Appendix for Day-Ahead Wind Power Ramp Events Prediction with Extreme-Value–Driven Learning and Confidence-Aware Detection

Chongyang Wan, *Student Member, IEEE*, Yuan Luo, *Senior Member, IEEE*, Weimin Wu, *Student Member, IEEE*, Nikos D. Hatziargyriou, *Life Fellow, IEEE*, Shunbo Lei, *Senior Member, IEEE*

Appendix A
Proof of Theorem 1

Lemma 1. *If the wind power prediction results and true data* $\left(\hat{X}(T_i), Y(T_i)\right), \ldots, \left(\hat{X}(T_{Q+i}), Y(T_{Q+i})\right)$ *are independent and identically distributed (i.i.d. or more generally, exchangeable) the prediction intervals obtained from conformal inference satisfy [1]:*

$$1-\alpha \leq P\left\{Y(T_{Q+i+1}) \in C(T_{Q+i+1})\right\} \leq 1-\alpha+\frac{1}{Q+1} \tag{A.1}$$

□

In essence, conformal inference operates by computing a transformation of the model's prediction errors over the holdout set $\hat{X}(T_i : T_{Q+i})$, and treating the empirical quantiles of these transformed errors as valid approximations of the quantiles for the subsequent prediction $\hat{X}(T_{Q+i+1})$. Since we assume that wind power and true data are i.i.d. and exchangeable, we can generalize Lemma 1 to Theorem 1:

Theorem 1. *If the wind power prediction results and true data* $\left(\hat{X}(T_i), Y(T_i)\right), \ldots, \left(\hat{X}(T_{Q+i}), Y(T_{Q+i})\right)$ *are i.i.d. (or more generally, exchangeable) the prediction intervals obtained from conformal inference satisfy [1]:*

$$1-\alpha \leq P\left\{Y(T_{2Q+i+1}) \in C(T_{2Q+i+1})\right\} \leq 1-\alpha+\frac{1}{Q+1} \tag{A.2}$$

□

However, it is important to note that this assumption holds only under the condition of distributional stability. When the distribution of wind power changes, the error distribution on the holdout set may no longer be representative of that on future data, thereby the distribution of the residuals on the holdout set and the next data are no longer guaranteed to be comparable. Therefore, We define the loss of the coverage rate between the desired and reached as coverage gap:

$$\text{Coverage gap} = 1-\alpha-P\left\{Y(T_{Q+i+1}) \in C(T_{Q+i+1})\right\} \tag{A.3}$$

When the wind power is independent and i.i.d. and the holdout set is sufficiently large, the coverage gap converges to 0.

We define the prediction and true data pair as $Z(t) = (\hat{X}(t), Y(t))$ and denote the full dataset as $Z = \{Z(T_i), Z(T_{i+1}), \ldots, Z(T_{Q+i})\}$. Then we swap the a-th data pair to the $Q+i$-th data pair and obtain the sequence $Z_a = \{Z(T_i), \ldots, Z(T_{i+a-1}), Z(T_{Q+i}), \ldots Z(T_{Q+i-1}), Z(T_a)\}$. The coverage gap is bounded as:

$$\text{Coverage gap} \leq \frac{\sum_{t=1}^{Q} \omega_t \cdot \mathrm{d_{TV}}(Z, Z_t)}{1+\sum_{t=1}^{Q} \omega_t}, \tag{A.4}$$

where $\mathrm{d_{TV}}$ is the variation distance between distributions. For wind power data, which exhibit strong temporal correlations, it is common for $\mathrm{d_{TV}}(Z, Z_t) \approx 1$ at many time steps. Therefore, wind power with large distributional discrepancies should be assigned smaller weights, while those with higher similarity should receive larger weights. Since the correlation between wind power data decays rapidly over time [2], an exponentially decaying function η^{n+1-t} can be used to model the temporal decay in wind power correlation. When Q is infinite, the coverage gap will converge to 0. In practical applications, although Q is finite, this weighting scheme effectively reduces the coverage gap, relaxes the assumptions of Theorem 1, and enhances its practical applicability.

Appendix B
Proof of Theorem 2

As described in this paper, the confidence of the proposed day-ahead WPREs prediction method depends jointly on the accuracy of extreme wind power value prediction and the confidence level of the confidence interval. Specifically, a higher confidence level implies that a greater proportion of extreme wind power values fall within the interval. Meanwhile, improved accuracy in predicting extreme wind power values reduces the proportion that extreme values lie outside the interval. Therefore, we will discuss the confidence of WPREs prediction by the extreme wind power values (i.e., the start and end value of a WPRE).

Based on principles from calculus and probability theory, the detection of WPREs is decomposed into the detection and comparison of wind power at individual time points within the WPRE duration. According to the literature [3], we assume that the error of day-ahead wind power prediction follows a Gaussian distribution. Since WPREs are extreme events and the temporal distance between the start and end points is hours, the temporal correlation between them is weak. We assume statistical independence between these two time points. Under this assumption, four possible cases are considered:

1) Both the start and end time points are accurately predicted: In this case, the detection algorithm is able to

TABLE I
OPTIMISED HYPERPARAMETERS

Hyperparameter	Search Range	Optimal Value
Historical time steps	1–3 days	2 days
IMF numbers	2–10	7
Encoder Layers	1–7	5
Attention Heads	1–64	32
Hidden Dimensions	1–2048	512
Activation function	Gelu	Gelu
Optimizer	Adam	Adam
Epoch	10–200	100
Batch Size	1–128	64
Learning rate	1e-6–1e-3	1e-3

identify the WPRE regardless of whether the true value lies within the confidence interval. The probability of such case is positively correlated with the wind power prediction accuracy β and decreases with increasing WPREs threshold. We model this probability as $\beta_c = 1/2 + \beta/2e^{0.2c}$, and the overall probability of this case is given by β_c^2. Specifically, for the minimum value, a negative prediction error does not impact the WPREs prediction, and under the assumption of a Gaussian distribution, this occurs with probability $1/2$. When the prediction error is positive, the probability that it falls within the acceptable range for WPREs detection is given by $\beta/(2e^{0.2c})$. In contrast, for the maximum value, a positive prediction error does not affect the detection, while a negative error must lie within the acceptable range with the same probability $\beta/(2e^{0.2c})$.

2) Only one of the two time points is accurately predicted: In this case, for time points with large prediction errors, the WPRE can only be successfully detected if the true value lies within the corresponding confidence interval. The probability that the true value falls within the confidence interval is $1 - \alpha$. Therefore, the overall probability of this case is $2(1 - \beta_c)\beta_c(1 - \alpha)$.

3) Both the start and end time points are inaccurately predicted. In this case, the WPRE can be correctly predicted only if the true wind power values at both the start and end points fall within the corresponding confidence intervals. Accordingly, the probability of this case is given by $(1 - \beta_c)^2(1 - \alpha)^2$.

4) At least one of the two time points exhibits a prediction error that falls outside the corresponding confidence interval: In this case, the detection process must incorporate an additional time point, leading to higher-order terms in the probability expression as $O(\alpha\beta^2)$.

Based on these four cases, the overall prediction confidence for day-ahead WPREs detection can be modeled as:

$$\beta_c^2 + 2(1-\beta_c)\beta_c(1-\alpha) + (1-\beta_c)^2(1-\alpha)^2 + O(\alpha\beta_c^2) = 1-\gamma \tag{B.1}$$

APPENDIX C
OPTIMAL HYPERPARAMETERS

The Optimal Hyperparameters of our model are shown in Table I.

REFERENCES

[1] A. N. Angelopoulos, S. Bates *et al.*, "Conformal prediction: A gentle introduction," *Found. Trends Mach. Learn.*, vol. 16, pp. 494–591, 2023.

[2] T. Konstantinou and N. Hatziargyriou, "Regional wind power forecasting based on bayesian feature selection," *IEEE Trans. Power Syst.*, 2024.

[3] G. Hao, Y. Li, Y. Li, L. Jiang, and Z. Zeng, "Lyapunov-based safe reinforcement learning for microgrid energy management," *IEEE Trans. Neural Netw. Learn. Syst.*, 2024.