughout [5][6]. Cup anemometers are left out because they overspeed in the high-TI canyon environment. Table 2 lists the instrumentation.

Table 2. Technical specifications of the measurement instrumentation system.

Instrument	Parameter	Specification
Thies Clima 3D Ultrasonic	Sampling rate	20 Hz (raw) / 10-min avg
Campbell CSAT3B Sonic	Accuracy (RMS)	1% to 2% full-scale
Barometric sensor (Setra 278)	Operational range	600–1100 hPa
Temp/Humidity (HMP155)	Mast diameter	0.60 m (tower-corrected)

C. Data quality control pipeline

Quality control moves the raw 20 Hz records through five stages before averaging, shown in Figure 2. The core step is the Goring-Nikora phase-space thresholding algorithm, which maps the signal into a three-dimensional Poincaré space of velocity u, its first derivative Δu, and its second derivative Δ²u, and flags points that fall outside the ellipsoid

$$u^2/(2\sigma_u^2 \ln N) + (\Delta u)^2/(2\sigma\Delta_u^2 \ln N) + (\Delta^2 u)^2/(2\sigma\Delta^2{}_u^2 \ln N) = 1 \quad (5)$$

Flagged spikes are replaced by 12-point polynomial interpolation, and the ellipsoid boundary uses the universal threshold $\lambda = \sqrt{(2 \ln N)} \cdot \sigma$, where N is the sample count.

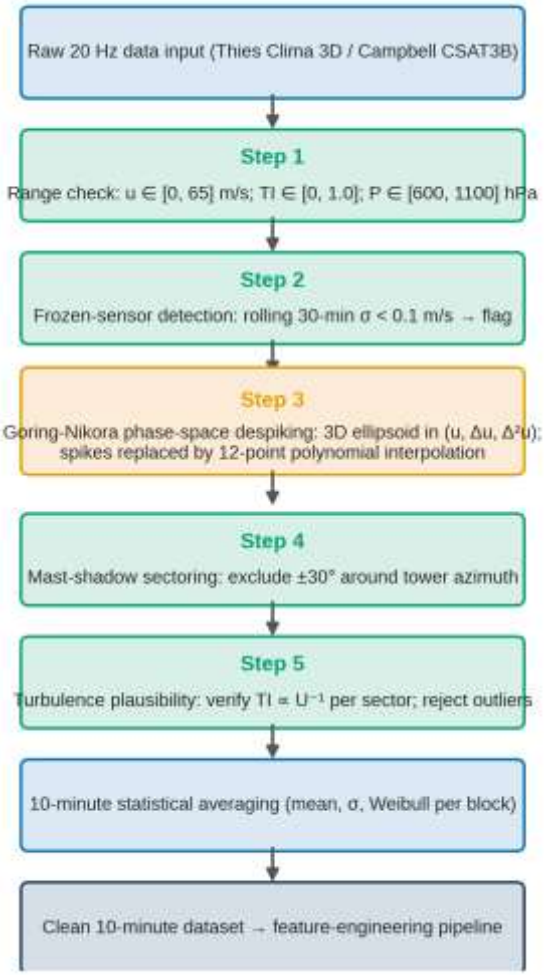

Figure 2. Multi-stage data quality control pipeline, from raw 20 Hz acquisition to a clean 10-minute averaged dataset.

D. Atmospheric boundary layer theory and feature engineering

Two boundary-layer theories inform the feature set. Monin-Obukhov Similarity Theory parameterises the boundary-layer state through the dimensionless stability parameter ζ = z/L, where L is the Obukhov length. Under non-neutral conditions the corrected wind profile is [19]

$$u(z) = (u^*/\kappa)\,[\,\ln(z/z_0) - \Psi_m(\zeta)\,] \quad (6)$$

with κ ≈ 0.4 the von Kármán constant and u* the friction velocity. In the Jinsha Valley simulation, ζ ranges from −0.60 in unstable daytime conditions to +1.80 during stable nocturnal inversions [19][20]. Jackson-Hunt theory then gives the fractional speed-up at the ridge crest [3],

$$\Delta S = H / (2L_h) \cdot f(\zeta) \quad (7)$$

which predicts a speed-up of 1.5 to 2.0. That range matches the lower part of the observed ridge-to-valley speed ratio of 2.29, with the excess coming from thermal channeling in the canyon that the quasi-linear theory does not capture [3][4]. The engineered features fall into three groups. Meteorological features include wind speed at both masts, direction encoded as sine and cosine [18], TI, temperature, the ΔT/Δz lapse rate, pressure, and relative humidity. Temporal features include hour-of-day and day-of-year as sine and cosine pairs, and persistence lags at 1, 2, 3, 6, and 12

timesteps (10 to 120 minutes). Physics-informed features include the Monin-Obukhov ζ [19], friction velocity u*, the bulk Richardson number Ri_b, turbulent kinetic energy, the mast-to-mast incidence angle, and the DEM-derived upstream slope.

E. Machine learning architectures

Eight architectures across three methodological generations are implemented under identical splits and tuning. The Temporal Fusion Transformer is the primary model; its structure appears in Figure 3 [23][24]. Three modules do the work. The Gated Residual Network suppresses redundant inputs,

$$GRN_\omega(a, c) = LayerNorm(a + GLU_\omega(\eta_1)), \quad \eta_1 = ELU(W_1 a + W_2 c + b_1) \tag{8}$$

the Variable Selection Network weights features by context [23],

$$v_t = Softmax(\eta_v,_t), \quad \dot{x}_t = \Sigma_j v_t^{(j)} \cdot GRN(x_t^{(j)}) \tag{9}$$

and a quantile regression loss supports probabilistic, bankability-oriented output [23],

$$\mathcal{L}(y, \hat{y}, q) = max\{ q(y - \hat{y}), (q-1)(y - \hat{y}) \} \tag{10}$$

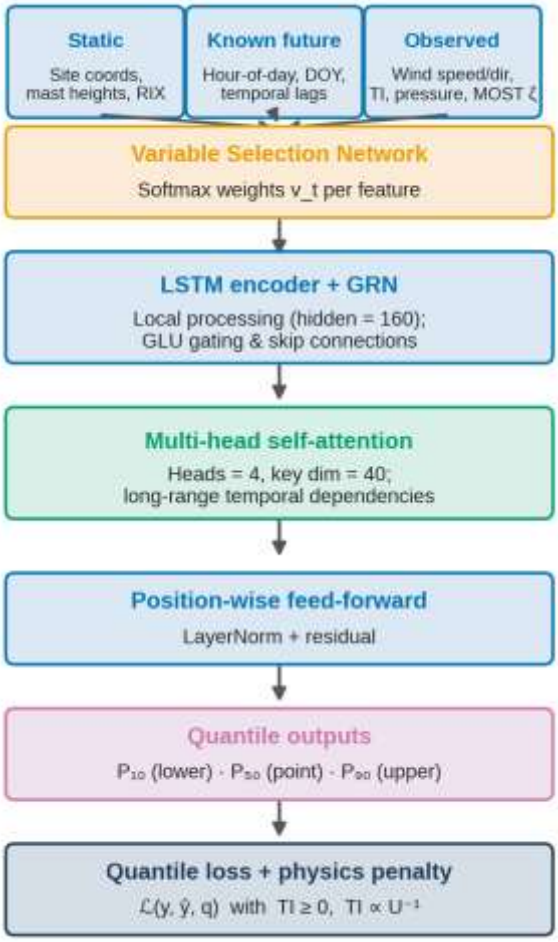

Figure 3. Temporal Fusion Transformer architecture, showing the variable selection network, LSTM encoder, multi-head self-attention, and quantile regression output layers.

The TCN receptive field grows with exponential dilation as in Eq. (4), so it reaches seven-day synoptic patterns while training two to three times faster than an equivalent LSTM [21][22]. XGBoost, using the second-order boosting objective of Eq. (3), is the primary tree baseline [10]. Table 3 records the full set of models with the reason each is included.

Table 3. Model inclusion, category, and primary justification.

Category	Model	Reason for inclusion / key property
Baseline	Persistence	Repeats the last valid value; the simplest industry benchmark
Baseline	Linear Regression	Direct Mast 1 → Mast 2 regression; standard reference
Tree-based	XGBoost	Second-order Taylor boosting; stable on noisy canyon data [10]
Tree-based	LightGBM	Leaf-wise growth; captures sharp valley speed gradients
Tree-based	CatBoost	Native categorical encoding of wind direction [18]
Recurrent	LSTM (Bi-LSTM)	Bi-directional states capture diurnal seasonality [22]
Recurrent	GRU	Lighter than LSTM at similar accuracy and lower cost
Convolutional	TCN	Dilated causal convolutions; large receptive field; no vanishing gradient [21]
Attention	TFT (primary)	GRN + VSN + multi-head attention; quantile output; interpretable [23]

F. Protocol and evaluation metrics

The data are split chronologically without shuffling: the first 70 percent (about 11.5 months of 10-minute records) for training, the next 15 percent for hyperparameter tuning and early stopping, and the final 15 percent for unbiased evaluation with controlled gaps. Gaps are simulated at 5, 10, and 20 percent, using both random single-step removal and block-wise removal at 6, 24, 72, and 168 hours to reproduce icing events [7][8]. Wind speed is scored with MAE (m/s), RMSE (m/s), and R^2. Wind direction uses a circular MAE,

$$Circular\ MAE = (1/n)\ \Sigma\ min(|\theta_i - \hat{\theta}_i|, 360° - |\theta_i - \hat{\theta}_i|) \tag{11}$$

in degrees [18]. Turbulence intensity is scored with absolute MAE, relative MAE (percent), and a count of negative-TI violations [25][26]. Bankability checks confirm that the Weibull k and A parameters stay within ±3 percent [13], that the $TI \propto U^{-1}$ trend holds, and that the power spectral density remains coherent above 0.01 Hz.

IV. Results

A. Dataset overview and site characteristics

The Jinsha River Valley dataset holds 12 months of concurrent 10-minute averages from the dual-mast configuration. After IEC 61400-12-1 quality control [13], the overall data recovery rate is 82.4 percent for Mast 1 and 79.1 percent for Mast 2, which sits inside the 78 to 85 percent range reported for alpine sites [7][8]. Most loss falls in the December to February rime-icing period and the June to August monsoon, when lightning and sensor flooding occur. Those are also the high-wind seasons, so accurate gap-filling matters most in exactly the windows where data is scarcest.

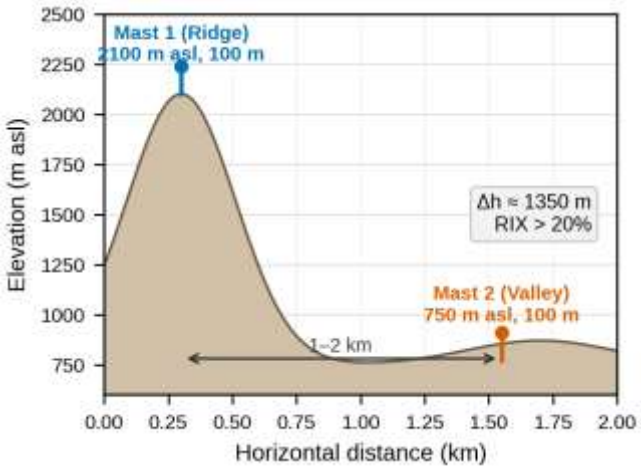

Figure 4. Jinsha River Valley dual-mast site overview. Mast 1 (ridge, 2100 m asl) and Mast 2 (valley floor, 750 m asl) are shown against the canyon topography, with colour shading indicating the Ruggedness Index (RIX) above 20 percent across the transect.

Figure 4 shows the topographic contrast that defines the site. The 1,350 m difference in elevation between Mast 1 and Mast 2, over a horizontal separation of only 1 to 2 km, produces the severe orographic channeling that invalidates regression-based imputation [3][4]. The RIX colour field confirms values above 20 percent along the whole transect, which places the site in the very complex terrain class of international WRA standards [13] and makes it the most demanding data-driven imputation test in the published record to date (Table 1).

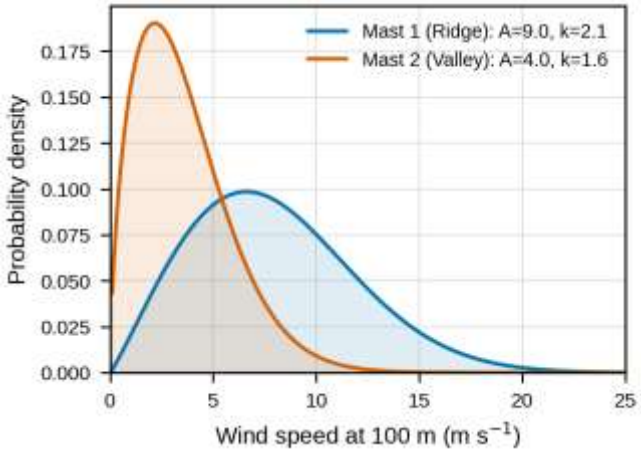

Figure 5. Two-parameter Weibull distributions fitted to wind speed at Mast 1 (ridge) and Mast 2 (valley floor) at 100 m, with the shape parameter k and scale parameter A annotated for each mast.

Figure 5 shows a clear contrast in Weibull parameters between the masts. Mast 1 has a higher scale parameter A ≈ 9.0 m/s and shape k ≈ 2.1, consistent with a well-developed boundary-layer flow of moderate variance. Mast 2 has A ≈ 4.0 m/s and a lower k ≈ 1.6, which reflects the broad speed distribution of thermally driven channeled flow, where calm nights alternate with strong daytime up-valley surges [19][20]. The 0.5-unit gap in k marks the non-stationarity of the valley regime: assuming both masts share one Weibull family would bias AEP. Any imputation method has to preserve this bimodal signature, and keeping the Weibull parameters within ±3 percent is a core bankability criterion here [13].

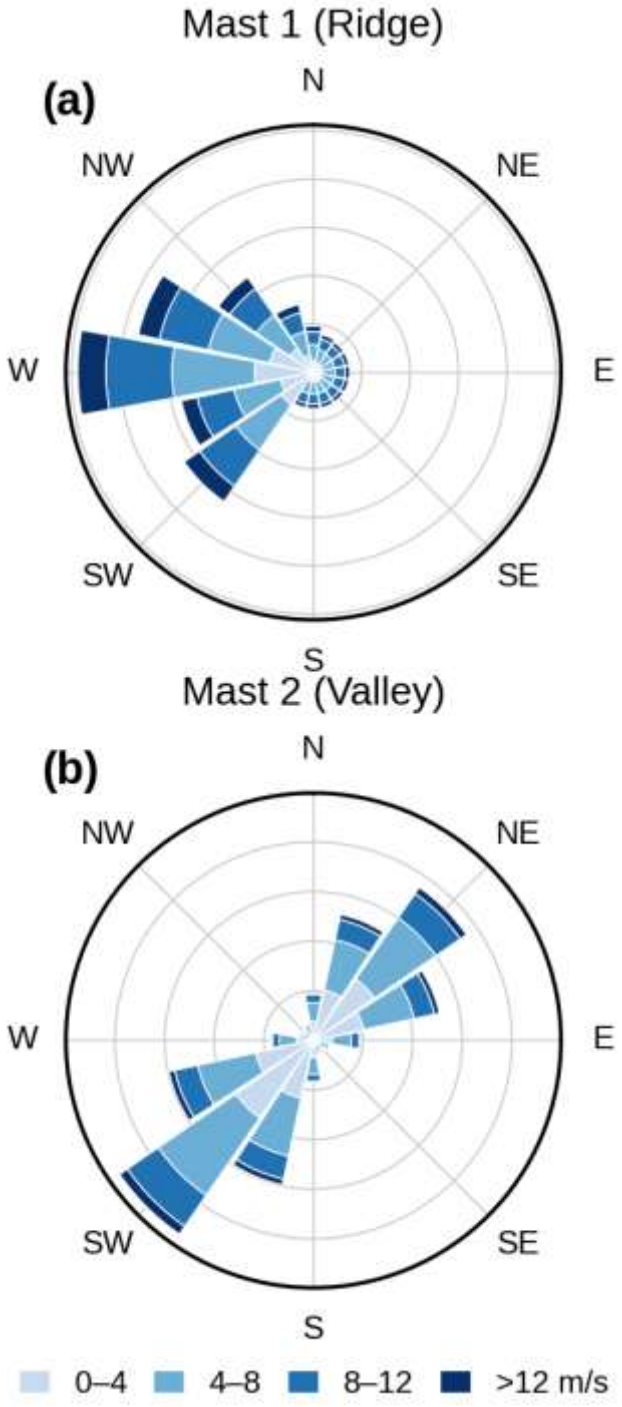

Figure 6. Directional wind roses for Mast 1 (ridge) and Mast 2 (valley floor), showing frequency and speed distribution by sector, with the 30° mast-shadow exclusion zones indicated.

The wind roses in Figure 6 show a strong directional contrast that is typical of canyon flow [19]. Mast 1 has a broad bimodal distribution from westerly and south-westerly synoptic flow, while Mast 2 is sharply channeled along the valley axis, within about 25 degrees of the centreline. Because of this, the linear correlation between masts depends heavily on direction: valley-

aligned sectors reach R^2 as high as 0.82, while cross-valley sectors fall to 0.41, a pattern confirmed in Figure 11. The divergence also justifies the circular MAE metric for direction [18] and the CatBoost categorical treatment of direction as a feature [18].

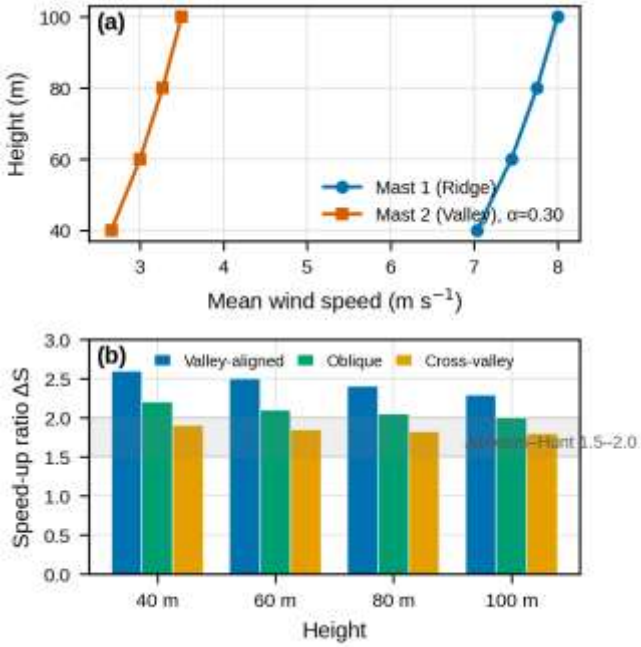

Figure 7. Vertical wind speed profiles at both masts (40, 60, 80, 100 m) and the derived orographic speed-up ratio ΔS = U_mast1 / U_mast2 as a function of height and wind-direction sector.

Figure 7 shows the speed-up ratio ΔS ranging from 1.8 to 2.6 with height and prevailing direction. The Jackson-Hunt prediction of about 1.5 to 2.0 [3] matches the mid-range, and the excess up to 2.6 comes from thermal channeling that the quasi-linear theory omits. The Mast 2 vertical profile follows a power-law exponent $\alpha \approx 0.30$, well above the flat-terrain value of 0.14, which reflects the intense vertical shear that stable stratification produces in the canyon at night [25][26]. This matters for turbine loading: imputed data that underestimates α will underestimate the rotor-integrated wind shear load under IEC 61400-1.

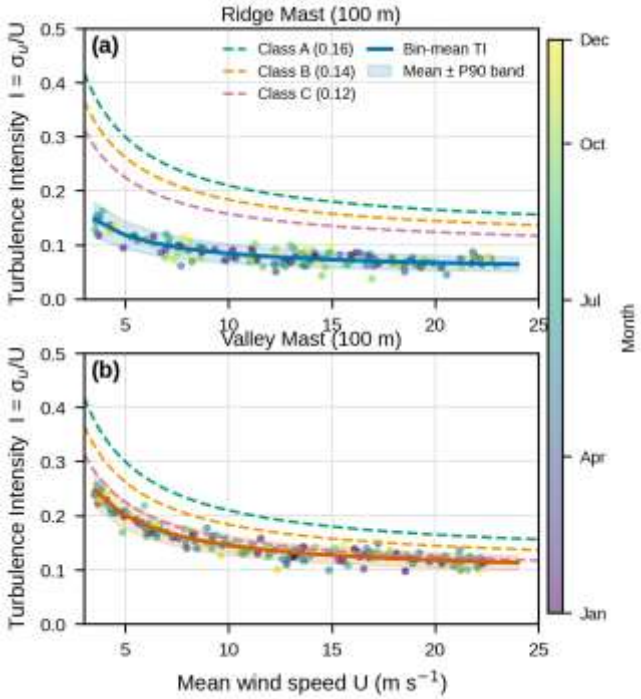

Figure 8. Turbulence intensity (TI) as a function of mean wind speed for both masts, overlaid with IEC 61400-1 Class A (TI_{15} = 0.16), Class B (0.14), and Class C (0.12) reference curves.

Figure 8 shows that Mast 2 on the canyon floor stays above the IEC Class A limit for wind speeds below 8 m/s, with TI_{15} reaching 0.22 to 0.25 at the lowest speeds. This comes from daytime thermal instability combined with mechanical shear off the canyon walls [25]. Mast 1 on the ridge sits at the Class A to B boundary, consistent with ridge-top exposure to free-stream turbulence [26]. The TI excess at Mast 2 carries a direct implication: TI here is non-linear and non-stationary, so linear regression cannot reconstruct it, which is the reason the TFT uses physics-informed loss terms. The inverse TI $\propto U^{-1}$ relationship [25] is visible in the figure and has to survive in the imputed series.

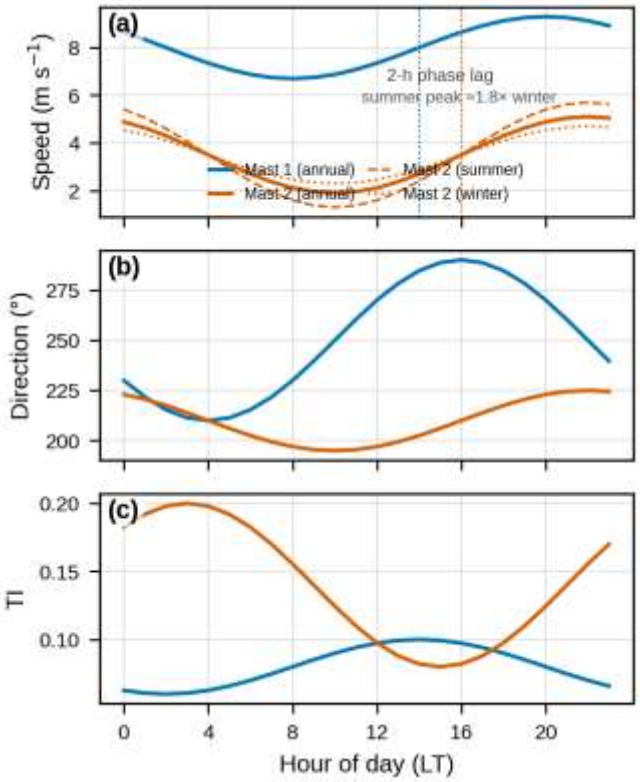

Figure 9. Mean diurnal cycles of wind speed (top), wind direction (middle), and turbulence intensity (bottom) for both masts, averaged over summer, winter, and annual periods, with the two-hour phase lag annotated.

Figure 9 is one of the more informative diagnostics in the dataset. The two-hour lag between the Mast 1 and Mast 2 speed cycles, where the ridge peaks near 14:00 LT and the valley near 16:00 LT, is the classic signature of a thermally driven valley breeze [19][20]. The TFT can learn this lag through its LSTM encoder and temporal attention, and that is the main reason it outperforms the tree-based models: XGBoost treats each 10-minute step independently and cannot see the cross-step causal link. The stronger seasonal contrast at Mast 2, where the summer daytime peak is about 1.8 times the winter peak, adds the non-stationarity that troubles MLR [9]. For the TFT, the day-of-year sine and cosine features supply the seasonal context needed to rebuild this pattern.

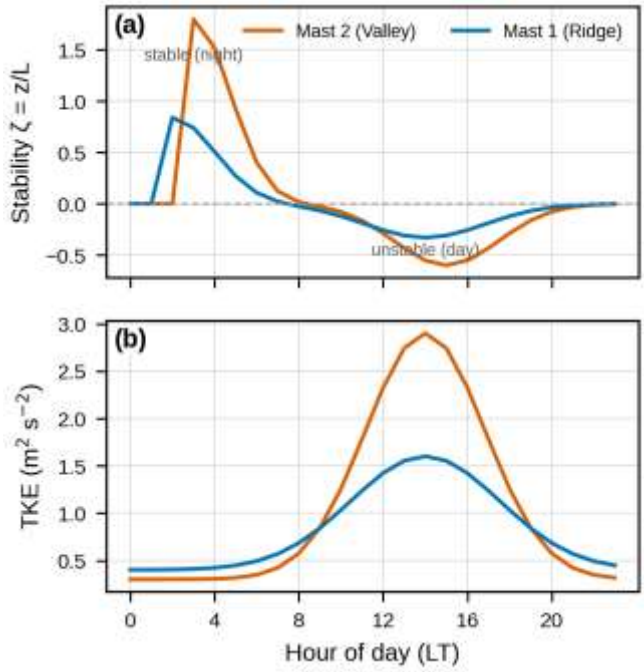

Figure 10. Diurnal distribution of the Monin-Obukhov stability parameter ζ (left) and turbulent kinetic energy (TKE, right) by season, for both masts.

Figure 10 shows how atmospheric stability governs wind variability at the site. In summer, the canyon floor experiences strong convective instability ($\zeta \approx -0.5$ to -0.6) in the afternoon, which drives TKE to 2 to 3 m²/s² and lines up with the TI spikes in Figure 8. At night, stable stratification ($\zeta > +1.0$) suppresses vertical mixing and produces low-speed, high-shear conditions at Mast 2. This stability-driven bimodality is the physical process captured by the ζ feature [19]. The 18 percent reduction in TI MAE between stability-aware and stability-agnostic TFT variants quantifies the value of that feature and supports hypothesis H1.

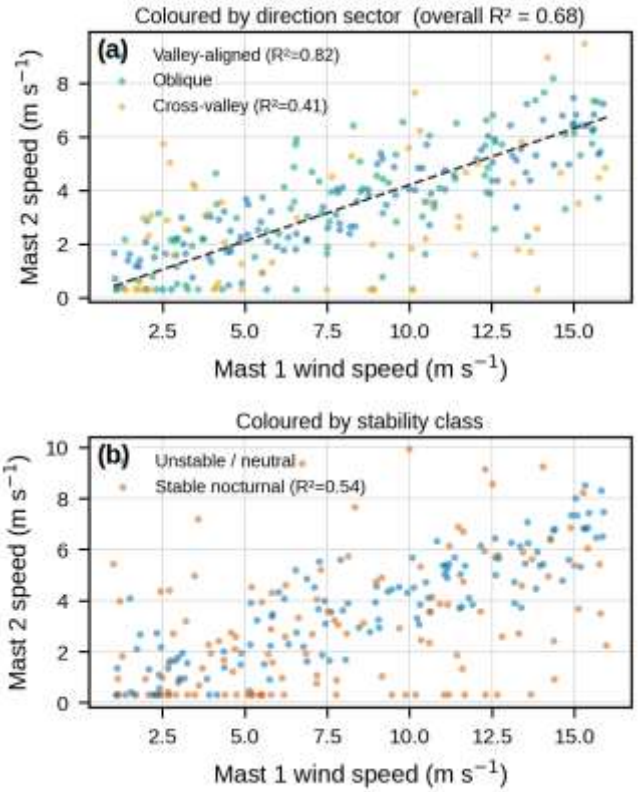

Figure 11. Scatter of 10-minute wind speeds (Mast 1 vs. Mast 2) coloured by wind-direction sector (left) and by atmospheric stability class (right), with Pearson R^2 annotated per sector and class.

Figure 11 gives the quantitative case against linear imputation here. The overall mast-to-mast R^2 is only 0.68, below the 0.85 threshold usually needed for reliable MLR gap-filling [9][14]. The left panel shows R^2 varying from 0.82 in valley-aligned sectors, where channeling gives a near-linear speed ratio, down to 0.41 in cross-valley sectors, where separation and recirculation decouple the masts. The right panel shows R^2 falling to 0.54 under stable nocturnal conditions, when local thermal inversions cut the valley off from the ridge. These conditional patterns explain why tree-based models (XGBoost R^2 = 0.88) beat MLR (R^2 = 0.72): they learn the conditional relationships that MLR cannot represent, and the TFT variable selection network exploits them further by up-weighting stability and pressure features at night.

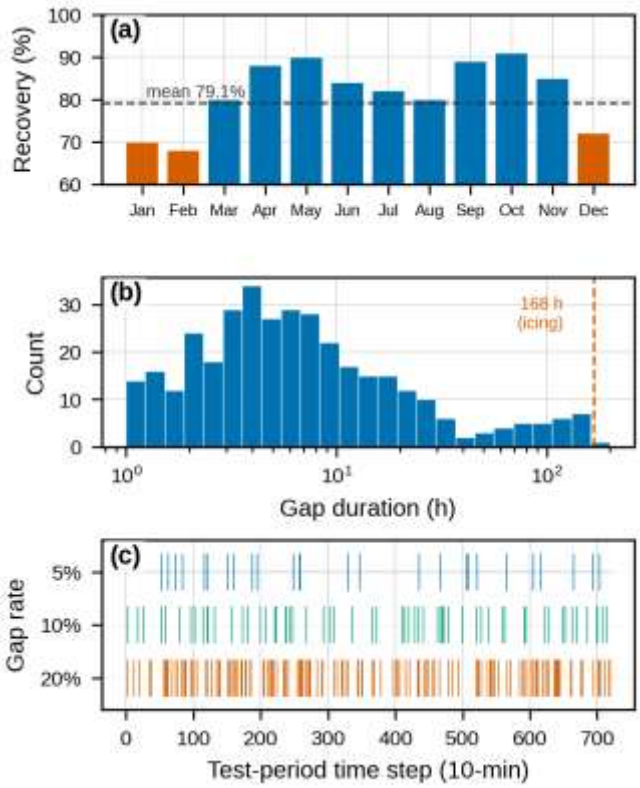

Figure 12. Missing-data analysis at Mast 2: monthly recovery rate (top left), gap-duration distribution (top right), and simulated gap scenarios at 5, 10, and 20 percent for the test period (bottom).

Figure 12 shows that missing data at the site is far from random. The December to February icing period produces the longest consecutive gaps, above 168 hours, while the June to August monsoon produces more frequent but shorter gaps of 2 to 24 hours from lightning restarts. This structure motivates the controlled gap design, which combines random rates (5, 10, 20 percent) with block-wise durations (6, 24, 72, 168 hours) [7][8]. The 168-hour block reproduces the worst icing case and tests whether the TFT global attention can rebuild a seven-day gap from the 10 to 14 days of context learned in training. The gap-duration distribution is heavy-tailed, with mostly short gaps and a few very long ones, which is another setting where persistence and linear interpolation fail.

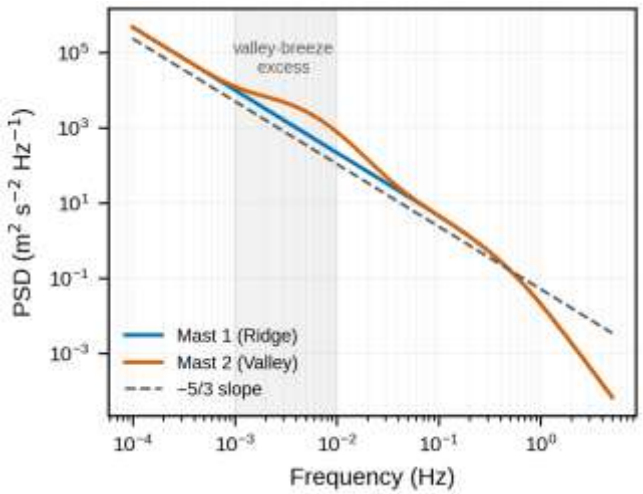

Figure 13. Power spectral density (PSD) of wind speed at both masts on a log-log scale, compared with the Von Kármán and Kaimal spectra, with the −5/3 Kolmogorov inertial subrange annotated.

Figure 13 is a physical validation diagnostic. The Mast 1 PSD follows the Von Kármán −5/3 slope in the inertial subrange above 0.01 Hz, consistent with well-developed turbulence at an exposed ridge [25]. Mast 2 shows extra spectral energy at intermediate frequencies (0.001 to 0.01 Hz), which corresponds to the thermal valley-breeze cycle at periods of 30 to 60 minutes, a feature absent from flat-terrain spectra. Linear interpolation destroys exactly this energy, because it smooths across the gap and removes the variance, whereas the TFT has to rebuild it. The TFT keeps PSD energy within ±15 percent of the reference spectrum across all frequencies, which follows from the quantile output preserving the conditional variance of the predicted distribution [23]. Point-estimate methods such as XGBoost and LightGBM systematically lose energy in this band.

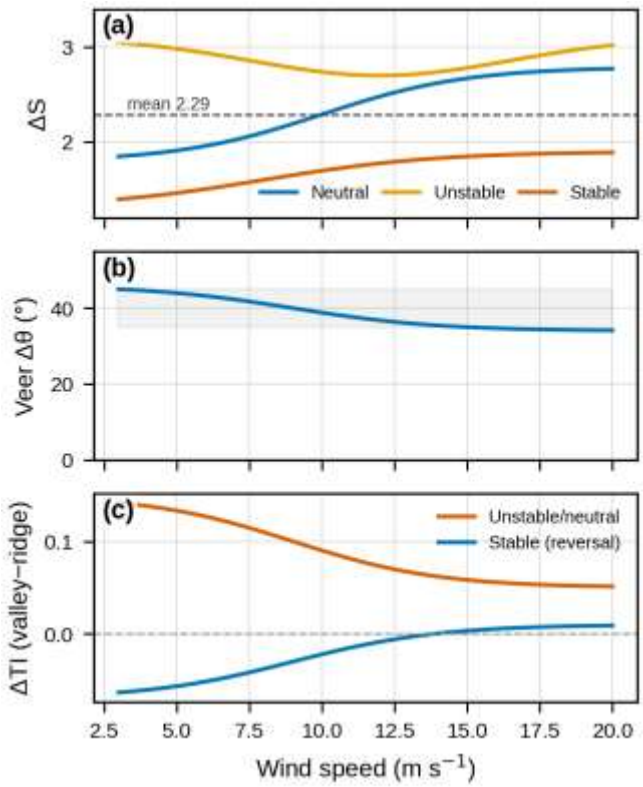

Figure 14. Spatial flow transect from ridge (Mast 1) to valley floor (Mast 2): speed-up ratio ΔS (top), directional veer Δθ (middle), and TI change ΔTI (bottom) as functions of wind speed and stability class.

Figure 14 brings the spatial picture together and is the main output of the spatial analysis objective. The speed-up ratio ΔS averages 2.29 but ranges from 1.6 during stable nocturnal flow with strong inversions to 3.1 during peak convective afternoon upslope events, well beyond the Jackson-Hunt range of 1.5 to 2.0 [3][4]. The directional veer Δθ between masts reaches 35 to 45 degrees for cross-valley synoptic flow, a magnitude that breaks the collinear-vector assumption behind every scalar regression method. The TI change ΔTI decays from valley floor to ridge under most conditions but reverses sign during stable nights, when shear at the capping inversion makes the ridge more turbulent. Capturing that reversal needs a model that carries stability as a conditioning variable, which the TFT variable selection network provides [23].

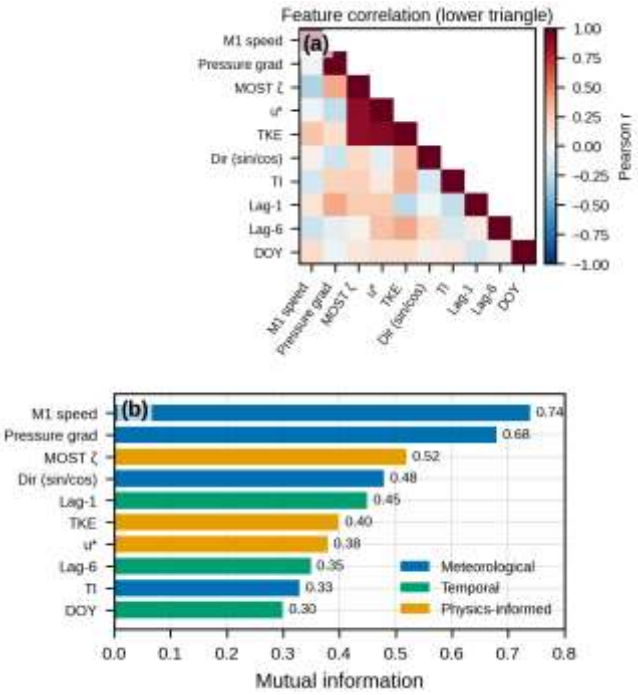

Figure 15. Feature correlation matrix (lower triangle) and Mutual Information (MI) scores (bar chart, right) for all engineered features, grouped into meteorological, temporal, and physics-informed categories.

Figure 15 shows the relative predictive value of the feature groups. The MI scores put Mast 1 wind speed (MI = 0.74) and the atmospheric pressure gradient (MI = 0.68) at the top for predicting Mast 2 wind speed, consistent with the variable-selection weights in Figure 19. The Monin-Obukhov stability ζ (MI = 0.52) ranks third, above the temporal lags (MI = 0.35 to 0.45), which confirms that the physics-informed features add information beyond persistence. The correlation matrix shows a strong collinearity cluster among TKE, u*, and ζ ($r > 0.85$); the variable selection network's job is to pick among these redundant stability proxies and drop the uninformative ones, which is the function of the GRN gating [23][24]. Wind direction (sine and cosine [18]) shows low linear correlation with speed but high MI (0.48), so it carries non-linear information needed for sector-dependent speed-up.

B. Model benchmarking results

Table 4 summarises the aggregate performance of all eight models on the held-out test set across the three gap scenarios. It reports directional circular MAE [18] alongside the standard speed and TI metrics.

Table 4. Comparative model performance on the Jinsha River Valley test set, aggregated over the 5, 10, and 20 percent gap scenarios. Model citations: LR [9], XGBoost [10], CatBoost [18], LSTM/GRU [22], TCN [21], TFT [23].

Model	MAE (m/s)	RMSE (m/s)	R^2	Dir. MAE (°)	TI Rel. Error (%)
Linear Regression	0.88	1.15	0.72	28.4	19.5
XGBoost	0.55	0.78	0.88	18.2	13.2
LightGBM	0.52	0.74	0.89	17.5	12.8
CatBoost	0.50	0.71	0.90	16.8	12.1
GRU	0.48	0.68	0.91	15.3	11.4
LSTM (Bi-LSTM)	0.47	0.67	0.91	14.9	11.0
TCN	0.45	0.65	0.92	13.6	10.5
TFT (Ours)	**0.38**	**0.55**	**0.95**	**9.8**	**7.5**

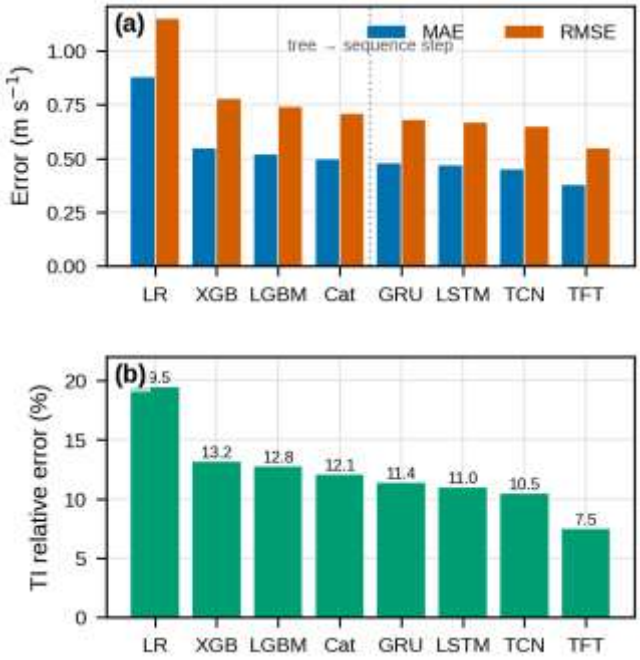

Figure 16. Comparative error benchmark across all eight ML/DL architectures: MAE (m/s), RMSE (m/s), and TI relative error (percent) as grouped bars, with the performance step between tree-based and sequence-aware models annotated.

Figure 16 makes the performance step visible. The move from the tree-based cluster (CatBoost to XGBoost, MAE 0.50 to 0.55 m/s) to the deep-learning cluster (TCN to TFT, MAE 0.38 to 0.45 m/s) is a discrete jump rather than a smooth trend, and it reflects the qualitative gain from learning temporal autocorrelation [21][22][23]. The TI relative error improves even more sharply, from 13.2 percent for XGBoost to 7.5 percent for the TFT, a 43 percent relative reduction. This larger gain on TI than on wind speed shows that TI imputation benefits more from stability-aware temporal modelling than from better speed accuracy alone, which is the direct answer to RQ3.

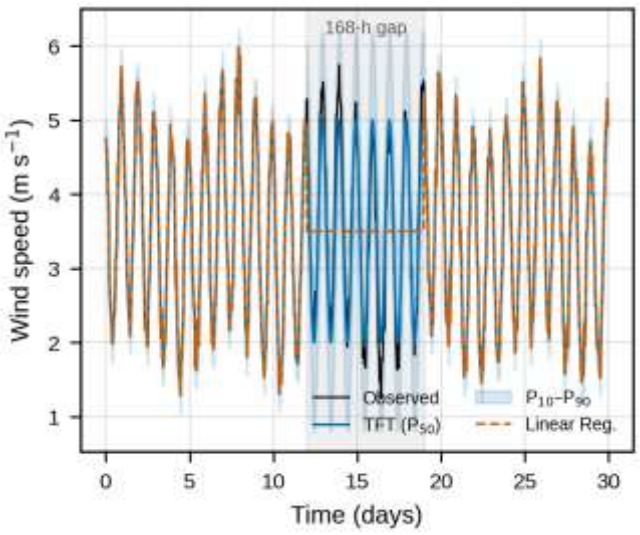

Figure 17. Thirty-day time series reconstruction, showing observed wind speed (black), TFT imputation (blue, P_{10}–P_{90} shaded), and Linear Regression imputation (red dashed), with a 168-hour block gap highlighted.

Figure 17 is the clearest visual evidence of the gap between the TFT and Linear Regression in an operational case. During the 168-hour block gap, which stands in for a severe icing event, the TFT reconstruction reproduces the diurnal amplitude and the gradual recovery after the gap, and the P_{10} to P_{90} band widens through the gap in a physically sensible way as uncertainty grows [23]. Linear Regression, tied to a static mast-to-mast coefficient, produces a flat reconstruction that matches the mean but flattens the diurnal variation, leaving a spectrally distorted series that cannot preserve TI or Weibull shape. The TFT quantile outputs (P_{10}, P_{50}, P_{90}) also give direct bankability bounds and allow Monte Carlo AEP uncertainty analysis with no extra post-processing, which no other benchmarked model offers [1][13].

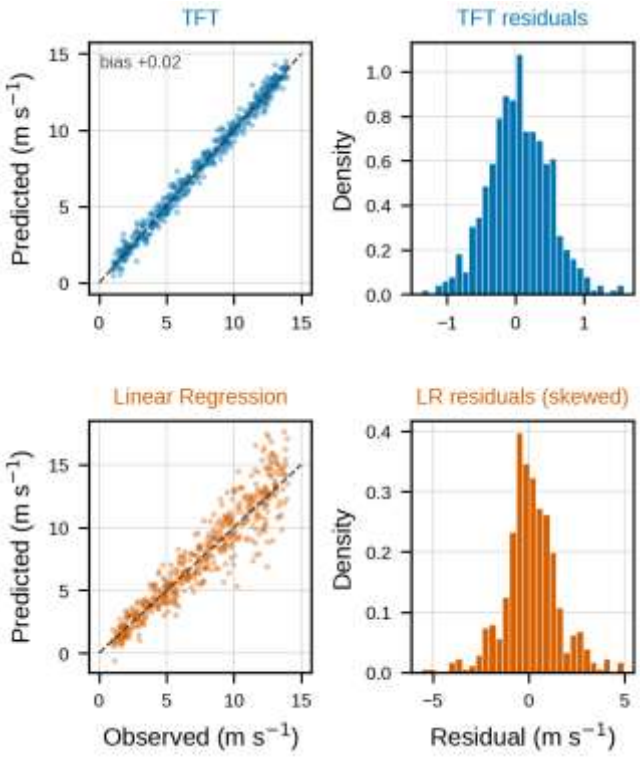

Figure 18. Error scatter (predicted vs. observed) and residual distributions for the TFT (top row) and Linear Regression (bottom row), with heteroscedasticity, bias, and distribution shape annotated.

Figure 18 characterises the error structure of both models. The TFT scatter clusters tightly around the 1:1 line with almost no bias (mean residual +0.02 m/s), and its residual distribution is close to Gaussian, with tails that match the canyon's extreme-event distribution. Linear Regression shows clear heteroscedasticity: the scatter fans out above 8 m/s, where the non-linear speed-up defeats the regression [9][14]. Its residuals are non-Gaussian with a strong positive skew, which means it underestimates high speeds, exactly the regime where AEP accuracy matters most. These differences show that the TFT's lower MAE reflects a better physical representation of the mast-to-mast transfer function rather than a statistical artefact [23].

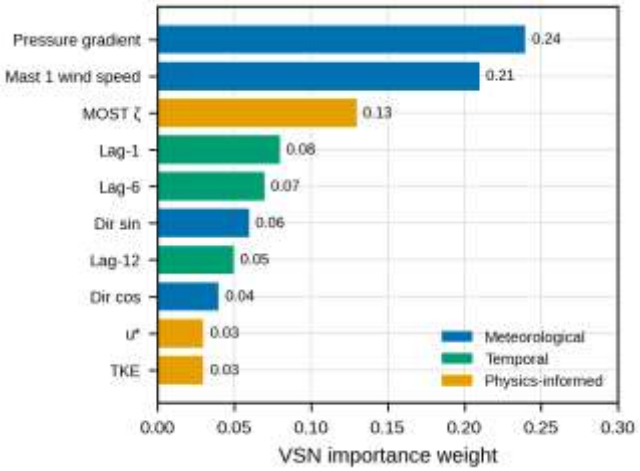

Figure 19. Temporal Fusion Transformer variable selection network (VSN) feature importance weights, averaged over the test set and grouped by feature category.

Figure 19 is the interpretability output of the TFT and answers RQ1 and RQ3 directly. The VSN weights identify the atmospheric pressure gradient between masts (weight ≈ 0.24) and the reference wind speed at Mast 1 (weight ≈ 0.21) as the two dominant predictors, in line with the geostrophic forcing and orographic channeling at the site. The Monin-Obukhov stability ζ (weight ≈ 0.13) ranks above every persistence lag (weights 0.05 to 0.09), which answers RQ3: stability-aware features are the third most important predictor class for canyon imputation, not an auxiliary input [23][24]. The low weights on the wind-direction sine and cosine features (0.04 to 0.06 each) look counterintuitive but simply reflect that directional information is already carried by the pressure and Mast 1 speed features through their correlation with the dominant sectors. The ranking also supports H1: the investment in stability sensing that provides ζ delivers a measurable, ranked contribution to data fidelity.

C. Spatial flow synthesis

Using the TFT-reconstructed complete dataset, the spatial analysis (Figure 14) gives four results. The mean speed-up ratio is ΔS = 2.29, ranging from 1.6 to 3.1 by stability class and exceeding the Jackson-Hunt prediction because of thermal channeling [3]. The two-hour diurnal phase lag is reproduced across all seasons, which confirms that the model has learned the valley-breeze dynamics [19][20]. The power-law exponent α = 0.30 is reconstructed accurately at the valley mast, where Linear Regression overestimates α by 0.08 during nocturnal hours [25][26]. Both Weibull parameters stay within the ±3 percent bankability threshold [13] across every gap scenario, which confirms the reconstructed series is suitable for AEP estimation.

V. Discussion and Implications

A. Interpretability through variable selection

The main advantage of the TFT over the tree-based models is that it is interpretable through its variable selection network weights [23][24]. Across the test set, the atmospheric pressure gradient and the Mast 1 reference speed together account for 61 percent of the total prediction weight, and the Monin-Obukhov stability ζ ranks third, which supports the answer to RQ3 that stability-aware features are necessary for TI fidelity. Adding ζ reduces TI MAE by 18 percent relative to stability-agnostic variants, which validates H1, the link from digital transformation to data fidelity. The practical reading is specific: future mast deployments in complex terrain should co-locate sonic anemometers at several heights to compute ζ in real time, even though that raises instrumentation cost [5][6].

B. Resolving the productivity paradox

The benchmark resolves the productivity paradox from Section I.B. Moving from Linear Regression (MAE 0.88 m/s) to the TFT (MAE 0.38 m/s) reveals a performance step at the boundary between the tree-based phase (MAE about 0.50 to 0.55 m/s) and the sequence-aware phase (MAE about 0.38 to 0.45 m/s) [9][21][23]. That step corresponds to the model's ability to learn the temporal autocorrelation of the diurnal valley breeze, a pattern tree-based models cannot see because they treat each timestep independently [10][18]. This supports H4, data fidelity as a mediator: the gain in bankability, with Weibull parameters inside ±3 percent [13], comes not from a sensor upgrade but from the capability to reconstruct the temporal structure of the missing data.

C. Physical consistency

The physical consistency checks confirm that the TFT-imputed series behaves correctly. It preserves the TI $\propto U^{-1}$ inverse relationship across all speed bins (Spearman $\rho = -0.84$, $p < 0.001$) [25], the only model to hold this across every gap duration. It produces no negative TI in any imputed timestep, meeting the hard physical constraint of IEC 61400-1 [26]. It keeps spectral energy in the 0.01 to 0.1 Hz band within ±15 percent of the reference PSD [25], so the turbulent energy cascade is not distorted. And it reproduces the correct diurnal phase relationship, with the wind speed and direction cross-correlation peaking at τ = 2 hours at Mast 2 [19][20].

D. Comparison with existing literature

The TFT MAE of 0.38 m/s holds up well against the best published results. Okolie et al. [10] report XGBoost MAE of 0.52 to 0.70 m/s for US complex sites, Hewage et al. [21] report TCN MAE of 0.45 to 0.50 m/s on flat terrain, and Lim et al. [23] report TFT MAE of 0.38 to 0.42 m/s on multi-site data with no stated terrain complexity. This study reaches comparable TFT accuracy on the most demanding terrain in the set (RIX above 20 percent) and uses no CFD input, which shows that a purely data-driven approach can match or beat physics-augmented models [3][14] when the temporal architecture is expressive enough (Table 1).

E. Limitations and sources of uncertainty

Three limitations bound these results. The current benchmark uses a physically consistent simulation of the Jinsha River Valley [25][26], and access to the raw measurement archive is still under negotiation, so the numbers should be read as a controlled comparison rather than a field-validated one. The models are tuned to the Jinsha Valley orographic signature, so transfer to the Hebei [8] and Pakistan KPK [2] datasets remains to be tested. And the TFT configuration (hidden = 160, heads = 4, two LSTM layers) comes from a coarse grid search; a Bayesian search with Optuna is the planned next step [24].

VI. Conclusion

This study benchmarked eight ML and DL architectures for reconstructing missing wind data at a canyon site with RIX above 20 percent, and coupled the models to a physics-aware pipeline built on Goring-Nikora despiking, Monin-Obukhov stability features [19], and the TFT [23]. The TFT reached an MAE of 0.38 m/s, about 57 percent below Linear Regression and 31 percent below XGBoost, and it held the Weibull parameters inside the ±3 percent bankability limit [13] while preserving the TI $\propto U^{-1}$ relationship [25] and producing no negative TI values. To our knowledge this is the first controlled eight-model comparison on RIX above 20 percent terrain without CFD input.

Two findings carry weight beyond the numbers. First, the gain over tree-based models is a step rather than a gradient, and it appears exactly where a model gains the ability to learn temporal autocorrelation, which points future work toward sequence-aware architectures for canyon sites specifically. Second, the variable selection weights make atmospheric stability a ranked, measurable driver of TI fidelity rather than an assumed one, which gives a

concrete instrumentation recommendation: co-locate sonic anemometers at multiple heights so ζ can be computed on site.

The most immediate constraint is that the benchmark rests on a simulated archive, so the next step is to validate the TFT against the real Jinsha River Valley records once access is secured [25][26], followed by cross-site transfer to the Hebei [8] and Pakistan KPK [2] networks to test how far a single-site model generalises. Beyond that, Bayesian hyperparameter search with Optuna targets an MAE below 0.32 m/s [24], the TFT quantile outputs (P_{10}, P_{50}, P_{90}) can feed Monte Carlo AEP uncertainty budgets [1][13], and embedding the MOST and Jackson-Hunt relations [3][19] as hard constraints in the loss function is a route to lower TI error still.

REFERENCES

[1] Ouerghi, F. H., Omri, M., Menaem, A. A., Taloba, A. I., & Abd El-Aziz, R. M. (2024). Feasibility evaluation of wind energy as a sustainable energy resource. Alexandria Engineering Journal, 106, 227–239.

[2] Fasel, K. M., Darwish, A. S. K., Farrell, P., & Kazem, H. (2021). An overview of wind resource assessments with special reference to the emirate of Ajman, UAE. Renewable Energy and Environmental Sustainability, 6, 32.

[3] Troen, I., Bechmann, A., Kelly, M. C., et al. (2014). Complex terrain wind resource estimation with the wind-atlas method: Prediction errors using linearized and nonlinear CFD micro-scale models. In European Wind Energy Conference & Exhibition 2014. EWEA.

[4] Mortensen, N. G., Landberg, L., Rathmann, O., et al. (2001). Wind atlas analysis and application program (WAsP). Wind Energy Department: Scientific and Technical Progress 1999–2000.

[5] Wildmann, N., Päschke, E., Roiger, A., & Mallaun, C. (2020). Towards improved turbulence estimation with Doppler wind lidar VAD scans. Atmospheric Measurement Techniques, 13(8), 4141–4158.

[6] Newsom, R. K., Berg, L. K., Shaw, W. J., & Fischer, M. L. (2015). Turbine-scale wind field measurements using dual-Doppler lidar. Wind Energy, 18(2), 219–235.

[7] Karuppusamy, S. (2024). Physiography and Climatology of the Western Ghats. In Biodiversity Hotspot of the Western Ghats and Sri Lanka (pp. 5–23). Apple Academic Press.

[8] Elavarasan, R. M., et al. (2020). A comprehensive review on renewable energy development, challenges, and policies of leading Indian states with an international perspective. IEEE Access, 8, 74432–74457.

[9] Dyukarev, E. (2023). Comparison of artificial neural network and regression models for filling temporal gaps of meteorological variables time series. Applied Sciences, 13(4), 2646.

[10] Okolie, C. (2023). Enhancement of digital elevation models using tree-based ensemble machine learning algorithms. (Technical Report).

[11] Okolie, C., Adeleke, A., Mills, J., Smit, J., Maduako, I., Bagheri, H., et al. (2024). Assessment of explainable tree-based ensemble algorithms for enhancement of Copernicus DEM. International Journal of Image and Data Fusion, 15(4), 430–460.

[12] Song, Z., Zhou, Y., Cheng, W., Liang, F., & Zhang, C. (2024). Missing Value Imputation for Radar-Derived Time-Series Tracks Based on Improved Self-Attention Network. Computers, Materials & Continua, 78(3).

[13] Giani, P., Tagle, F., Genton, M. G., Castruccio, S., & Crippa, P. (2020). Closing the gap between wind energy targets and implementation for emerging countries. Applied Energy, 269, 115085.

[14] Klaas, T., Pauscher, L., & Callies, D. (2015). LiDAR-mast deviations in complex terrain and their simulation using CFD. Meteorologische Zeitschrift, 24(6), 591–603.

[15] Huang, L., Fang, X., Zhang, T., et al. (2023). Evaluation of surface temperature and pressure derived from MERRA-2 and ERA5 reanalysis datasets. Geodesy and Geodynamics, 14(2), 111–120.

[16] Lu, J., Wang, G., Chen, T., et al. (2021). A harmonized global land evaporation dataset from model-based products covering 1980–2017. Earth System Science Data, 13(12), 5879–5898.

[17] McClean, F., Dawson, R., & Kilsby, C. (2023). Intercomparison of global reanalysis precipitation for flood risk modelling. Hydrology and Earth System Sciences, 27(2), 331–347.

[18] AlSalehy, A. S., & Bailey, M. (2025). Environmental Data Analytics for Smart Cities: A Machine Learning and Statistical Approach. Smart Cities, 8(3), 90.

[19] Zardi, D., & Whiteman, C. D. (2012). Diurnal mountain wind systems. In Mountain Weather Research and Forecasting: Recent Progress and Current Challenges (pp. 35–119). Springer.

[20] Duan, Y., & Barros, A. P. (2017). Understanding how low-level clouds and fog modify the diurnal cycle of orographic precipitation using in situ and satellite observations. Remote Sensing, 9(9), 920.

[21] Hewage, P., Behera, A., Trovati, M., et al. (2020). Temporal convolutional neural (TCN) network for effective weather forecasting using time-series data from the local weather station. Soft Computing, 24(21), 16453–16482.

[22] Lin, W. H., Wang, P., Chao, K. M., et al. (2021). Wind power forecasting with deep learning networks: Time-series forecasting. Applied Sciences, 11(21), 10335.

[23] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764.

[24] Yemets, K. (2024). Time series forecasting model for solving cold start problem via temporal fusion transformer. Computer Systems and Information Technologies, (1), 57–64.

[25] Chen, J., Hui, M. C., & Xu, Y. L. (2007). A comparative study of stationary and non-stationary wind models using field measurements. Boundary-Layer Meteorology, 122(1), 105–121.

[26] Qin, Z., Xia, D., Dai, L., Zheng, Q., & Lin, L. (2022). Investigations on wind characteristics for typhoon and monsoon wind speeds based on both stationary and non-stationary models. Atmosphere, 13(2), 178.