Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Weather–power mapping inference for wind power forecasting across heterogeneous sites

Weiye Song
North China Electric Power University
Jie Yan
yanjie@ncepu.edu.cn
North China Electric Power University
Shuang Han
North China Electric Power University
Kaiyun Liu
North China Electric Power University
Bo Wang
China Electric Power Research Institute
David Infield
University of Strathclyde
Yongqian Liu
North China Electric Power University
Jinyue Yan
Hong Kong Polytechnic University

Article

Keywords: wind power forecasting, transferable learning, sample-conditioned mapping, retraining-free deployment, wind farm heterogeneity, extreme events, reserve procurement, imbalance settlement

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11417289/v1

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

Additional Declarations: No competing interests reported.

Weather–power mapping inference for wind power forecasting across heterogeneous sites

Weiye Song[1], Jie Yan[1,*], Shuang Han[1], Kaiyun Liu[2], Bo Wang[3], David Infield[4], Yongqian Liu[1], Jinyue Yan[5]

[1] State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of New Energy, North China Electric Power University, Beijing 102206, China

[2] MOE Key Laboratory of Resources and Environmental Systems Optimization, College of Environmental Science and Engineering, North China Electric Power University, Beijing 102206, China

[3] China Electric Power Research Institute, Beijing, China

[4] Electronic and Electrical Engineering Department, University of Strathclyde, Glasgow, UK

[5] Department of Building Environment and Energy Engineering, Hong Kong Polytechnic University
Kowloon, Hong Kong, China

***Corresponding author.** E-mail: yanjie@ncepu.edu.cn. Fax: +86 6177 2046.

Abstract

Accurate wind power forecasting is critical for increasingly weather-sensitive power systems. Forecast accuracy depends on the meteorology–power mapping, which varies across farms and evolves with operating conditions. Existing approaches either fit site-specific mappings from limited local data, making them vulnerable to poorly represented operating conditions, or learn a shared mapping across farms and risk blurring local response differences. The challenge is to exploit cross-farm data while retaining mappings that adapt to local and evolving conditions. Here we develop UniWind, which infers target-specific mappings from a shared cross-farm prior rather than relearning each farm's complete mapping. Using data from 576 wind farms spanning all nine wind-energy environments in China, UniWind learns a shared weather–power prior to reduce reliance on conditions represented at any single farm, and uses target-farm observations to infer local deviations, adapting this prior to the current site and operating regime. On the held-out test sets of these farms, UniWind achieves 82.98% mean accuracy, outperforming the strongest baseline by 1.28 percentage points. It further achieves 84.05% accuracy on 142 farms excluded from training. UniWind also improves forecasts of rare ramp and low-power events, reduces estimated upward-reserve requirements by 5.0 MW per 100 MW at 95% reliability, and cuts fleet-wide model-update computation by 96.7%.

Keywords: wind power forecasting; transferable learning; sample-conditioned mapping; retraining-free deployment; wind farm heterogeneity; extreme events; reserve procurement; imbalance settlement

Wind power has come to play a pivotal role in the global transition toward sustainable energy [1], [2]. By 2025, China's installed wind power capacity had increased nearly twentyfold over the preceding 15 years, and wind power accounted for 10.67% of the country's electricity generation [3]. Looking further ahead to 2060, under the overarching goal of carbon neutrality, wind power is expected to play a dominant role in the energy mix [4]. However, this rapid expansion poses significant operational challenges for the power system, as it becomes increasingly susceptible to weather-driven variability with higher levels of wind power integration [5]-[7]. Accurate wind power forecasting is critical for mitigating weather-driven variability, as improved accuracy reduces reserve procurement, real-time balancing needs, and redispatch requirements, thereby lowering operational costs [8], [9]. Crucially, at high wind penetration, sudden wind ramps can erode operational flexibility and increase stability risk [10]. Forecasting therefore becomes a security-critical requirement, not merely an economic tool [11]. A National Renewable Energy Laboratory (NREL) report shows that savings from a 10% forecast-accuracy improvement grow sharply with wind penetration in the California Independent System Operator, increasing from ~ USD 5.05 million annually at ~8% penetration to ~USD 25.10 million at 25% (~fivefold) [12].

Most existing methods are wind-farm-specific models (WFSMs), focusing on individual farms and using their operational data to model site-level weather-to-power relationships. These approaches perform well when data are abundant, and conditions are stable, with modeling approaches spanning physical models, traditional machine learning, and deep learning [13], [14], [15], [16]. However, because the learned mapping is tied to the source site and its observed weather regimes, accuracy depends heavily on the quantity and quality of local data. For new wind farms or rare weather regimes, generalization is often poor [17].

The prevailing farm-specific paradigm scales poorly with fleet expansion because each additional site requires its own model, while knowledge learned elsewhere is not directly reusable. This limitation becomes increasingly consequential as deployment accelerates: in 2025, China added approximately 120 GW of wind projects for construction and grid connection [18]. The same reliance on local data also weakens performance at newly commissioned farms and under rare weather regimes, where historical samples are limited [19], [20]. Forecast degradation in these settings can increase operating costs and complicate performance assessment, reserve scheduling, and balancing dispatch [21], [22].

To overcome this scaling limitation, recent work has explored time-series foundation

models (TSFMs) [23]–[25], which are pretrained on large, heterogeneous time-series datasets [26], [27]. By sharing parameters and representations across sites, a single pretrained model can serve multiple wind farms, reducing the need for repeated site-specific training and maintenance. However, parameter sharing alone does not resolve cross-farm heterogeneity: similar weather conditions can produce different power responses at different sites [28], and a fully shared model may average over these site-specific differences and limit forecasting accuracy.

The central challenge is therefore to exploit what is shared across wind farms without averaging away their site-specific weather–power relationships [29]. This suggests a different forecasting principle: learn what is shared across farms, while inferring what remains site specific [30]. Based on this principle, we develop UniWind, a fleet-scale forecasting framework that learns a shared weather–power prior from multi-farm data and adapts it to individual farms at inference time, as shown in Fig. 1. Using recent weather–power observations from the target farm as context, UniWind infers how the shared prior should be adjusted to the local forecasting condition without updating model parameters. We evaluate UniWind across 576 wind farms spanning all nine wind-energy environments in China, followed by a strict generalization test on 142 additional farms withheld entirely from training. We further examine whether cross-farm knowledge sharing remains effective during rare low-power and ramp events and whether its forecasting gains translate into operational and computational benefits at fleet scale.

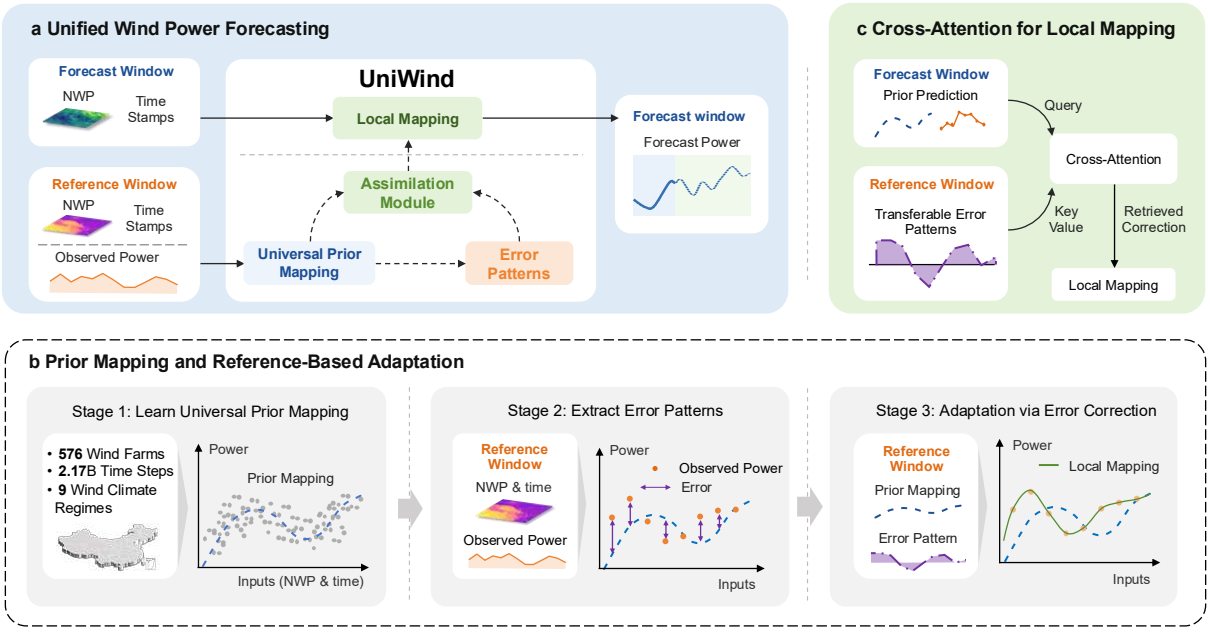


Fig. 1 | Overview of the UniWind framework. (a) UniWind replaces farm-specific

forecasting models with a unified model that learns a local weather–power mapping from forecast-window NWP inputs and reference weather–power samples. (b) UniWind first learns a shared prior mapping from large-scale multi-farm data, then extracts reference error patterns and uses them to adapt the prior to the target forecasting scenario. (c) Cross-attention retrieves the most relevant error patterns from reference memory and injects them as correction signals to form the local mapping.

Results

Fleet-scale forecasting across heterogeneous wind farms in China

UniWind replaces the traditional farm-specific modeling paradigm with a single fleet-level model that shares forecasting knowledge across heterogeneous wind farms while preserving high site-level accuracy. As shown in Fig. 2a, UniWind is trained on weather–power data from 576 wind farms spanning all nine wind-energy environments in China [31]. The model consists of a shared base predictor and a reference-based adaptation module. The shared predictor learns a weather–power prior from large-scale multi-farm data, while the adaptation module uses historical target-farm context to retrieve transferable error patterns and correct the prior prediction. During inference, UniWind requires no site-specific retraining or parameter updates [32], [33]. This design allows shared cross-farm knowledge to be adapted to local forecasting conditions and enables generalization across diverse terrains, climates, and wind-energy environments.

UniWind leverages historical mapping samples from the target wind farm as reference information to model its weather–power relationship under the current forecasting scenario, thereby enabling effective adaptation to varying environmental conditions. We used the final 180 days of previously unseen data from all 576 wind farms as the test set. Province-level averages of wind-farm prediction accuracy are shown in Fig. 2b, with representative forecasting trajectories provided in Supplementary Fig. 1. UniWind improves accuracy over the best WFSM by 1.28 percentage points (pp), achieving an average accuracy of 82.98% (Fig. 2c and Supplementary Table 1). The WFSM baselines include iTransformer [34], DLinear [35], and XGBoost [36]. UniWind achieves a prediction accuracy that is 3.16 percentage points higher than that of mainstream time series foundation models, with all models trained on the same data. The TSFMs compared in this paper include Moirai [37], TimeLLM [38], and TimeMixer [39]. Although both UniWind and these foundation models are trained on large-scale data, UniWind explicitly emphasizes the learning of

heterogeneous features of wind farms, particularly the wind-to-power conversion characteristics. In contrast, the evaluated foundation models primarily learn temporal mappings from historical sequences and weather forecasts to future power output. This distinction likely contributes to UniWind's accuracy advantage.

Performance gains varied across terrain types and wind-resource levels (Fig. 2d,e). Relative to the strongest WFSM, UniWind achieved the largest mean gain on plains (1.45 pp), followed by offshore sites (1.20 pp; Supplementary Table 2). Across wind-power-density classes, the gains were 1.72 pp in the lowest-resource category (<70 W m^{-2}) and 1.61 pp in the highest-resource category (Supplementary Table 3), whereas performance on plateaus was comparable to that of WFSMs. Overall, the magnitude of the improvement depended on the operating environment.

The observed advantages also occurred in regions relevant to China’s wind power development. Major wind power provinces such as Gansu, Hebei, and Shandong are dominated by plains [40]. At the same time, the future development plan also prioritizes offshore regions with higher wind-resource levels. These are precisely the scenarios in which UniWind performs best. Overall, UniWind provided improvements across most terrain and resource categories, although the magnitude of the gains varied among operating environments.

Five representative wind farm case studies further validated UniWind’s generalization capability (Fig. 2f–j). These farms are located across Gansu, Guangdong, Yunnan, Hebei, and Anhui and differ in location, terrain, and wind-resource conditions. UniWind produces more accurate short-term forecasts across plains, mountains, coastal regions, and inland high-wind regions, capturing power fluctuations better than the strongest WFSM. Across these five cases, UniWind achieves the largest gain of 9.03 pp over the strongest baseline, and the gain remains 2.14 pp even in the most challenging case. These results support UniWind as a scalable and general forecasting framework.

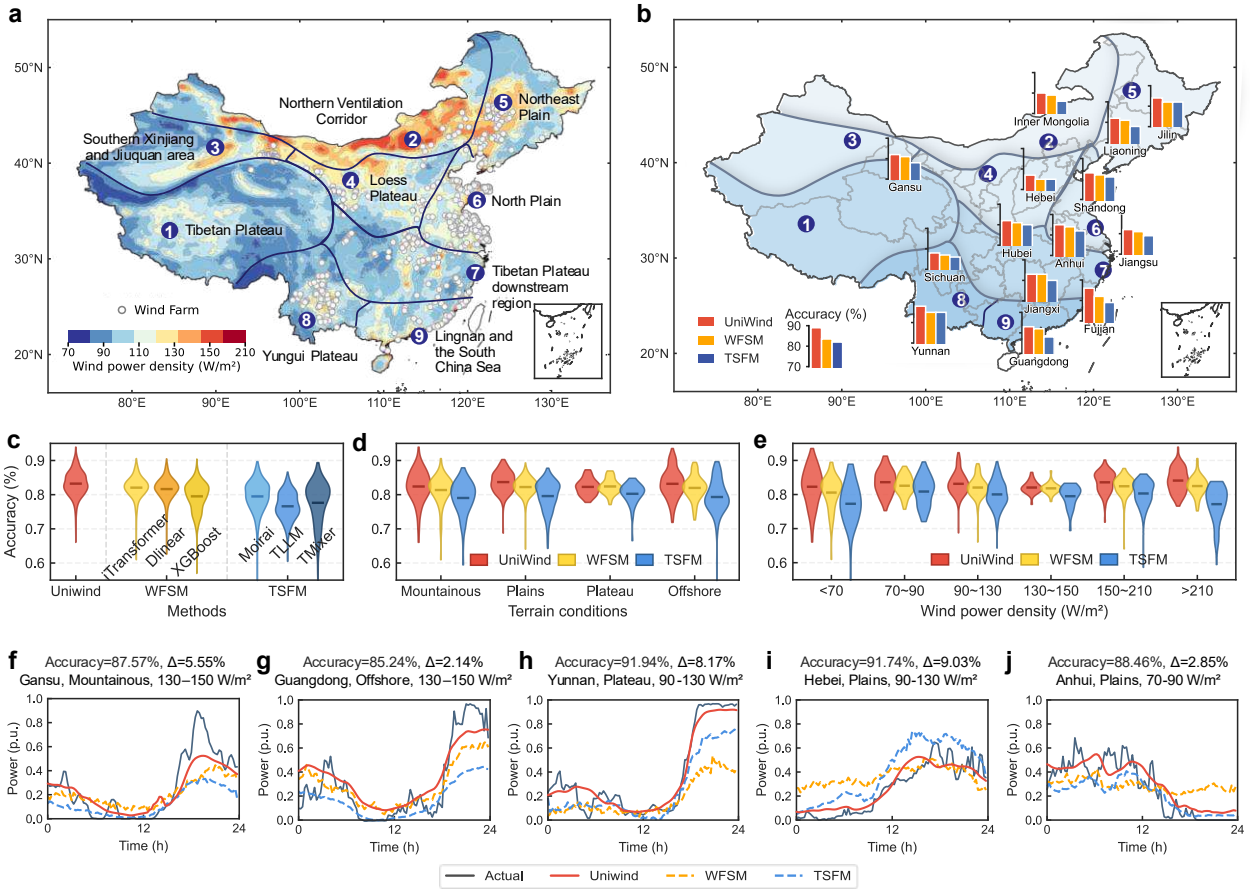


Figure 2 | Spatial distribution and stratified evaluation of wind power forecasting accuracy in China. (a) National wind power density map and nine representative regions; dots indicate wind farm locations. (b) Province-level accuracy comparison across forecasting paradigms (UniWind, WFSM, and TSFM). (c) Accuracy distributions across all wind farms for each method. (d) Accuracy distributions by terrain type (mountain, plain, plateau, offshore). (e) Accuracy distributions by wind power density class. (f–j) Case studies for five representative wind farms. Curves show observations and predictions from each method.

Generalization to unseen wind farms without site-specific retraining

UniWind can be deployed to unseen wind farms with high accuracy without site-specific retraining or parameter tuning. During forecasting, recent weather–power pairs from the target farm provide contextual information about its local weather–power response. Guided by the shared prior learned across farms, UniWind uses this context to adapt the forecast to the target site. We evaluate generalization under a strict no-parameter-update

setting, in which target-farm data are excluded from training and used only as contextual input during inference. This setting directly tests whether knowledge learned across farms can be combined with target-site context to generalize to out-of-distribution wind farms [41].

We evaluated UniWind on 142 wind farms that were completely withheld from training. These sites span all nine wind-energy environments in China and cover major terrain and climate zones (Fig. 3a). UniWind achieved an average accuracy of 84.05%, exceeding the strongest WFSM trained independently with 10% local data (82.38%; Fig. 3b and Supplementary Table 4). Thus, cross-farm pretraining can provide useful site-level forecasts before a separate local model is trained.

Fig. 3c examines how forecasting accuracy changes with target-site data availability. UniWind maintains the highest accuracy and the most stable performance across the full range of data availability. This robustness indicates that cross-farm training captures transferable weather–power structure, reducing dependence on extensive target-site training data. By contrast, WFSMs are more sensitive to data availability, with performance degrading sharply in small-sample settings and remaining below UniWind as more local data become available. TSFMs also benefit from large-scale pretraining and are less sensitive to target-site data volume (Fig. 3b–c). However, their lower overall accuracy indicates that large-scale pretraining alone is insufficient; effective cross-farm forecasting also requires explicit adaptation to heterogeneous weather–power responses [42].

UniWind outperforms the strongest baseline across the test provinces, further demonstrating its robust performance across diverse terrains and climatic conditions (Fig. 3d). This advantage is particularly pronounced in provinces where WFSMs perform relatively poorly (Supplementary Table 4). In Jilin (83.08%), Hebei (83.17%), and Jiangsu (83.18%), where WFSMs perform worst, UniWind achieves larger accuracy gains than in other provinces when trained on the full dataset, with improvements of 1.69, 2.05, and 2.24 percentage points, respectively. This pattern is consistent with the design of UniWind. By leveraging shared structure learned from large-scale cross-site data, UniWind depends less on target-site data and remains more robust under unfavorable data conditions. In contrast, because WFSMs rely solely on single-site data, they are more susceptible to performance degradation when samples are limited, scenario coverage is insufficient, or observations are noisy or incomplete.

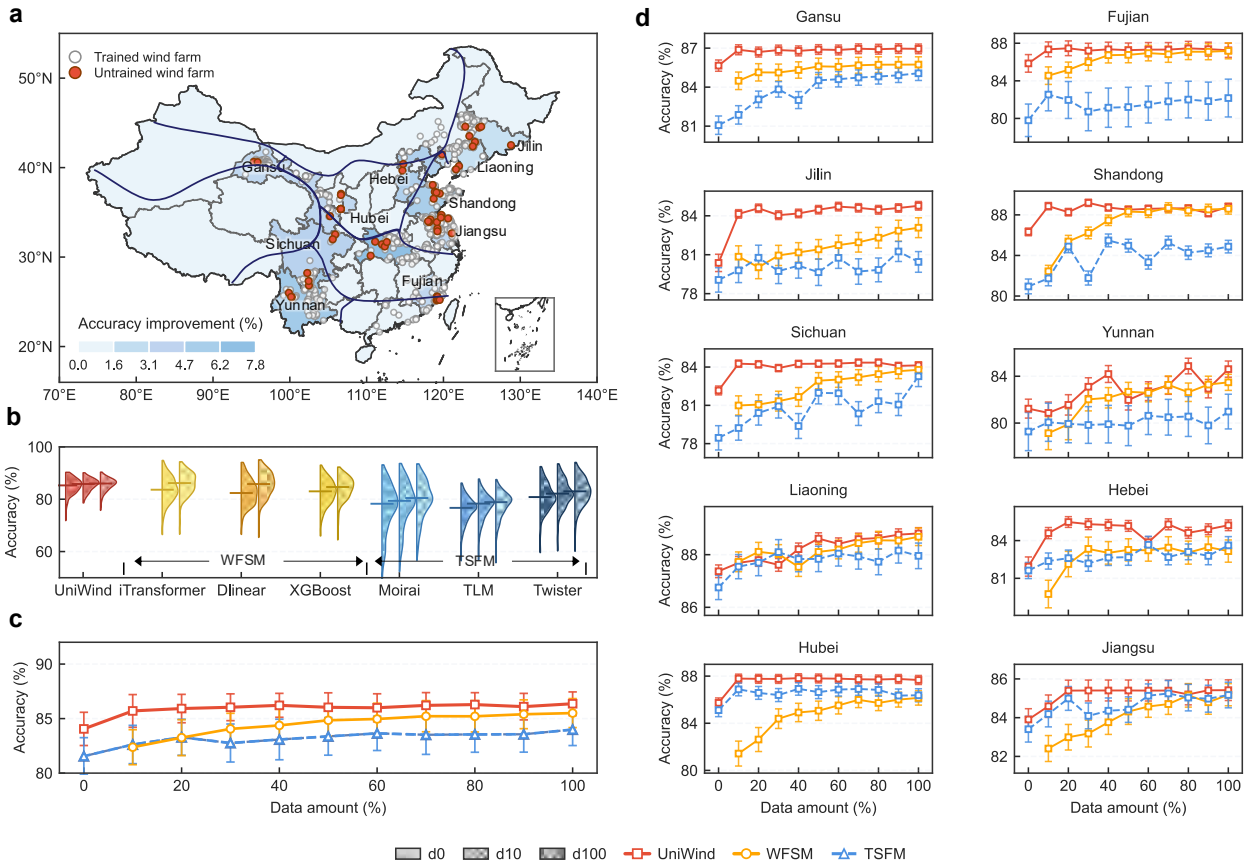


Figure 3 | Parameter-free generalization to unseen wind farms and cross-province stability (accuracy measured as (1 − NRMSE) × 100%). (a) Spatial distribution of training and test wind farms, and the geographic distribution of parameter-free gains. The test set includes 142 unseen wind farms across all nine wind-energy environments in China. (b) Overall comparison. UniWind achieves 84.05% mean accuracy under parameter-free conditions. It exceeds WFSMs trained with 10% local data (82.38%) and TSFM baselines. (c) Sensitivity to data availability. UniWind remains accurate and stable as target-site data increases; WFSM improves but stays below UniWind at matched target-data availability; TSFM gains are limited. (d) Province-level performance. UniWind outperforms WFSMs in all provinces.

Forecasting rare but system-critical wind-power events

Rare but severe wind-power events, including prolonged low-output periods and sharp ramps, can challenge resource adequacy and real-time balancing [43], [44]. Their scarcity in historical data makes them a stringent test of forecasting generalization [45].

We focus on two operationally critical event types: long-lasting low wind power events

(LLWPEs) [46] and large wind power ramp events (LWPREs) [47]. LLWPEs cause sustained power deficits, increase reserve needs, and elevate operational risk during tight supply–demand periods, yet account for only 2.15% of the samples (Fig. 4a). LWPREs produce sharp short-term power changes and increase real-time balancing and corrective-action requirements, yet account for only 0.46% of the samples (Fig. 4e). Event definitions and metrics are provided in the Supplementary Information.

On average, UniWind outperformed the strongest WFSM in both event detection and within-event power forecasting (Fig. 4b,c,f,g). For LLWPE detection, UniWind improved AUPRC and AUROC by 7.41 and 5.38 percentage points, respectively (Supplementary Table 5; Fig. 4b). For LWPRE detection, the corresponding gains were 1.52 and 5.26 percentage points (Supplementary Table 6; Fig. 4f). Within event windows, 1−NRMSE improved by 3.15 percentage points for LLWPEs and 3.63 percentage points for LWPREs relative to the strongest WFSM (Supplementary Tables 5–6; Fig. 4c,g). These results show that UniWind improves both event identification and power-trajectory forecasting during rare high-risk periods.

Aggregate scores can mask whether forecast gains persist across individual event windows. We therefore quantified the gain over WFSMs for each forecast window (Fig. 4d,h). Across both LLWPEs and LWPREs, improvements were observed for the majority of forecast windows under all four metrics. For LLWPEs, the fraction of improved windows ranged from 61.3% to 76.2%; for LWPREs, it ranged from 72.5% to 78.0% (Fig. 4d,h). Thus, the average gains are not driven by a small subset of events but persist across most individual forecast windows.

Representative cases illustrate how these gains appear in individual events. For the LLWPE case, the predicted onset and end of the low-output period aligned more closely with observations, while predicted power remained closer to observed levels throughout the event window (Fig. 4i). For the LWPRE case, UniWind more accurately captured the ramp magnitude and showed less lag and bias around the sharp transition (Fig. 4j). Together, the aggregate, per-window, and event-level analyses show that cross-farm knowledge sharing remains effective precisely in the rare conditions where individual wind farms provide the least training data.

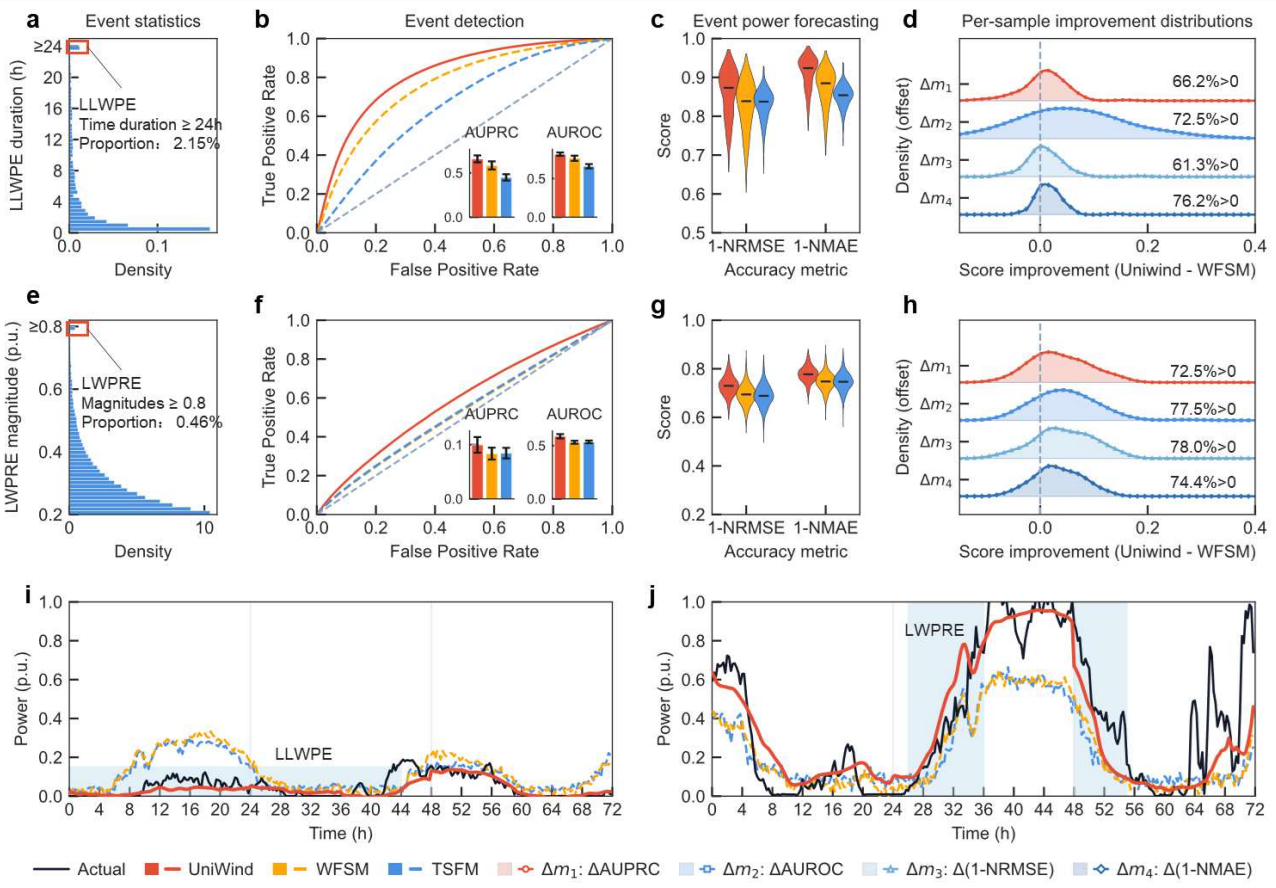


Figure 4 | Detection and within-event power forecasting for rare high-risk events. (a,e) Distributions and thresholds for LLWPE (a) and LWPRE (e). (b,f) ROC curves and key detection metrics (AUROC, AUPRC). (c,g) Within-event power forecasting scores (1 − NRMSE, 1 − NMAE). (d,h) Distributions of per-forecast-window gains relative to WFSMs. (i,j) Case studies showing forecasts for LLWPE and LWPRE events.

Operational value and scalability of fleet-wide forecasting

We next assessed whether fleet-level forecasting gains translate into operational value as wind deployment scales. We quantify these benefits through two measurable channels: grid-assessment penalties on the wind-farm side and additional upward-reserve requirements on the system side. Under the assumptions described below, these gains correspond to an estimated CNY 3.33 billion in annual operational savings when extrapolated to China's 640 GW of installed wind capacity in 2025. Under the installed-capacity projections reported by China Energy Investment Corporation [48], this estimate rises to CNY 15.59 billion by 2060 (Fig. 5a). The savings reported in Fig. 5a include only these two quantified channels. We additionally assess the computational cost of scaling

forecasting through fleet-wide model training and updates.

Penalty reduction is the most direct operational benefit for wind farm owners. In China, grid-assessment rules typically set accuracy thresholds for power forecasts to support stable grid operation and dispatch. Wind farms below the threshold face penalties. Fig. 5c summarizes the annual average penalties by region. Values are reported as “penalty per 100 MW wind farm”. Relative to WFSM, UniWind reduces penalties in all regions. On average, the annual penalty decreases by CNY 0.36 million per 100 MW, corresponding to an approximately 20% reduction.

On the system side, we estimate the additional upward reserve implied by forecast shortfalls. For each timestamp, we compute the shortfall as the difference between predicted and measured power when the prediction exceeds the measured power. Otherwise, the shortfall is set to zero. We then sort all shortfalls and use high quantiles to represent reserve needs under a reliability target. For example, a 95% reliability target uses the 95th percentile. A 99% target uses the 99th percentile (Fig. 5d). Per 100 MW wind farm, UniWind reduces the required upward reserve by 5.0 MW at 95% reliability and by 5.5 MW at 99% reliability. Using *capacity savings* × *reserve price* × *equivalent annual utilization hours*, we convert this into an annual saving of approximately CNY 0.16 million per 100 MW. We assume an upward *reserve price* of ~CNY 10 per MW·h [52] and *equivalent annual utilization hours* of 3119 h [18]. This estimate is a proxy derived from error statistics rather than a full-system simulation with unit commitment and real-time dispatch constraints. Sensitivity to reliability targets and price assumptions is shown in Supplementary Fig. 2.

Beyond operational gains, UniWind reduces the computational burden of fleet-wide model training and updates. It replaces repeated per-site model training with a centralized fleet-wide training process, avoiding duplicated training computation as the fleet grows. In our cost accounting, we compare only the training stage; site-level data capture and deployment are assumed to be equivalent across methods. Within this scope, one UniWind fleet-wide training run requires 192 GPU·h, whereas WFSM training requires 10 GPU·h per farm, as shown in Fig. 5b. Across 576 wind farms, one complete WFSM update therefore requires 5,760 GPU·h. UniWind reduces the training computation required for a fleet-wide update by 96.7%, with the computational advantage increasing as the number of wind farms grows (Fig. 5e).

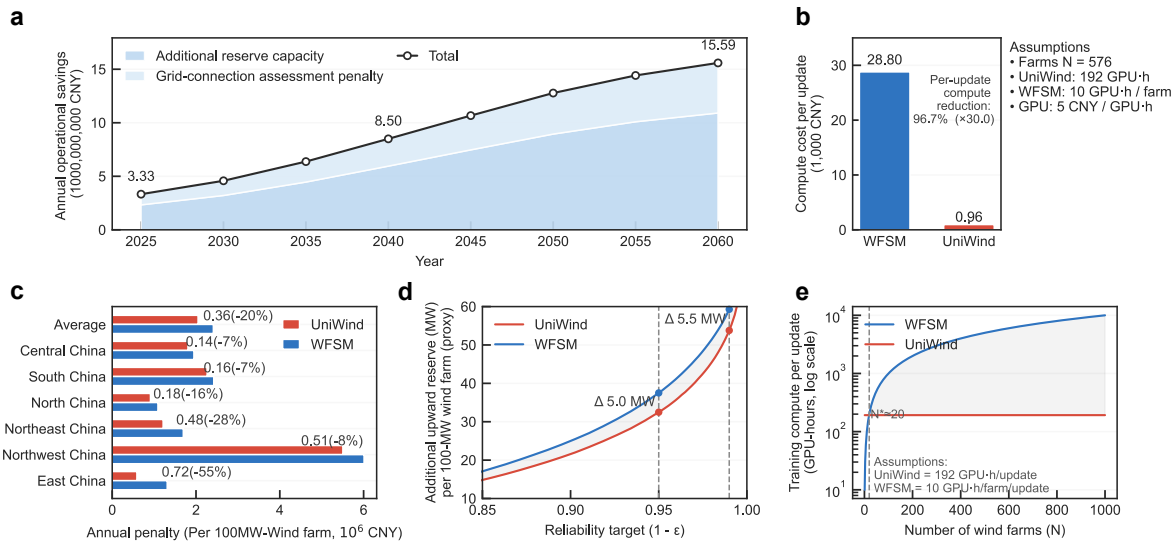


Figure 5 | Operational value and deployment scalability of fleet-wide wind forecasting. (a) Illustrative annual operational savings extrapolated along a capacity growth pathway (2025–2060; unit: 10^9 CNY). Savings are decomposed into grid-assessment penalty reduction and reduced additional upward reserve. (b) Training compute cost for one full model update over 576 wind farms (unit: 10^3 CNY; rounded). Assumptions: UniWind 192 GPU·h per update; WFSM 10 GPU·h per site per update; GPU price 5 CNY per GPU·h. Relative savings are shown. (c) Regional average grid-assessment penalties (per 100 MW wind farm; unit: 10^6 CNY). UniWind is compared with WFSM, with percentage reductions annotated. (d) Additional upward reserve needs versus reliability target ($1 - \varepsilon$) (per 100 MW wind farm). Reductions at 0.95 and 0.99 are highlighted. (e) Training compute per update versus number of wind farms N (GPU·h per update; log scale). WFSM scales linearly with N, whereas UniWind uses one centralized training run. The gap and break-even point N* are highlighted.

Discussion

As wind penetration and the number of wind farms grow, forecasting capability must scale across an expanding and increasingly heterogeneous fleet. Under the prevailing farm-specific paradigm, however, each additional site adds another independently trained model, while the knowledge accumulated across the fleet remains fragmented. This creates a scalability mismatch between rapid wind deployment and the forecasting infrastructure needed to support it. UniWind addresses this mismatch by combining fleet-level knowledge sharing with target-site adaptation. Across the evaluated wind farms, the

gains over both farm-specific and time-series foundation models show that cross-farm knowledge sharing need not come at the expense of site-specific forecasting accuracy.

The results suggest that cross-farm generalization does not require a fixed universal weather–power mapping. Instead, wind farms share transferable weather–power structure, while their site-specific deviations remain important for accurate forecasting. UniWind operationalizes this view by learning a shared prior from large-scale multi-farm data and using target-site weather–power context to adapt that prior during forecasting. This design enables the model to exploit cross-farm diversity without averaging away local response differences, thereby improving accuracy while avoiding farm-specific retraining.

This generalization capability changes how forecasting models can be deployed at new wind farms. On 142 previously unseen wind farms, UniWind achieved 84.05% accuracy, outperforming WFSMs trained on 10% of the local data and approaching fully trained site-specific models. This shifts the operational workflow from repeated local model development to centralized training and direct deployment, allowing forecasting capability to expand without adding a separately trained model for every new wind farm. It is particularly relevant to newly commissioned or data-scarce wind farms, where sufficient local records for conventional model development may not yet be available.

The event-level results suggest that cross-farm knowledge sharing may be particularly valuable when local examples are scarce. By learning recurring event patterns from a large and diverse set of wind farms, UniWind improves both event detection and within-event power forecasting for LLWPE and LWPRE, even when such events are rarely observed at individual sites. More reliable identification of event onset and occurrence can support earlier risk awareness, while improved trajectory forecasts provide information on event duration, severity, and recovery. These capabilities may offer more informative inputs for reserve scheduling, balancing actions, and operational preparedness during rare but high-impact periods.

The improved forecasting accuracy also translates into measurable operational value. For wind farm owners, lower forecast errors can reduce deviation penalties, while for the power system they can reduce the upward reserves required to cover forecast shortfalls. Under representative assumptions, the corresponding annual savings are estimated at CNY 0.36 million and CNY 0.16 million per 100 MW, respectively. Under the same assumptions, a simple capacity-based extrapolation corresponds to approximately CNY 3.33 billion in annual operational savings at China’s 2025 installed wind capacity and CNY 15.59 billion by 2060 under the assumed growth scenario. Centralized fleet-wide training

further reduces the computation required for a full model update by 96.7% relative to separate farm-specific training. Thus, fleet-level knowledge sharing can reduce the operational burden of forecast uncertainty while avoiding proportional growth in model-training effort as the wind fleet expands.

This study has several limitations. UniWind depends on the diversity and representativeness of its training data; extension beyond China will require datasets spanning a broader range of wind regimes. Performance also remains sensitive to meteorological-input quality and common SCADA issues, including missing data, sensor drift, compression artifacts, and non-stationarity caused by control updates or equipment ageing. Finally, although UniWind improves point forecasts and rare-event detection, risk-aware operation ultimately requires well-calibrated probabilistic forecasts; extending the framework to predictive distributions is therefore an important next step.

In summary, our results show that site heterogeneity does not require wind farms to be treated as independent forecasting problems. By learning transferable cross-farm structure while adapting to local weather–power responses, UniWind shifts forecasting from repeated per-site model development toward fleet-level knowledge accumulation and deployment. Across a national-scale testbed, this approach reduces reliance on local training data, generalizes to previously unseen farms, improves forecasts of rare system-critical events, and lowers both estimated reserve requirements and fleet-wide training burden. These results suggest that forecasting capability can scale with wind-power expansion through shared cross-farm knowledge rather than a proportional increase in independently trained models.

Methods

Mathematical setting

Short-term wind power forecasting uses regional numerical weather prediction (NWP) as the main input. Given an initialization time t_0, the goal is to predict the wind-farm power sequence over a forecast horizon of length L. We set $L = 96$ by default, corresponding to 24 h at 15 min resolution. Both inputs and outputs are sampled every 15 min.

For each forecast sample from wind farm s initialized at time t_0, the input has three parts. First, we use a spatiotemporal NWP grid centered at the wind farm within a $\pm 0.3^\circ$ neighborhood, $X_{s,t_0}^{\mathrm{nwp}} \in \mathbb{R}^{L \times H \times W \times M}$. The grid has spatial resolution $0.1^\circ$and includes variables such as wind speed, wind direction, temperature, and irradiance. Second, we use

time covariates, $X_{t_0}^{\text{time}} \in \mathbb{R}^{L\times 3}$, including month, day, and hour. Third, we use a historical reference set ζ_{s,t_0}, which provides recent paired samples of meteorological inputs and measured power. To avoid temporal information leakage, the reference set ζ_{s,t_0} is constructed strictly using only historical observations available prior to the forecast initialization time t_0.

Specifically, for each forecast instance, we include reference windows at $\{t_0 - gD\}_{g=1}^{G}$, where all corresponding meteorological inputs and power measurements are observed before t_0.

We construct the reference set by aligning reference windows by time of day across preceding calendar days. For a given initialization time t_0, we collect G historical windows at the same time-of-day, $\{t_0 - 1\,\text{day}, \dots, t_0 - G\,\text{day}\}$. For each window, we take its NWP inputs, time covariates, and measured power sequence:

$$\zeta_{s,t_0} = \{(X_{s,t_0-gD}^{\text{nwp}}, X_{t_0-gD}^{\text{time}}, y_{s,t_0-gD})\}_{g=1}^{G}, y_{s,\cdot} \in \mathbb{R}^{L\times 1},$$

where $D = 1$ day. The reference set characterizes the recent site-specific weather-to-power response under similar diurnal conditions.

We formalize the predictor as conditional regression:

$$\hat{y}_{s,t_0} = f_\theta(X_{s,t_0}^{\text{nwp}}, X_{t_0}^{\text{time}} \mid \zeta_{s,t_0}), \hat{y}_{s,t_0} \in \mathbb{R}^{L\times 1}.$$

Rather than retraining a separate site-specific predictor for each wind farm, the model learns a shared cross-site prior and uses the reference set to infer a local error-correction rule for the current forecasting scenario.

Shared meteorology–time encoder and base predictor

UniWind first maps the target and reference meteorology–time inputs into a shared representation space [53], [54]. At each 15 min step, NWP features and time covariates are concatenated and projected into a d_{model} -dimensional latent space. A dual-axis Transformer encoder is then applied. The encoder combines temporal positional encoding and learnable feature-axis positional encoding, allowing the model to capture both temporal evolution and feature interactions.

The same encoder is shared by the target window and all reference windows. Therefore, meteorological evolution from different sites and different days is embedded in a common feature space [55]. On top of this shared representation, a base prediction head produces a prior forecast:

$$\hat{y}_{s,t_0}^{\mathrm{prior}} = h_{\mathrm{base}}(H^{\mathrm{tgt}}),$$

where $H^{\mathrm{tgt}} \in \mathbb{R}^{L \times d_{\mathrm{model}}}$ is the encoded target representation. This prior forecast represents the globally shared weather-to-power mapping learned from multi-site data.

Reference-error memory construction

Instead of directly retrieving reference power trajectories, the current model explicitly constructs a reference-error memory. For each reference window, the shared base predictor is first applied to obtain its prior forecast:

$$\hat{y}_{s,t_0-gD}^{\mathrm{prior}} = h_{\mathrm{base}}(H_g^{\mathrm{ref}}).$$

Given the observed reference power y_{s,t_0-gD}, the model computes the reference error sequence:

$$e_g = y_{s,t_0-gD} - \hat{y}_{s,t_0-gD}^{\mathrm{prior}},$$

and its temporal difference Δe_g. The model further extracts statistical descriptors of the error trajectory, including mean, standard deviation, extrema, range, slope, peak/valley positions, median, and energy.

These components are encoded jointly with the reference meteorology representation. Specifically, the memory source for each reference token contains the reference feature representation, encoded error trajectory, prior prediction embedding, error embedding, error-difference embedding, and error-statistics embedding. Two separate projections generate memory keys and values:

$$K^{\mathrm{err}} = \psi_{\mathrm{k}}(\cdot), V^{\mathrm{err}} = \psi_{\mathrm{v}}(\cdot).$$

The key emphasizes when a reference error pattern is transferable, whereas the value stores the correction information to be read out. This design makes the reference set an explicit error-aware memory rather than a simple NWP–power lookup table.

Target error query and group-level reference selection

For the target window, the model constructs an error-aware query from three components [56]: the target meteorology representation H^{tgt}, the target prior forecast $\hat{y}_{s,t_0}^{\mathrm{prior}}$, and the temporal difference of the prior forecast. These components are projected, normalized, and fused to obtain a target error query:

$$Q^{\mathrm{err}} = \phi_{\mathrm{err}}(H^{\mathrm{tgt}}, \hat{y}_{s,t_0}^{\mathrm{prior}}, \Delta\hat{y}_{s,t_0}^{\mathrm{prior}}).$$

The model also computes a global query for selecting relevant reference groups. Each reference window is summarized into a group representation using its encoded meteorology, encoded error, prior prediction, error embedding, and error statistics. A softmax similarity between the target group query and reference group keys produces group weights:

$$\alpha_g = \mathrm{softmax}\left(\frac{q_{\mathrm{grp}}^{\top} k_{\mathrm{grp},g}}{\sqrt{d_{\mathrm{model}}}\tau}\right).$$

The memory keys and values are then fused across reference groups using α_g. This allows the model to adaptively select the most transferable historical error patterns for the current target scenario.

Gated cross-attention error correction

The final forecast is produced by gated error correction [57], [58]. The target error query Q^{err}attends to the fused reference-error memory:

$$\mathrm{Attn}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^{\top}}{\sqrt{d_{\mathrm{model}}}}\right)V.$$

A cross-attention block first refines the target query by self-attention and then retrieves correction context from the reference-error memory. The correction head outputs an estimated error adjustment $\Delta\hat{e}$. To avoid over-correction, a learnable gate controls the correction magnitude:

$$\hat{y}_{s,t_0} = \hat{y}_{s,t_0}^{\mathrm{prior}} + g_{\mathrm{corr}} \cdot \Delta\hat{e}.$$

The correction gate is initialized with a small bias so that the model starts close to the base predictor and gradually learns when reference-based correction should be trusted. This improves training stability and robustness.

Conceptually, UniWind first learns a universal prior mapping from weather to power, then uses recent reference windows to infer how that prior is likely to be biased under the current site and weather regime. Since the transferable object is the error pattern rather than site-specific model parameters, the method can generalize across scenes without retraining.

Training objectives

The current implementation uses a two-stage training strategy. In stage 1, the model trains the shared encoder and base prediction head using mean squared error between

the prior forecast and observed target power:

$$\mathcal{L}_{\text{base}} = \text{MSE}(\hat{y}^{\text{prior}}, y).$$

This stage learns the cross-site universal weather-to-power prior.

In stage 2, the reference-error memory and gated correction module are activated. The final prediction is optimized by:

$$\mathcal{L}_{\text{pred}} = \text{MSE}(\hat{y}, y).$$

The model learns to correct the base forecast by retrieving transferable reference-error patterns.

Optimization

Unless otherwise stated, we set $L = 96$, $G = 7$, and $d_{\text{model}} = 256$. The Transformer encoder uses 4 attention heads, 2 layers, and feed-forward dimension $2d_{\text{model}}$. We use dropout during training and disable it during inference. The optimizer is Adam with learning rate 5×10^{-4}. We train for at most 100 epochs with early stopping based on validation performance. The batch size is 512. We evaluate using the best checkpoint.

Initialization and training stability

We apply layer normalization, residual connections, dropout, and gated correction to stabilize training. The correction gate is initialized with a negative bias, making the initial model close to the base predictor. As training proceeds, the network learns when and how strongly to apply reference-error correction. This residual correction design reduces the risk of unstable reference retrieval and improves cross-scene robustness.

Cross-scene generalization mechanism

Under cross-scene training, UniWind learns shared meteorology–time representations and a universal base predictor from many wind farms. During inference on a new scene, the model does not update its parameters. Instead, it uses the target site's recent reference windows as conditional context. By comparing the target prior forecast and meteorological representation with reference-window error memories, the model retrieves transferable bias patterns and applies a gated correction. High-accuracy generalization is achieved without retraining: the learned parameters provide the universal prior and retrieval mechanism, while the reference set supplies scene-specific error information at inference time.

Data availability

A subset of the data supporting the findings of this study is publicly available at [https://github.com/WindFarm803/UniWind] for benchmarking and testing purposes. The full dataset is subject to third-party restrictions and is therefore not publicly available. Access to the full dataset may be granted for research verification upon reasonable request to the corresponding author and, where applicable, with permission from the data provider, under a data use agreement that restricts redistribution.

Code availability

UniWind model and an end-to-end inference example are available at https://github.com/ WindFarm803/UniWind. The same repository includes the test dataset used for evaluation.

Acknowledgments

This work was supported by Smart Grid-National Science and Technology Major Project (2025ZD0805500).

Author contributions

Weiye Song and Jie Yan conceived and designed the experiments, performed the experiments, analyzed the data, and wrote the paper. Shuang Han conceived and designed the experiments and wrote the paper. Kaiyun Liu performed the experiments and wrote the paper. Bo Wang conceived and designed the experiments, performed the experiments, and wrote the paper. David Infield and Yongqian Liu conceived and designed the experiments and wrote the paper. Jinyue Yan contributed to writing the paper. All authors discussed the results and approved the final manuscript.

Additional information

Supplementary information is available for this paper at https://github.com/ WindFarm803/UniWind.

Competing interests

The authors declare no competing interests.

References

[1] Global Wind Energy Council. Global Wind Report 2025 (GWEC, 2025). https://www.gwec.net/reports/globalwindreport/2025

[2] “Renewables 2024: Analysis and forecast to 2030,” International Energy Agency (IEA), Oct. 2024.

[3] National Bureau of Statistics of China, “Statistical Communiqué of the People’s Republic of China on the 2025 National Economic and Social Development,” National Bureau of Statistics of China, Beijing, China, Feb. 2026. Accessed: Jul. 26, 2026. [Online]. Available: https://www.stats.gov.cn/english/PressRelease/202602/t20260228_1962661.html

[4] State Grid Corporation of China, “The situation,task,challenge and realization path faced by the State Grid under the goal of carbon peak and carbon neutralization,” State Grid Corporation of China, Beijing, 2021.

[5] Y. Wang et al., “Accelerating the energy transition towards photovoltaic and wind in China,” Nature, vol. 619, no. 7971, pp. 761–767, Jul. 2023, doi: 10.1038/s41586-023-06180-8.

[6] J. Yan, “Urban energy transformation through integrated systems,” Nat Energy, vol. 11, no. 1, pp. 16–17, Jan. 2026, doi: 10.1038/s41560-025-01922-6.

[7] R. Chen, B. F. Hobbs, Z. Lu, Y. Dvorkin, and Y. Qiao, “An hourly climate projection and renewable energy generation dataset for power system modeling in China,” Sci Data, vol. 13, no. 1, p. 86, 2026, doi: 10.1038/s41597-025-06396-5.

[8] J. Wang et al., “Inherent spatiotemporal uncertainty of renewable power in China,” Nat Commun, vol. 14, no. 1, p. 5379, Sep. 2023, doi: 10.1038/s41467-023-40670-7.

[9] J. M. T. Sauer, L. D. Anadón, J. Kirchherr, J. Plummer Braeckman, and V. Schulhof, “Determinants of Chinese and Western-backed development finance in the global electricity sector,” Joule, vol. 6, no. 6, pp. 1230–1252, Jun. 2022, doi: 10.1016/j.joule.2022.05.003.

[10] J. Yan, X. Li, H. Wang, F. Si, W. Qiao, and Y. Liu, “The economic value of wind power forecasting: a data-driven method and its application in various scenarios,” Energy, vol. 340, p. 139151, Dec. 2025, doi: 10.1016/j.energy.2025.139151.

[11] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, “A review of wind speed and wind power forecasting with deep neural networks,” Applied Energy, vol. 304, p. 117766, Dec. 2021, doi: 10.1016/j.apenergy.2021.117766.

[12] H. B.-M., S. J., M. M., and M. D., “The Value of Improved Short_Term Wind Power Forecasting,” National Renewable Energy Laboratory, Technical Report NREL/TP-5D00-63175, Feb. 2015.

[13] F. Chen, J. Yan, Y. Liu, Y. Yan, and L. B. Tjernberg, “A novel meta-learning

approach for few-shot short-term wind power forecasting," Applied Energy, vol. 362, p. 122838, May 2024, doi: 10.1016/j.apenergy.2024.122838.

[14] M. Yang, H. Zhou, M. Li, B. Wang, and F. Wang, "A correction method for wind power forecast considering the dynamic process of wind turbine icing," Electric Power Systems Research, vol. 246, p. 111669, Sep. 2025, doi: 10.1016/j.epsr.2025.111669.

[15] L. Ye et al., "Combined Approach for Short-Term Wind Power Forecasting Based on Wave Division and Seq2Seq Model Using Deep Learning," IEEE Transactions on Industry Applications, vol. 58, no. 2, pp. 2586–2596, Mar. 2022, doi: 10.1109/TIA.2022.3146224.

[16] M. Yang, J. Niu, B. Wang, D. Huang, X. Su, and C. Ma, "FDCA-DSTGCN: A Wind Farm Cluster Power Day-Ahead Prediction Model Based on Frequency Domain Information Gain and Dynamic Trend Sensing," IEEE Trans. Sustain. Energy, vol. 16, no. 4, pp. 2782–2796, Oct. 2025, doi: 10.1109/TSTE.2025.3566592.

[17] R. Geirhos et al., "Shortcut learning in deep neural networks," Nat Mach Intell, vol. 2, no. 11, pp. 665–673, Nov. 2020, doi: 10.1038/s42256-020-00257-z.

[18] China National Energy Administration, "National electricity statistics for 2025." Accessed: Feb. 10, 2026. [Online]. Available: https://www.nea.gov.cn/20260129/6874f211acd0417eab7ac10c3061a7c2/c.html

[19] K. Jiang, N. Liu, K. Wang, Y. Chen, J. Wang, and Y. Liu, "Spatiotemporal assessment of renewable adequacy during diverse extreme weather events in China," Nat Commun, vol. 16, no. 1, p. 5198, Jun. 2025, doi: 10.1038/s41467-025-60264-9.

[20] M. Qu et al., "Prolonged wind droughts in a warming climate threaten global wind power security," Nat. Clim. Chang., vol. 15, no. 8, pp. 842–849, Aug. 2025, doi: 10.1038/s41558-025-02387-x.

[21] J. Zhao, F. Li, and Q. Zhang, "Impacts of renewable energy resources on the weather vulnerability of power systems," Nat Energy, vol. 9, no. 11, pp. 1407–1414, Nov. 2024, doi: 10.1038/s41560-024-01652-1.

[22] L. van der Most, K. van der Wiel, R. M. J. Benders, P. W. Gerbens-Leenes, and R. Bintanja, "Temporally compounding energy droughts in European electricity systems with hydropower," Nat Energy, vol. 9, no. 12, pp. 1474–1484, Dec. 2024, doi: 10.1038/s41560-024-01640-5.

[23] D. Kochkov et al., "Neural general circulation models for weather and climate," Nature, vol. 632, no. 8027, pp. 1060–1066, Aug. 2024, doi: 10.1038/s41586-024-07744-y.

[24] H. Shi, S. Du, Y. Yang, J. Zhang, T. Li, and Y. Zheng, "A Knowledge-Guided

Pre-Training Temporal Data Analysis Foundation Model for Urban Computing," IEEE Trans. Knowl. Data Eng., vol. 37, no. 11, pp. 6259–6271, Nov. 2025, doi: 10.1109/TKDE.2025.3607026.

[25] J. Yan, Y. Li, H. Wang, S. Han, W. Shang, and Y. Liu, "Wind Power Forecasting Based on Large Time Series Model," Engineering, Nov. 2025, doi: 10.1016/j.eng.2025.11.007.

[26] Y. Zhang et al., "Skilful nowcasting of extreme precipitation with NowcastNet," Nature, vol. 619, no. 7970, pp. 526–532, Jul. 2023, doi: 10.1038/s41586-023-06184-4.

[27] H. Wu, H. Zhou, M. Long, and J. Wang, "Interpretable weather forecasting for worldwide stations with a unified deep model," Nat Mach Intell, vol. 5, no. 6, pp. 602–611, Jun. 2023, doi: 10.1038/s42256-023-00667-9.

[28] J. Zhou et al., "SDWPF: A Dataset for Spatial Dynamic Wind Power Forecasting over a Large Turbine Array," Sci Data, vol. 11, no. 1, p. 649, Jun. 2024, doi: 10.1038/s41597-024-03427-5.

[29] Y. Liu and J. Wang, "Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting," Applied Energy, vol. 312, p. 118729, Apr. 2022, doi: 10.1016/j.apenergy.2022.118729.

[30] A. Meng et al., "A novel few-shot learning approach for wind power prediction applying secondary evolutionary generative adversarial network," Energy, vol. 261, p. 125276, Dec. 2022, doi: 10.1016/j.energy.2022.125276.

[31] Xiaofeng H., Rongwei Z., and Rong Z., "Study of classification method for the wind energy resource numerical simulation as-sessment," Acta Meteorologica Sinica, vol. 73, no. 5, pp. 975–984, Apr. 2015, doi: 10.11676/qxxb2015.061.

[32] Vaswani, A. et al. Attention is all you need. In Advances in Neural Information Processing Systems 30 (NeurIPS, 2017).

[33] C. Liu, X. Zhang, S. Mei, Q. Zhou, and H. Fan, "Series-wise attention network for wind power forecasting considering temporal lag of numerical weather prediction," Applied Energy, vol. 336, p. 120815, Apr. 2023, doi: 10.1016/j.apenergy.2023.120815.

[34] Liu, Y. et al. iTransformer: Inverted transformers are effective for time series forecasting. In International Conference on Learning Representations (ICLR, 2024).

[35] Zeng, A., Chen, M., Zhang, L. & Xu, Q. Are transformers effective for time series forecasting? In Proceedings of the AAAI Conference on Artificial Intelligence 37, 11121–11128 (2023). https://doi.org/10.1609/aaai.v37i9.26317

[36] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in

Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[37] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo, “Unified Training of Universal Time Series Forecasting Transformers,” in Proceedings of the 41st International Conference on Machine Learning, PMLR, Jul. 2024, pp. 53140–53164. Accessed: Feb. 10, 2026. [Online]. Available: https://proceedings.mlr.press/v235/woo24a.html

[38] Jin, M. et al. Time-LLM: Time series forecasting by reprogramming large language models. In International Conference on Learning Representations (ICLR, 2024).

[39] Wang, S. et al. TimeMixer: Decomposable multiscale mixing for time series forecasting. In International Conference on Learning Representations (ICLR, 2024).

[40] Beijing Climate Centre, “Wind and solar energy resources assessment service,” Apr. 2022.

[41] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, “Domain Generalization: A Survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 4, pp. 4396–4415, Apr. 2023, doi: 10.1109/TPAMI.2022.3195549.

[42] Y. Tang, K. Yang, S. Zhang, and Z. Zhang, “Wind power forecasting: A temporal domain generalization approach incorporating hybrid model and adversarial relationship-based training,” Applied Energy, vol. 355, p. 122266, Feb. 2024, doi: 10.1016/j.apenergy.2023.122266.

[43] L. Xu et al., “Resilience of renewable power systems under climate risks,” Nat Rev Electr Eng, vol. 1, no. 1, pp. 53–66, Jan. 2024, doi: 10.1038/s44287-023-00003-8.

[44] J. Li, Z. Song, X. Wang, Y. Wang, and Y. Jia, “A novel offshore wind farm typhoon wind speed prediction model based on PSO–Bi-LSTM improved by VMD,” Energy, vol. 251, p. 123848, Jul. 2022, doi: 10.1016/j.energy.2022.123848.

[45] C. Shyalika, R. Wickramarachchi, and A. P. Sheth, “A Comprehensive Survey on Rare Event Prediction,” ACM Comput. Surv., vol. 57, no. 3, pp. 1–39, Mar. 2025, doi: 10.1145/3699955.

[46] W. Song et al., “A Self-Supervised Pre-Learning Method for Low Wind Power Forecasting,” IEEE Transactions on Sustainable Energy, vol. 16, no. 3, pp. 1723–1736, Jul. 2025, doi: 10.1109/TSTE.2025.3529199.

[47] H. Baki, S. Basu, and G. Lavidas, “Modeling frontal low-level jets and associated extreme wind power ramps over the North Sea,” Wind Energ. Sci., vol. 10, no. 8, pp. 1575–1609, Aug. 2025, doi: 10.5194/wes-10-1575-2025.

[48] China Energy Investment Corporation, “China Energy Outlook (2025-2060),” Sep. 2025. Accessed: Jan. 31, 2026. [Online]. Available: https://epaper.ceic.com/pc/content/202509/12/content_17081.html

[49] A. Dinler, “Reducing balancing cost of a wind power plant by deep learning in market data: A case study for Turkey,” Applied Energy, vol. 289, p. 116728, May 2021, doi: 10.1016/j.apenergy.2021.116728.

[50] J. Yang, Z. Y. Dong, F. Wen, Q. Chen, and B. Liang, “Spot electricity market design for a power system characterized by high penetration of renewable energy generation,” Energy Conversion and Econom, vol. 2, no. 2, pp. 67–78, Jun. 2021, doi: 10.1049/enc2.12031.

[51] X. Fang, B.-M. Hodge, H. Jiang, and Y. Zhang, “Decentralized wind uncertainty management: Alternating direction method of multipliers based distributionally-robust chance constrained optimal power flow,” Applied Energy, vol. 239, pp. 938–947, Apr. 2019, doi: 10.1016/j.apenergy.2019.01.259.

[52] China National Energy Administration, “Xinjiang provincial-level power backup ancillary services successfully completed,” Dec. 2025. Accessed: Feb. 10, 2026. [Online]. Available: https://xjb.nea.gov.cn/dtyw/jgdt/202512/t20251224_293048.html

[53] L.-L. Bao, J.-S. Zhang, and C.-X. Zhang, “Spatial multi-attention conditional neural processes,” Neural Networks, vol. 173, p. 106201, May 2024, doi: 10.1016/j.neunet.2024.106201.

[54] A. Vaughan, W. Tebbutt, J. S. Hosking, and R. E. Turner, “Convolutional conditional neural processes for local climate downscaling,” Geoscientific Model Development, vol. 15, no. 1, pp. 251–268, Jan. 2022, doi: 10.5194/gmd-15-251-2022.

[55] C. Tian, T. Niu, and W. Wei, “Developing a wind power forecasting system based on deep learning with attention mechanism,” Energy, vol. 257, p. 124750, Oct. 2022, doi: 10.1016/j.energy.2022.124750.

[56] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” Neurocomputing, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

[57] Zhou, H. et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence 35, 11106–11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325

[58] Z. Ma and G. Mei, “A hybrid attention-based deep learning approach for wind power prediction,” Applied Energy, vol. 323, p. 119608, Oct. 2022, doi:

702 10.1016/j.apenergy.2022.119608.

Supplementary Files

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

- Supplementaryinformation.docx