

Review of WES-Manuscript: “Chronologically calibrated PCA for wind-turbine SCADA health monitoring: a leakage-controlled CARE v6 benchmark“

General Comments

The manuscript represents a contribution to comparative evaluations of early fault detection approaches for wind turbines. This addresses the WES-relevant topics of “operation and maintenance”, “condition monitoring” and “reliability”. The main contributions lie in the extension of an existing benchmark through the evaluation of established methods using a chronological development/test split intended to reduce data leakage.

However, in its current form, the manuscript does not yet provide a sufficiently clear, well-justified, and reproducible comparison of the evaluated methods. The manuscript appears to be primarily a benchmark contribution rather than a methodological contribution. Such a contribution can be valuable, but the benchmark design, candidate configurations, method-selection rationale, and limitations need to be described much more clearly.

I believe that substantial revision is required before the manuscript can be properly assessed and published.

Major Comments

1. Writing quality: The manuscript requires substantial language editing. Several sentences are difficult to interpret, which makes it challenging to understand the methodology and assess the conclusions. Terms and phrases, including “views”, and the extensive usage of dash-connected expression such as “farm-turbine”, should be defined clearly and used consistently throughout the paper.
2. Description of Methods: The descriptions of the evaluated methods vary in detail and clarity. It is currently difficult to determine which combinations of algorithms, views, thresholds, preprocessing steps, and calibration procedures constitute each evaluation candidate.
3. Rationale for method selection: The rationale for focusing on PCA-based approaches requires further justification. The manuscript cites the avoidance of unnecessarily complex models as motivations. However, neither interpretability nor explainability is systematically evaluated or discussed in the results. If interpretability is a central advantage of the selected methods, the manuscript should demonstrate this advantage. The authors should also explain why the selected approaches provide an appropriate and sufficiently broad comparison set for complex wind-turbine systems
4. Scope and contribution of the benchmark: The evaluation currently appears to include mainly variants of established baseline methods. This does not necessarily weaken the manuscript, provided that the paper is clearly positioned as a carefully controlled benchmark study. However, the authors should explain more explicitly what additional knowledge the extension provides beyond previous benchmark versions. The contribution would be strengthened by either including a well-motivated additional comparators, or by clearly justifying the restricted scope of the benchmark and

discussing the resulting limitations. If the manuscript contains a genuinely new methodological component, this component should be described more clearly and distinguished from the established baseline methods.

Additionally, the table below lists some technical questions and remarks which need further attention and clarification.

Technical Comments

Page/Row	Topic	Comment
6-7/158-204	Evaluation candidates	Please include a table or flowchart showing how algorithms, views, thresholds, and calibration procedures are combined to form each candidate configuration.
6/193-199	Alternative views	The procedure used to combine the different views is difficult to understand. Please provide an explicit equation or step-by-step description. In particular, clarify whether the physics view is a 50/50 combination of residual families (a) and (b), and how the weights are selected.
6/195-198	Physics views	Please explain why one-step ridge-regression models and histogram-based gradient-boosting models were selected to construct the “physics residuals.” It would be helpful to clarify in what sense these residuals are physics-informed rather than purely data-driven. Please also report the input variables, hyperparameters, training procedure, and especially the model-selection criteria.
12/347	Physics effectiveness	The statement about the effectiveness of physics-informed approaches is too broad. Physics-informed methods include a wide range of approaches for incorporating domain knowledge into data-driven models. The reported results support conclusions only about the specific physics-informed residual construction tested in this study.

Conclusion

Addressing these points would substantially improve the clarity, reproducibility, and scientific value of the manuscript. A clearer description of the benchmark protocol, used methods and a rework of the contribution section would allow readers to assess the relevance of the findings.