Probabilistic forecasting of wind turbine remaining useful life using conformalised quantile regression

General comments

- The literature review could be broadened beyond conformal approaches to include state-of-the-art probabilistic methods based on Gaussian processes/Kriging, Bayesian neural networks, stochastic degradation models, and reliability-based RUL estimation. The authors should better position the proposed contribution relative to this broader uncertainty quantification literature.
- What is the significance of the methodology presented in this paper compared to Romano et al (2019), Mao et al. (2024) and Piao et al. (2025), that the authors cited in the literature review?
- Several hyperparameters are said to be chosen by trial-and-error without reporting the search range or final values; reproducibility would benefit from more explicit settings and/or code availability.
- The contribution of each pipeline component is currently difficult to isolate. In particular, the incremental benefit of the normal-behaviour residual features proposed in Section 2.1.1 is not demonstrated. Please provide an ablation study for the second case study comparing, at minimum, CAE inputs based on raw SCADA signals and inputs based on the proposed residual features, using the same training and evaluation protocol.
- How was the healthy operating period used for normal-behaviour modelling identified in the SCADA dataset? Please explain how the authors verified that this period preceded degradation onset and was not contaminated by incipient fault behaviour.
- For supervised RUL training, the model needs run-to-failure histories, or at least histories where the failure time is known. This might make the approach less ideal for practical deployment where failure data is not available. Please discuss how the framework could be trained or updated for fleets with few or no historical failures, and whether transfer learning or semi-supervised learning could be accommodated.
- Since the three quantiles are produced by parallel output heads, please report whether quantile crossing occurred and how it has been prevented or handled.
- The signed nonconformity score in Equation (8) appears consistent with the CQR formulation of Romano et al. (2019), and the conformal correction may therefore be negative when the initial quantile intervals are overly conservative. However, this should be explained more clearly, since many readers may expect nonconformity scores to be non-negative residual magnitudes.
- Equation 11. Consider using a different variable for the empirical coverage. ‘Cov’ can be misinterpreted with the coefficient of variation or covariance.
- The predicted median RUL trajectories exhibit noticeable local increases and decreases. Please discuss whether these variations represent forecast updates caused by new observations or undesirable instability induced by noise. The authors may also consider evaluating temporal smoothness or introducing a temporal-consistency constraint, while noting that strict monotonicity may not always be appropriate for sequential RUL estimates.
- The conformal validity argument requires further clarification, particularly for the second case study. Standard split-conformal guarantees rely on exchangeability between calibration and test observations. However, the SCADA samples include multiple, potentially overlapping time windows from the same turbine, which are

strongly dependent, while turbines may also differ in fault type, operating conditions, and degradation trajectory. Please define what constitutes one calibration observation, explain how within-turbine dependence is handled, and state precisely what type of coverage guarantee is claimed under this data structure.

- Please clarify whether the benchmark model was also evaluated using the residual features. Otherwise, the comparison may conflate the benefit of preprocessing with the benefit of the proposed approach.

Specific comments

- In Line 292, Page 14, “calibration set” or “validation set”?
- The conformal correction values q_c are not reported in the case studies.
- Do the losses shown in Figures 7 and 11 correspond to the total pinball loss in Equation (7)? Please clarify whether the values are averaged per sample, normalized, dimensionless, or reported in physical RUL units.
- Please revise the terminology “confidence interval” in Lines 299-300, Figures 8,12 and elsewhere throughout manuscript. The intervals reported in Table 2 are prediction intervals for RUL, not classical confidence intervals. Moreover, “the RUL prediction was repeated under different confidence intervals” is ambiguous. It would be clearer to state that the quantile-regression and conformal-calibration procedure was repeated for different nominal prediction-interval coverage, 1- α, and that empirical coverage and mean interval width were then evaluated for each case.