

Six-Hour-Ahead Wind Power Forecasting Across Iraqi Cities Using a NASA-GFS Aligned Multi-Model Framework

Abdulqader Jasim Mohamed^{1*} , Khalid Mohammed Saffer² 

^{1,2} Department of Computer Science, College of Science, University of Diyala, 32001, Iraq

¹ scicomps242506@uodiyala.edu.iq, ² dr.khaledmoh@uodiyala.edu.iq

Abstract

Short-term wind-power forecasting is essential for renewable-energy planning because wind generation changes rapidly with atmospheric conditions and turbine power output responds nonlinearly to wind speed. This study evaluates six-hour-ahead gross wind-power forecasting across 15 Iraqi cities. We derived the target power series from NASA POWER wind data by estimating hub-height wind speed and converting it to power using a 3370-kW reference turbine. We divided the operational forecasting dataset chronologically into 2021-2023 for training, 2024 for validation, and 2025 for testing. The evaluated methods include raw GFS power-law baselines, MOS-calibrated variants, standalone neural models, history-only forecasting, and multi-model forecast blending. Blend v2 + History LSTM achieved the best overall performance. Across the full H6 window, the model achieved an RMSE of 530.75 kW, an MAE of 320.41 kW, a bias of -8.38 kW, and an R2 of 0.8395 over 122,538 forecast rows. This corresponds to a 15.01% RMSE reduction relative to the history-only LSTM baseline and a 53.71% reduction relative to the raw GFS power-law baseline. Lead-step results show the expected increase in Error from h+1 to h+6. Speed-regime diagnostics identify the 9-12 m/s range as the most difficult operating region because it lies near the rated-power transition, where small wind-speed errors can produce large power deviations. Overall, the results indicate that forecast blending provides a stable and interpretable short-term wind-power forecasting framework for Iraqi cities.

Keywords: wind power forecasting; six-hour-ahead forecasting; NASA POWER; GFS; forecast blending; LSTM; MOS; Iraq; renewable energy planning.

Article history: Received: 22 Jun 2026; Accepted: 10 Sep 2026; Published: 15 Sep 2026.

This article is open-access under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate wind-power forecasting is essential for renewable-energy planning. Because wind production varies with atmospheric conditions and turbine operating behavior, six-hour-ahead forecasts can support power system scheduling, storage coordination, reserve planning, and wind energy integration. The problem is particularly relevant in Iraq, where electricity demand can exceed reliable supply and renewable energy resources remain underused relative to their potential [5]. Previous studies on hybrid renewable systems in Iraq show that wind and solar resources can help meet local electricity demand and reduce dependence on conventional generation. However, performance

depends strongly on location and resource variability [6].

Wind-power forecasting is more difficult than wind-speed forecasting because turbine power curves are nonlinear. Around the rated-power transition, small wind-speed errors can lead to large power errors. In contrast, errors at very low or very high speeds may have a smaller effect on the final power estimate. This paper focuses on six-hour-ahead forecasting of a NASA-derived gross reference-turbine power proxy for 15 Iraqi cities. The experiments use a consistent H6 framework and compare raw forecast baselines, MOS calibration, recurrent models, history-based forecasting, and multi-model blending.

The main contributions are: (1) an issue-time-aligned NASA-GFS framework for h+1 to h+6 forecasting across 15 Iraqi cities; (2) a common 3370 kW reference-turbine target that allows the same power definition to be used at every location; and (3) a leakage-controlled comparison of raw GFS, MOS calibration, recurrent forecasting, and Ridge blending under a chronological 2021-2025 operational protocol.

2. Related Work

Mahmood et al. [1] provide a regional deep-learning reference for weather forecasting in the Middle East. Their study compares RNNs and hybrid LSTM-GRU models for climate variables, including wind speed, solar radiation, humidity, temperature, and rainfall. It is relevant to the present work because it demonstrates the use of recurrent and hybrid deep-learning models for Middle Eastern weather and climate time series. Most wind-energy studies in Iraq have focused on resource assessment, turbine suitability, hybrid-system feasibility, and general weather-variable prediction. These studies provide useful local context, but they do not fully address short-term multi-city wind-power forecasting using a common turbine-based power target.

For Duhok city, Hussain et al. used a recurrent neural network to predict wind speed, showing the relevance of recurrent learning for wind-speed prediction in an Iraqi context [14]. The present study extends this line of research by forecasting wind gross power rather than wind speed alone and by evaluating a shared model across 15 cities. Deep learning has also become an important tool for solar and wind forecasting because it can learn nonlinear temporal patterns that purely statistical models struggle to capture [15]. A review of renewable-energy forecasting further underscores the need to shift from single-site to multi-site forecasting frameworks [16].

Recent short-term forecasting research also shows a shift toward integrated and hybrid architectures. Wang et al. proposed a variable-attention and multi-task learning method for short-term forecasting of wind power, solar power, and system load [17]. Similarly, Bashir et al. developed a multi-component wind and solar forecasting system

based on CNN-ABiLSTM and CNN-transformer-MLP models [18].

Feasibility and resource-assessment studies in Iraq provide the energy-planning background for this work. Alshamri et al. optimized on- and off-grid hybrid renewable-energy systems for house units in Iraq and showed that local wind speed affects the feasibility and payback period [6]. Al-Mamory et al. assessed 60 MW hybrid wind-solar systems in Iraq. They reported that wind-solar projects can support energy export and emissions reduction when resource conditions are suitable [7]. At the resource-assessment level, Abbas et al. analyzed measured wind data at 10, 30, and 50 m heights in southern Iraq and found that the dominant wind directions were from the west and northwest [9]. Hadi et al. developed GIS-based classification maps for turbine characteristics in central and southern Iraq, further emphasizing the need to match turbine characteristics to local wind regimes [8].

Several Iraqi studies address related but narrower forecasting tasks. Mohammed et al. [10] considered general weather forecasting over Iraq, while Aziz et al. [11] focused on wind-speed estimation in a case study. Saleh [12] estimated wind power from wind-speed records and turbine characteristics, and Aldarraj et al. [13] forecast Iraqi electricity supply and demand. These studies support data-driven forecasting in the Iraqi context. However, they do not provide the same issue-time-aligned h+1 to h+6, multi-city gross reference-turbine power experiment used here.

The present paper therefore addresses a more specific forecasting problem than general weather prediction, resource assessment, or system-level demand forecasting. It evaluates a common NASA-derived gross reference-turbine power proxy over an H6 multi-city test set and reports model ranking, lead-step behavior, city-level diagnostics, speed-regime errors, and forecast-row energy accounting within one chronological framework.

Study Area and Data Sources

The study area includes 15 Iraqi cities: Anbar, Babylon, Baghdad, Basra, Dhi Qar, Diyala, Karbala, Kirkuk, Maysan, Muthanna, Najaf, Nineveh, Qadisiyyah, Salah al-Din, and Wasit. We treat the

task as a global multi-city forecasting problem rather than a collection of independent city-specific models.

The two data sources serve different purposes. We used NASA POWER hourly data to construct the historical reference target, not to provide future forecast information. We used NASA wind fields at 10 m and 50 m to estimate wind speed at the 110 m turbine hub height. We then used the reference turbine power curve to obtain the NASA-derived gross wind-power proxy [2,4]. We used GFS data only as issue-time-aligned numerical weather prediction (NWP) inputs, providing future forecast information for lead times $h+1$ to $h+6$ [3].

Because the two sources have different roles and time coverage, we defined the chronological split at two levels. We split the full NASA-derived target period, 2015-2025, into 2015-2021 for training, 2022-2023 for validation, and 2024-2025 for testing, using NASA-only historical or diagnostic experiments. The final operational H6 experiments were restricted to the common NASA-GFS-aligned period, 2021-2025. In this setting, we used 2021-2023 for training, 2024 for validation and blend calibration, and 2025 for independent testing. Table 1 outlines the data-source roles, temporal coverage, and chronological evaluation design.

Table 1: Evaluation design, chronological and data-source roles.

Item	Value
Cities	15 Iraqi cities
Forecast horizon	$h+1$ to $h+6$
NASA-derived target coverage	2015-2025
NASA-only historical split	Train 2015-2021; validation 2022-2023; test 2024-2025
GFS-aligned forecasting period	2021-2025
Adopted H6 operational split	Train 2021-2023; validation 2024; test 2025
Official target	NASA-derived 110 m wind speed converted to gross reference-turbine power (proxy)
Reference capacity	3370 kW (3.37 MW)

3. NASA-Derived Gross Wind-Power Target Construction

The official output variable, `wind_gross_power_kw`, is a NASA-derived gross reference-turbine power proxy, not measured turbine or wind-farm output. For each city and hour, we used NASA POWER wind

speeds at 10 and 50 m to estimate hub-height wind speed. The time-dependent shear exponent and the 110 m speed were calculated as follows:

$$\alpha_t = \ln(v_{50,t} / v_{10,t}) / \ln(50 / 10) \quad (1)$$

$$v_{110,t} = v_{50,t} (110 / 50)^{\alpha_t} \quad (2)$$

The shear exponent was evaluated only when both source wind speeds were at least 1.0 m/s. Invalid values were repaired within the location using the location median and, when necessary, a fallback of 0.143, then constrained to 0.05-0.50 before extrapolation. The estimated 110 m speed was finally bounded to 0-80 m/s as a numerical safeguard. The resulting speed was mapped to the tabulated IEA reference-turbine electrical curve by linear interpolation; values below 3.0 m/s or above 25.0 m/s were assigned zero power, with output limited to 0-3370 kW. The 3.0 m/s value is the lower boundary of the adopted electrical table. In contrast, the turbine specification reports an operational cut-in speed of 4.0 m/s [4].

The 3.37 MW IEA Task 37 reference turbine was selected because its published specification provides the 110 m hub height, rotor dimensions, operating limits, and tabulated electrical power curve required for a transparent and repeatable conversion [4]. Applying one turbine to all cities standardizes the target; it does not imply that this turbine is installed or optimal at every location. Different turbine designs and local wind regimes would change the speed-to-power mapping and therefore the absolute error magnitudes. Plant-level wake losses, availability, curtailment, maintenance, and electrical losses are not represented. Fig. 1 illustrates the three-stage workflow from NASA-derived target construction through GFS-aligned H6 forecasting to final blending and evaluation. Table 2 lists the reference-turbine specifications and target-construction parameters used in the study.

Table 2: Reference turbine and NASA-derived target-construction parameters.

Parameter	Value
Reference turbine	IEA 3.4 MW / 3.37 MW reference turbine
Rated power used	3370 kW
Hub height	110 m
Rotor diameter	130 m
Rated wind speed	about 9.8 m/s

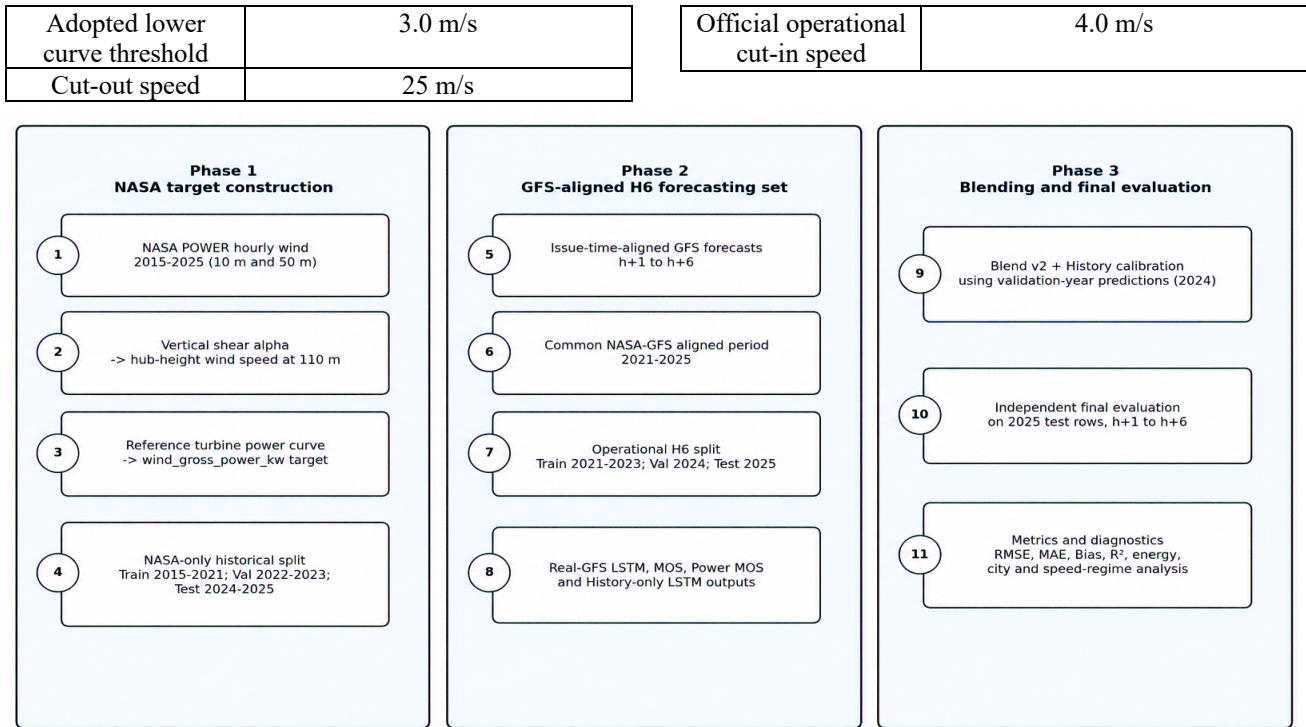


Fig. 1 Workflow and chronological splitting of the NASA-derived target, issue-time-aligned H6 forecasting, and final blending/evaluation stages.

4. Forecasting Framework, Blending, and Leakage Control

Blend v1 used a Ridge meta-model to combine the Real-GFS LSTM, Multi-grid Speed MOS, Multi-grid Power MOS, selected raw multi-grid power estimates, forecast lead, cyclical hour/month variables, and city one-hot indicators. Blend v2 retained those inputs and added the History-only LSTM forecast. The blend was therefore learned rather than assigned fixed manual weights. For case i , the final prediction is $P_i = \beta_0 + z_i^T \beta$, where z_i contains the component forecasts and meta-features. Ridge coefficients were estimated from aligned 2024 validation predictions by minimizing $\Sigma(P_i - \beta_0 - z_i^T \beta)^2 + \alpha \|\beta\|^2$, with $\alpha = 10.0$. We standardized and fit Ridge using 2024 only, then froze the fitted transformation and coefficients before the 2025 test. The retained Blend v2 coefficients are publicly available in the reproducibility repository and Zenodo archive identified in the Data Availability statement.

We enforced leakage control through GFS issue-time, lead-time, and valid-time alignment. For every forecast origin, the latest archived GFS cycle issued at or before that origin was used; a later cycle was never substituted. For example, at a 10:00 UTC forecast origin, the 06:00 UTC cycle is eligible, but

the 12:00 UTC cycle is not. The $h+1$ target at 11:00 UTC is therefore matched to the $+5$ h field from the 06:00 cycle, while the $h+6$ target at 16:00 UTC uses the $+10$ h field from that same eligible cycle. We retained future NASA-derived wind and power values only as reference targets.

NASA-only diagnostic and architectural experiments use a different information and chronological protocol. They are therefore excluded from the official NASA-GFS operational ranking and final Ridge blend. This separation prevents models with different future-information availability or test periods from being treated as directly comparable operational baselines.

5. Performance Metrics

We evaluated performance using RMSE, MAE, bias, R^2 , percentage-capacity error (PCE), and cumulative energy error (CEE). In the following equations, A_t denotes the NASA-derived target wind power, F_t denotes the forecasted wind power, A_{bar} denotes the mean of the target values, and n is the number of evaluated forecast rows. Division by n indicates that the Error is averaged across all evaluated samples.

$$\text{RMSE} = \sqrt{(1/n) \Sigma(A_t - F_t)^2} \quad (3)$$

$$\text{MAE} = (1/n) \Sigma|A_t - F_t| \quad (4)$$

$$\text{Bias} = (1/n) \sum (F_t - A_t) \quad (5)$$

$$R^2 = 1 - \sum (A_t - F_t)^2 / \sum (A_t - \bar{A})^2 \quad (6)$$

Equations used for wind-power forecasting evaluation. For consistency, RMSE, MAE, and R^2 use the same error difference, $A_t - F_t$. Bias is written in the opposite direction, $F_t - A_t$, so that a positive bias indicates over-prediction and a negative bias indicates under-prediction. Percentage-capacity errors express each power error relative to the turbine's 3370 kW rated capacity.

6. Model Implementation

The official NASA-GFS comparison retains baselines with distinct roles but a common operational period: raw GFS measures uncalibrated NWP-to-power performance; Speed MOS and Power MOS measure statistical calibration; History-only LSTM tests forecasting without future NWP; Real-GFS LSTM combines recent History with issue-time-aligned future GFS; and Blend v2 tests whether these sources are complementary. These operational components use 2021-2023 for development, 2024 for validation/calibration, and 2025 for independent testing. The NASA-only strict diagnostic, TCN, and TCCNN experiments remain separate because their predictor availability and chronology differ.

The Real-GFS LSTM uses a 168 h historical sequence, three LSTM layers with 256 hidden units per layer and 0.20 dropout, plus an MLP that maps the future GFS inputs through 128 units to a 64-dimensional representation. The concatenated representation passes through a 128-unit prediction head to six direct wind-speed outputs. The History-only LSTM uses the same 168 h lookback, three 256-unit LSTM layers, 0.20 dropout, and a 128-unit head, but no future GFS predictors. Standard LSTM sigmoid/tanh gates are used, while the feed-forward hidden layers use ReLU activations. Continuous variables are standardized using training-fitted

scalers. The official recurrent runs use AdamW with learning rate 0.001, weight decay 10^{-4} , batch size 2048, gradient clipping at 1.0, seed 1, at most 30 epochs, and early stopping after five validation checks without improvement. The Real-GFS LSTM minimized a power-curve- and horizon-weighted Huber loss ($\beta = 1.0$), whereas the History-only LSTM used a power-curve- and horizon-weighted mean squared error loss; both weighting schemes emphasize steeper power-curve regions and more distant lead steps.

Speed MOS and Power MOS are fitted separately for $h+1$ to $h+6$ with Hist Gradient Boosting Regressor models using a local 3×3 GFS neighborhood (learning rate 0.05, maximum 500 iterations, maximum 31 leaf nodes, minimum 30 samples per leaf, L2 regularization 0.01, 10% internal validation, and early stopping after 30 non-improving iterations). Model settings and checkpoint choices were fixed during development/validation before the 2025 test; no test-period tuning was performed.

7. Results

A. Overall H6 model comparison

Blend v2 + History LSTM produced the best official NASA-GFS H6 result among the operational models, with an RMSE of 530.75 kW, an MAE of 320.41 kW, a bias of -8.38 kW, and an R^2 of 0.8395 across 122,538 forecast cases. Its RMSE was 15.01% lower than the History-only LSTM and 53.71% lower than the raw GFS power-law baseline. These percentage differences are reported descriptively; no formal paired significance claim is made because a common case-level prediction archive was not retained for every standalone comparator. Fig. 2 visualizes the RMSE ranking of the six official NASA-GFS-aligned operational models reported. Table 3 presents the official operational model comparison and RMSE-based ranking on the independent 2025 test set.

Table 3: Official NASA-GFS-aligned H6 operational model comparison on the independent 2025 test set.
 Rank is based on RMSE (lower is better).

Rank	Model	Type	RMSE kW	MAE kW	Bias kW	R ²
1	Blend v2 + History LSTM	Final ensemble	530.75	320.41	-8.38	0.8395
2	Real-GFS LSTM v1	Neural model	544.21	323.07	n/a	0.8300
3	Multi-grid Speed MOS	MOS calibration	584.09	327.39	-31.17	0.8100
4	Power MOS	Direct power calibration	585.73	380.23	88.02	0.8000
5	History-only LSTM H6	History model	624.47	378.24	4.91	0.7800
6	Raw GFS power-law baseline	Raw NWP baseline	1146.60	769.30	-438.44	0.2600

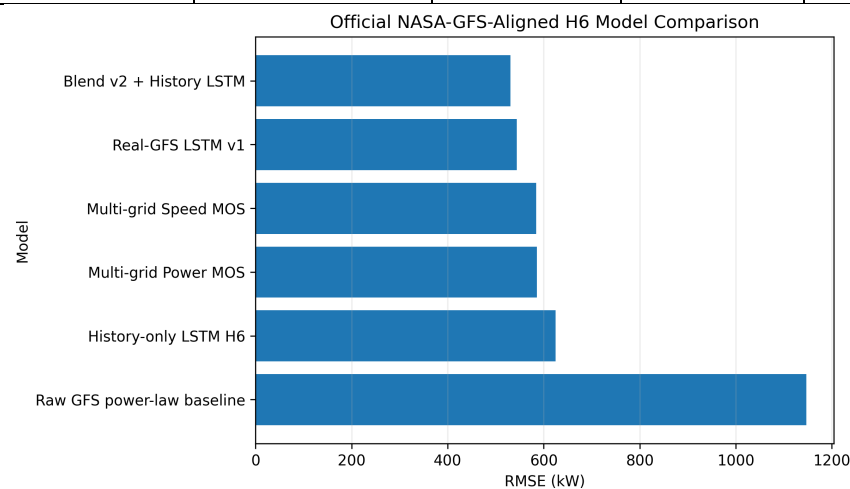


Fig. 2 RMSE ranking of the official NASA-GFS-aligned H6 operational models.

B. Lead-step performance within the H6 window

Table 4 separates the full H6-window result from the individual lead-step results. The overall row aggregates all forecast rows from h+1 to h+6, whereas the h+6 row represents only the sixth lead step and should not be interpreted as the overall H6 performance. The lead-step results clearly show that Error increases as forecast lead time grows. RMSE increases from 216.84 kW at h+1 to 669.60 kW at

h+6, while R² decreases from 0.9737 to 0.7505. This behavior is expected because the task moves from near-term correction toward increasingly uncertain short-term wind evolution. Fig. 3 shows how RMSE and MAE change from h+1 to h+6 for the adopted Blend v2 + History LSTM model. Fig. 4 shows the corresponding decline in R² as the forecast horizon increases from h+1 to h+6.

Table 4: Lead-step performance of the adopted Blend v2 + History LSTM model within the H6 forecasting window.

Horizon	RMSE kW	RMSE (% of 3370 kW)	MAE kW	Bias kW	R ²	Count
Overall h+1-h+6	530.75	15.75	320.41	-8.38	0.8395	122538
h+1	216.84	6.43	106.97	-2.89	0.9737	20505
h+2	379.71	11.27	212.05	-4.86	0.9175	20487
h+3	515.24	15.29	316.41	-8.68	0.8474	20470
h+4	608.40	18.05	404.80	-15.24	0.7831	20364
h+5	646.28	19.18	433.88	-13.46	0.7592	20362
h+6	669.60	19.87	450.58	-5.20	0.7505	20350

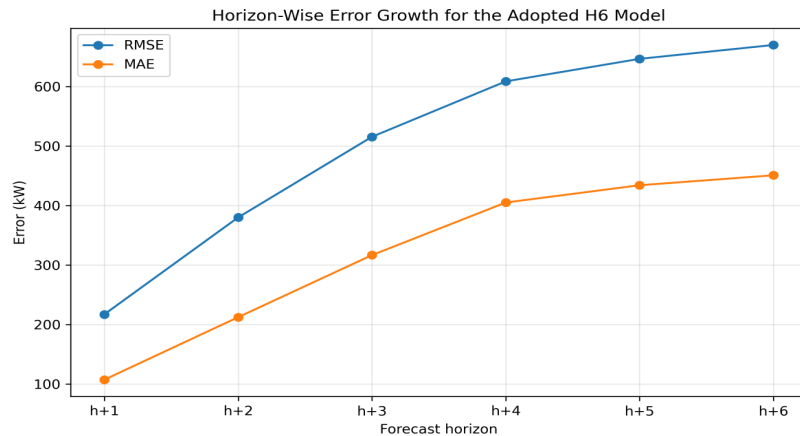


Fig. 3 Horizon-wise RMSE and MAE of Blend v2 + History LSTM from h+1 to h+6.

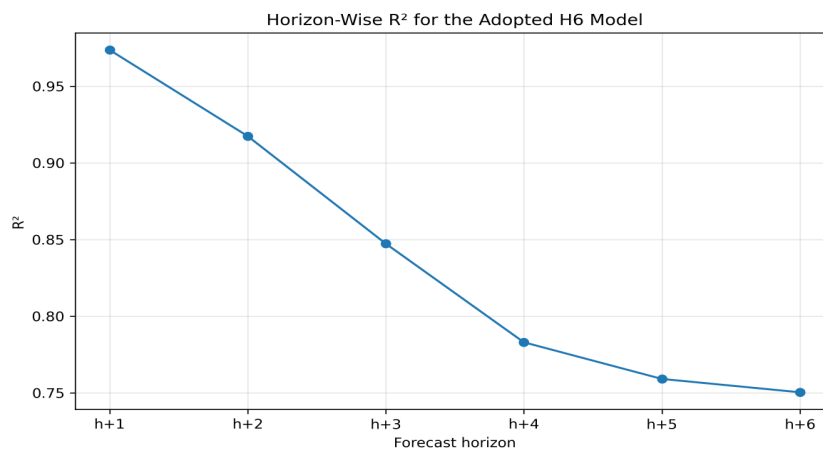


Fig. 4 Horizon-wise R² of Blend v2 + History LSTM from h+1 to h+6.

C. City-level diagnostics

City-level diagnostics confirm that performance was not uniform across the 15 locations. Qadisiyyah had the lowest RMSE (489.31 kW), while Wasit had the lowest MAE (286.14 kW) and highest R² (0.8626). Kirkuk had the highest RMSE and MAE, followed by Nineveh. Fig. 7 shows the corresponding bias across

the same wind-speed regimes, highlighting the negative bias in the 9-12 m/s band. Table 5 reports the complete city-level count, RMSE, MAE, bias, and R² results for all 15 Iraqi locations. Fig. 5 retains the compact RMSE visualization used for the displayed city-level diagnostic subset.

Table 5: City-level performance of Blend v2 + History LSTM across all 15 Iraqi locations.

City	Count	RMSE kW	MAE kW	Bias kW	R²
Anbar	8,150	555.14	342.25	-6.01	0.8195
Babylon	8,297	526.09	325.30	-6.20	0.8371
Baghdad	8,294	528.64	338.05	0.97	0.8251
Basra	8,292	524.66	293.52	-9.55	0.8462
Dhi Qar	8,302	512.62	293.37	4.52	0.8511
Diyala	8,315	513.03	315.54	-31.07	0.8483
Karbala	8,229	539.75	335.55	2.00	0.8232
Kirkuk	7,697	593.73	370.44	-39.83	0.7890
Maysan	7,862	514.92	312.44	-8.58	0.8526
Muthanna	8,280	529.15	300.36	-21.32	0.8406
Najaf	8,311	529.96	322.60	-10.26	0.8350
Nineveh	7,691	584.14	367.47	5.51	0.7917
Qadisiyyah	8,278	489.31	293.43	-9.78	0.8599
Salah al-Din	8,311	525.04	316.39	20.24	0.8335
Wasit	8,229	492.42	286.14	-17.59	0.8626

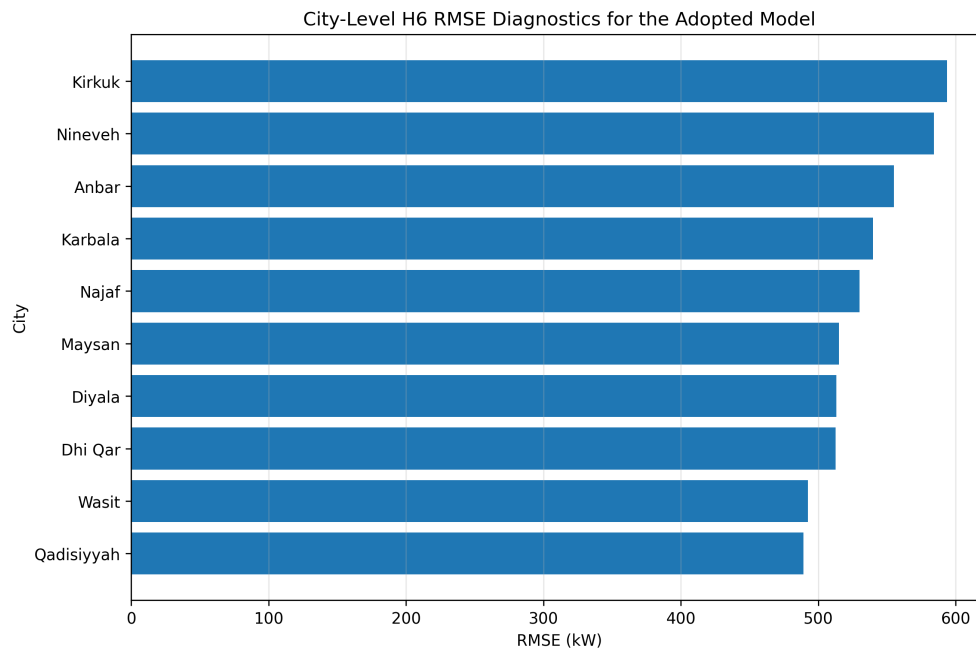


Fig. 5 City-level H6 RMSE diagnostics for the displayed city subset.

D. Speed-regime diagnostics

Speed-regime diagnostics are important because the turbine power curve is nonlinear. The highest RMSE occurred in the 9-12 m/s band, reaching 627.25 kW, with a bias of -337.98 kW. This band contains 32,182 forecast cases, so the behavior is not explained by a very small sample alone. The range lies on the steep transition toward rated power, where a modest wind-speed underestimation can produce a much larger negative power error; smoothing in the component forecasts or residual NWP wind-speed bias may also

contribute. The present outputs do not isolate a single cause, so this regime is treated as a specific limitation of the adopted system rather than attributed to one model component. Fig. 6 visualizes the RMSE differences across the five NASA-derived 110 m wind-speed regimes reported in Table 6. Table 6 details the sample count, RMSE, percentage-capacity RMSE, MAE, and bias across the five wind-speed regimes. Whereas, Fig. 7 displayed the bias of the adopted H6 model across NASA-derived 110 m wind-speed regimes

Table 6: Speed-regime performance of the adopted model using the NASA-derived 110 m reference wind speed.

Speed regime	Count	RMSE kW	RMSE (% of 3370 kW)	MAE kW	Bias kW
0-3 m/s	11097	431.87	12.82	223.17	223.17
3-6 m/s	29474	505.28	14.99	312.93	230.65
6-9 m/s	35240	559.76	16.61	411.33	52.24
9-12 m/s	32182	627.25	18.61	366.71	-337.98
12-25 m/s	14545	289.73	8.60	87.00	-87.00

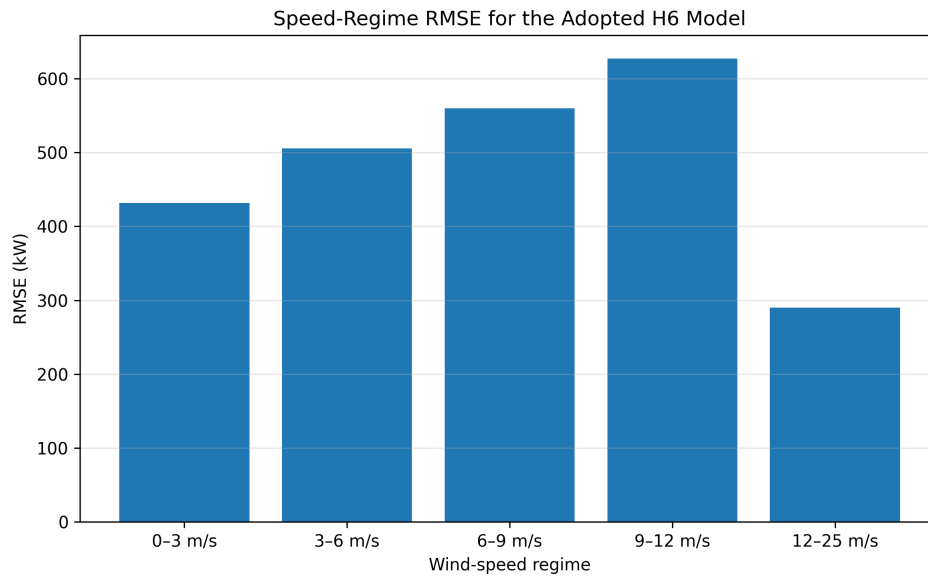


Fig. 6 RMSE of the adopted H6 model across NASA-derived 110 m wind-speed regimes.

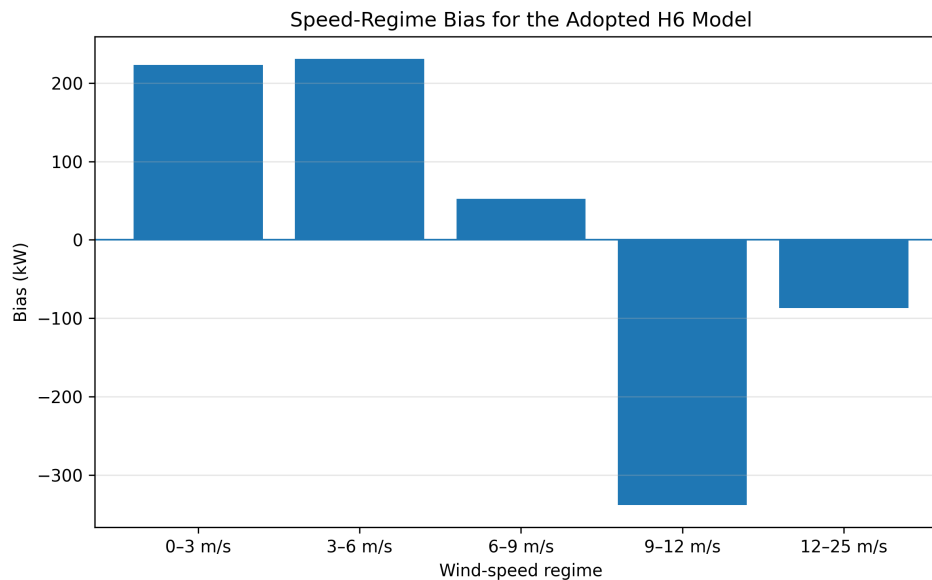


Fig. 7 Bias of the adopted H6 model across NASA-derived 110 m wind-speed regimes.

E. Energy accounting

Forecast-row energy accounting provides a cumulative view of model bias. Each forecast row represents one hourly power estimate, so its contribution is power (kW) $\times$ 1 h; row totals are summed and converted from kWh to GWh. Across all 122,538 forecast rows, the adopted model was 1.027 GWh below the NASA-derived reference total, a difference of -0.47%. Because forecasts from different origins and lead steps can refer to overlapping valid hours, the all-row total intentionally counts forecast cases rather than unique physical hours. It must not be interpreted as annual

turbine or wind-farm energy production. Table 7 provides the forecast-row energy accounting at each lead step and for all evaluated forecast rows. Fig. 8 visualizes the forecast-row energy difference at each forecast horizon. Fig. 9 compares the NASA-derived reference and predicted forecast-row energy totals at each horizon. Fig. 10 provides an illustrative h+6 time-series example for Anbar in January 2025, showing how the forecast follows the NASA-derived reference power and where larger deviations occur during rapid changes. The preceding tables provide the main quantitative evidence for model comparison.

Table 7: Forecast-row energy accounting by horizon for Blend v2 + History LSTM.

Horizon	Rows	Reference GWh	Predicted GWh	Error GWh	Error %
h+1	20505	35.94	35.88	-0.059	-0.16
h+2	20487	37.00	36.90	-0.100	-0.27
h+3	20470	38.52	38.34	-0.178	-0.46
h+4	20364	37.81	37.50	-0.310	-0.82
h+5	20362	35.65	35.37	-0.274	-0.77
h+6	20350	35.36	35.26	-0.106	-0.30
All forecast rows	122538	220.28	219.25	-1.027	-0.47

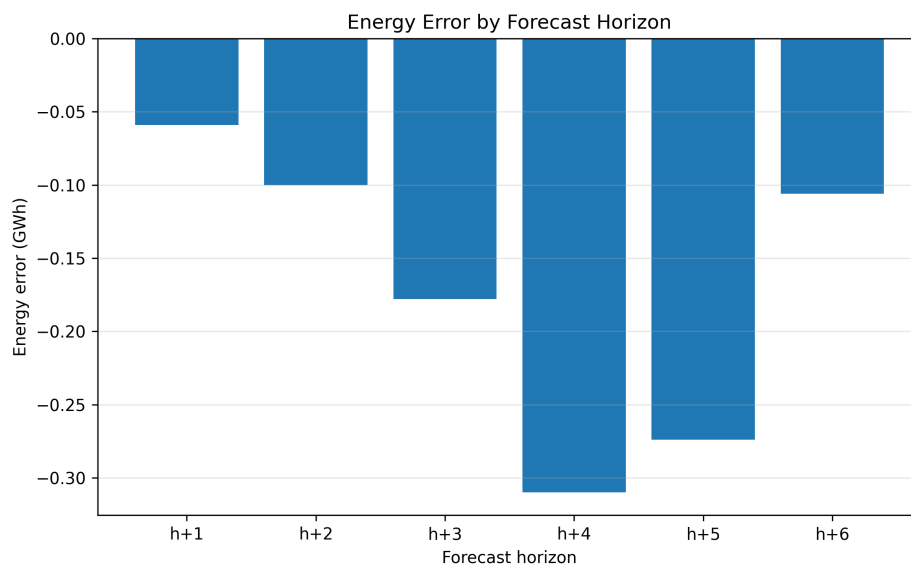


Fig. 8 Forecast-row energy difference by forecast horizon.

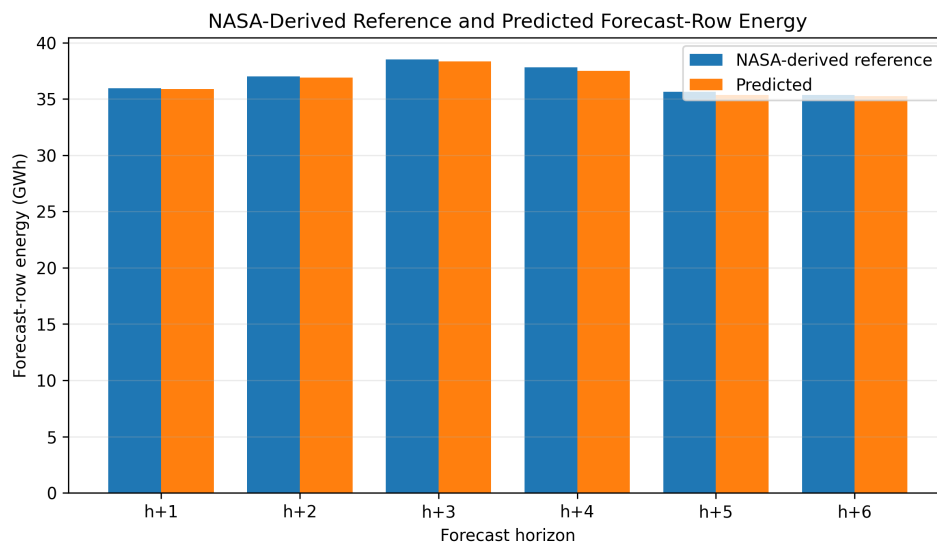


Fig. 9 NASA-derived reference and predicted forecast-row energy by forecast horizon.

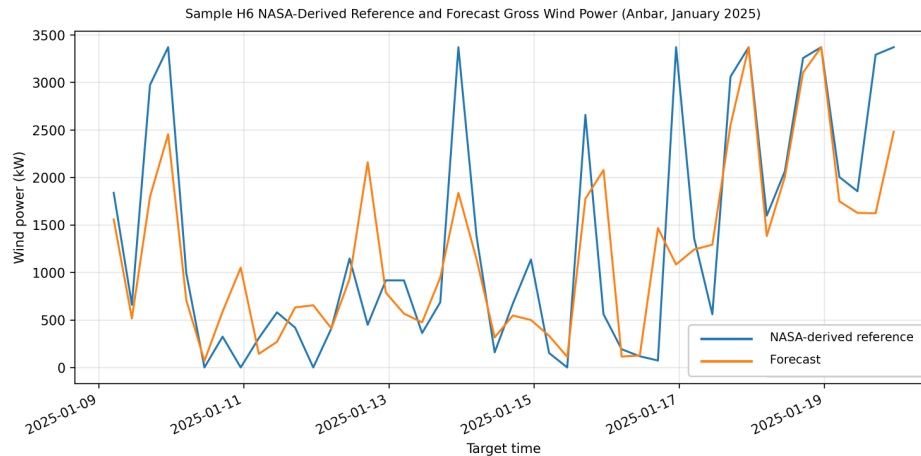


Fig. 10 Example h+6 NASA-derived reference and forecast gross wind power for Anbar in January 2025.

8. Discussion

The main operational result is that blending complementary forecast paths improved the overall H6 error relative to the individual history-only and raw-GFS references. Raw GFS showed a pronounced negative bias, whereas Power MOS showed a positive bias; calibration and Ridge blending therefore reduced dependence on the systematic Error of any single path. The comparison is an end-to-end system comparison because the branches do not all receive identical predictor information.

Forecast skill declined with lead time, while the 9-12 m/s operating region remained the main regime-specific weakness. The large negative bias in this band is consistent with the steep turbine-curve transition amplifying wind-speed underestimation. However, the retained outputs do not allow the effect to be separated cleanly from NWP bias or ensemble smoothing. City-level differences also remained after pooled training; city indicators can represent persistent offsets, but the framework does not model direct spatial dependence between cities.

The target must also be interpreted carefully. It is a NASA-derived gross-power proxy for one standardized reference turbine, not measured production from operating Iraqi wind farms. This common target is useful for controlled comparison across locations, but wake losses, availability, curtailment, maintenance, electrical losses, and alternative turbine characteristics are outside its scope. Likewise, the -0.47% energy difference is a forecast-row cumulative-bias diagnostic, not a physical annual-energy estimate.

The separate NASA-only strict diagnostic, TCN, and TCCNN experiments are not part of the official NASA-GFS ranking because their information protocol and chronology differ. Finally, the reported RMSE reductions should be read as descriptive differences rather than statistical significance claims; a future paired uncertainty analysis requires a common retained case-level prediction archive for all compared systems.

9. Conclusion

This study evaluated short-term forecasting of a NASA-derived gross reference-turbine power proxy across 15 Iraqi cities using a 3370 kW reference-turbine curve. The historical NASA-derived target covers 2015-2025, while the official GFS-aligned H6 experiment uses the common 2021-2025 period. The final split used 2021-2023 for training, 2024 for validation and blend calibration, and 2025 for independent testing. On the 2025 test set, Blend v2 + History LSTM achieved 530.75 kW RMSE, 320.41 kW MAE, -8.38 kW bias, and an R^2 of 0.8395.

The results show that multi-model forecast blending consistently improves over raw GFS baselines, MOS alternatives, standalone neural models, and history-only forecasting. The main remaining challenge is the 9-12 m/s speed range, where the nonlinear transition toward rated power leads to larger errors and stronger under-prediction.

Future work should evaluate the framework against measured turbine or wind-farm power, examine other turbine configurations, preserve common case-level prediction archives for paired uncertainty testing, improve rapid-change and 9-12

m/s regime performance, and extend the framework to probabilistic forecasting and longer archived NWP horizons.

Acknowledgment

The authors acknowledge NASA POWER and NOAA/NCEI for providing the meteorological and archived numerical-weather data services used in this study.

Data Availability

The code, retained model configurations, preprocessing procedures, final Ridge coefficients and fitted meta-model, supporting result summaries, and aligned 2025 final-blend prediction table for the H6 experiments are publicly available at <https://github.com/abdjasm87-art/iraq-h6-wind-power-forecasting> and are permanently archived on Zenodo as version v1.0.0 (DOI: 10.5281/zenodo.22070420). The underlying NASA POWER and archived GFS data are publicly available from the services cited in [2] and [3].

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

- [1] Mahmood, A. Y., Saffer, K. M., Al-Nuaimi, B. T., & Bakr, D. I. (2024). Climate change prediction in the Middle East using an RNN model and a hybrid LSTM & GRU model. In *2024 Antennas Design and Measurement International Conference (ADMInC)* (pp. 125-135). IEEE. <https://doi.org/10.1109/ADMInC63617.2024.10775334>
- [2] NASA Langley Research Center. (n.d.). *POWER Hourly API Documentation*. NASA Prediction of Worldwide Energy Resources Project. <https://power.larc.nasa.gov/docs/services/api/temportal/hourly/>
- [3] NOAA National Centers for Environmental Information. (n.d.). *Global Forecast System (GFS)*. National Centers for Environmental Prediction. <https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast>
- [4] Bortolotti, P., Tarres, H. C., Dykes, K., Merz, K., Sethuraman, L., Verelst, D., & Zahle, F. (2019). *IEA Wind TCP Task 37: Systems Engineering in Wind Energy - WP2.1 Reference Wind Turbines* (NREL/TP-5000-73492). National Renewable Energy Laboratory. <https://doi.org/10.2172/1529216>
- [5] Al-Yozbaky, O. S. A.-D., & Khalel, S. I. (2022). The future of renewable energy in Iraq: Potential and challenges. *Indonesian Journal of Electrical Engineering and Informatics*, 10(2), 273-291. <https://doi.org/10.52549/ijeei.v10i2.3756>
- [6] Alshamri, H., Cockerill, T., Tomlin, A. S., Al-Damook, M., & Al Qubeissi, M. (2024). On-off-grid optimal hybrid renewable energy systems for house units in Iraq. *Clean Technologies*, 6, 602-624. <https://doi.org/10.3390/cleantechnol6020032>
- [7] Al-Mamory, L. F., Akay, M. E., & Abdul Wahhab, H. A. (2025). Technical and economic assessment of the implementation of 60 MW hybrid power plant projects (wind, solar photovoltaic) in Iraq. *Sustainability*, 17(13), 5853. <https://doi.org/10.3390/su17135853>
- [8] Hadi, F. A., Abdulwahab, R. A., & Al-Khafaji, K. (2026). Building a classification map of wind turbine characteristics compatible with the winds of middle and southern regions in Iraq. *Wind*, 6(2), 15. <https://doi.org/10.3390/wind6020015>
- [9] Abbas, T. A., Al-Jiboori, M. H., & Altmimi, A. I. (2025). Wind analysis for power generation in the south of Iraq. *Nature Environment and Pollution Technology*, 24(1), D1687. <https://doi.org/10.46488/NEPT.2025.v24i01.D1687>
- [10] Mohammed, I. J., Al-Nuaimi, B. T., & Baker, T. I. (2022). Weather forecasting over Iraq using machine learning. *Journal of Artificial Intelligence and Metaheuristics*, 2(2), 39-45. <https://doi.org/10.54216/JAIM.020204>
- [11] Aziz, E. F., Daoud, R. W., Algburi, S., Ahmed, O. K., & Yassen, K. F. (2024). Estimation of wind speed by an artificial intelligence method: A case study. *Journal of Thermal Engineering*, 10(5), 1347-1361. <https://doi.org/10.14744/thermal.0000874>
- [12] Saleh, H. H. (2025). Wind power prediction based on polynomial regression method. *El-Cezeri Journal of Science and Engineering*, 12(3), 274-282. <https://doi.org/10.31202/ecjse.1722153>
- [13] Aldarraj, M., Vega-Márquez, B., Pontes, B., Mahmood, B., & Riquelme, J. C. (2024). Addressing energy challenges in Iraq: Forecasting power supply and demand using artificial intelligence models. *Heliyon*, 10, e25821. <https://doi.org/10.1016/j.heliyon.2024.e25821>

- [14] Hussain, Z. S., Danha, N. Y., Muheden, K. M., & Kareem, S. W. (2022). Wind speed prediction for Duhok city using a recurrent neural network. *International Journal of Intelligent Systems and Applications in Engineering*, 10(3s), 180-188. <https://www.ijisae.org/index.php/IJISAE/article/view/2429>
- [15] Shringi, S., Saini, L. M., & Aggarwal, S. K. (2025). A review of data-driven deep learning models for solar and wind energy forecasting. *Renewable Energy Focus*, 55, 100739. <https://doi.org/10.1016/j.ref.2025.100739>
- [16] Verdone, A., Panella, M., De Santis, E., & Rizzi, A. (2025). A review of solar and wind energy forecasting: From single-site to multi-site paradigm. *Applied Energy*, 392, 126016. <https://doi.org/10.1016/j.apenergy.2025.126016>
- [17] Wang, H., Yan, J., Zhang, J., Liu, S., Liu, Y., Han, S., & Qu, T. (2024). Short-term integrated forecasting method for wind power, solar power, and system load based on variable attention mechanism and multi-task learning. *Energy*, 304, 132188. <https://doi.org/10.1016/j.energy.2024.132188>
- [18] Bashir, T., Wang, H., Tahir, M., & Zhang, Y. (2025). Wind and solar power forecasting based on hybrid CNN-ABiLSTM and CNN-Transformer-MLP models. *Renewable Energy*, 239, 122055. <https://doi.org/10.1016/j.renene.2024.122055>