Dear Referee #1,

First and foremost, thank you for reviewing our paper *"Uniform Blade Pitch Misalignment in Wind Turbines: a learning-based detection and classification approach"* (Preprint wes-2025-153). We are glad that your feedback was positive.

We have carefully revised the original manuscript to accommodate all your suggestions, taking this opportunity to make minor improvements to the text. In the amended version of the paper, the changes marked in orange refer to your comments. The point-by-point reply to your comments is reported here below.

General comments

The paper proposed a data-driven method for detecting collective turbine pitch offsets. The topic is relevant to wind energy, and the methods are well described. Some minor changes to the manuscript are advised in order to improve its impact and readability.

[Reviewer] The literature review is a bit too concise, and the authors should spend some time finding more references for the topic (at least 10). The topic of wind turbine proactive monitoring is quite vast, and it feels that many important works were not mentioned.

[Answer]
We thank the reviewer for this comment. We agree that the original literature review was too concise and did not sufficiently reflect the breadth of existing research on wind turbine proactive monitoring. The manuscript has now been revised to substantially expand the related works section. In particular, more than ten additional references have been included, covering both physics-based and data-driven approaches for condition monitoring and fault detection in wind turbines, with a specific focus on pitch system monitoring and SCADA-based methods. The revised literature review now provides a broader and more balanced overview of the state of the art and better positions the proposed contribution within the context of existing work.

[Action] Section Related Works has been modified accordingly

[Reviewer] The titles of sections 2 and 3 are similar, which may disorient the reader. "Preliminary" and "Exploratory" are also repeated. The paper would benefit from a restructuring into a more common layout including only Introduction, Methods, Results, and Discussion.

[Answer] We thank the Reviewer for this insightful comment. We agree that the similarity between the titles of Sections 2 and 3, as well as the repeated use of the terms "Preliminary" and "Exploratory", may lead to confusion.

[Action] We have restructured the manuscript following a more standard and widely adopted layout. The paper is now organized into Introduction, Methods, Results, and Discussion, with the content of the former Sections 2 and 3 redistributed accordingly. Section titles have been revised to clearly reflect their scope and avoid redundancy. We believe that this reorganization significantly improves the clarity and readability of the manuscript.

[Reviewer] A point that was skimmed overis the definition of the inflow. In particular, the

turbulent class is just briefly mentioned, but it would be important to understand wind conditions and terrain type the inflow turbulence represents. The accuracy of each of the 4 models should be reported in a more consistent way..

[Answer]
We thank the reviewer for highlighting this important aspect. We agree that the description of the inflow conditions was too concise in the previous version of the manuscript. The turbulent inflow was generated according to the IEC 61400-1 (Ed.3) Normal Turbulence Model (NTM), considering Turbulence Class C. According to the IEC standard, Class C corresponds to a reference turbulence intensity $I_{ref} = 0.12$, representing relatively low ambient turbulence conditions. The turbulence intensity varies as a function of the mean wind speed following the IEC prescription. Such conditions are typically associated with low-roughness or offshore sites. To ensure statistical robustness, multiple turbulence realizations were generated by varying the turbulence seed across simulations.

[Action]The manuscript has been revised accordingly in Section 2 to provide a clearer and more complete description of the inflow characteristics. The accuracy of each of the 4 models has been reported in a more consistent way in a concise table.

[Reviewer] All sections should have a table reporting recall, precision, F1 score, and support, and possibly also a confusion matrix (particularly when using the binary detection, like in Region II).

[Answer]
We thank the Reviewer for this helpful suggestion. Tables reporting recall, precision, F1 score, and support were already included throughout the manuscript to quantitatively assess the classification performance. Following the Reviewer's recommendation, we have now added the confusion matrix for the binary detection case in Region II

[Action] We have now added the confusion matrix for the binary detection case in Region II

[Reviewer] Also, the existing tables and figures are missing important information in their captions, so it is hard to pinpoint the data they refer to.

[Answer]
The captions of all tables and figures have been revised and expanded to include clearer and more detailed descriptions of the corresponding datasets, experimental conditions, and analysis context. This ensures that each table and figure is self-contained and that the reader can easily identify the specific data and results being presented without ambiguity.

[Action] The captions of all tables and figures have been revised.

[Reviewer] To conclude, it would be interesting to discuss the real-world challenges that implementing this method on a real turbine would inevitably bring about. Think, for instance, about how you would train a model for full-scale application (using SCADA data or still simulations?), how different turbulence conditions may affect accuracy, if industry cares more about false negatives (i.e., missed alarms) or false positives (i.e., unnecessary inspections), etc..

[Answer]
We thank the Reviewer for this insightful comment. In real-world deployment, several practical aspects should be considered. First, model training would likely rely on a hybrid strategy combining high-fidelity simulations (digital twin) and SCADA data, since uniformly distributed pitch misalignment events are rare and rarely labeled in field data. Domain shift between simulated and real turbulence, controller tuning, and sensor noise may affect generalization and would require site-specific calibration. Second, turbulence intensity and wake interactions may increase variability in the collective pitch demand, thus requiring adaptive baselines and potentially longer time windows, as already observed in the NTM scenarios. From an industrial perspective, missed detections (false negatives) are generally more critical, as prolonged operation under uniform misalignment may lead to long-term fatigue damage and efficiency losses. However, excessive false positives should also be avoided to prevent unnecessary inspections and loss of operator trust. Therefore, in practice, threshold tuning would likely prioritize high recall while maintaining an acceptable false alarm rate.

[Action]

Specific comments

[Reviewer] L91: Should "Normal Weather Prediction" be "Numerical Weather Prediction". This is what NWP stands for, generally. Please add a reference for this as well.

[Answer] We thank the Reviewer for highlighting this point. In the original manuscript, we had indeed referred to Numerical Weather Prediction, and we have now revised and replaced the corresponding phrase accordingly.

[Action]Numerical Weather Prediction has been subsituted in L91

[Reviewer] L106: Please define 1xRev and make its formatting consistent.

[Answer We thank the reviewer for this comment. The term $1\times$Rev is now explicitly defined as the once-per-revolution harmonic associated with the rotor rotational speed, and its notation has been made consistent throughout the manuscript.

[Action]1xRev has been defined and formatted in the text

[Reviewer] L121: It is not clear why the pitch regulator would still maximize power in the case of a pitch misalignment. It sounds like the pitch offset, as defined, is indeed the difference between the pitch command and the actual pitch, so it should lead to suboptimal performance in Region II.

[Answer] We thank the Reviewer for raising this important point. In our definition, the pitch misalignment is modeled as a uniform offset in the blade pitch angle, rather than as a tracking error between the commanded and measured pitch. In other words, the controller still operates according to its nominal control law and computes the collective pitch demand as usual, but the aerodynamic equilibrium is altered due to the imposed offset. In Region III, the collective pitch controller actively regulates the turbine to its rated power. Therefore, even in the presence of a uniform offset, the controller compensates by adjusting the demanded collective pitch, resulting in comparable power production but different pitch-demand statistics. In Region II, the pitch controller is not active, since torque control regulates the turbine operation while the blades are kept at their fine pitch position. In this case, a uniform offset may indeed shift the operating point away from the optimal C_p–λ condition, potentially leading to suboptimal aerodynamic performance. However, due to the relatively flat shape of the C_p curve around its maximum, small uniform offsets ($\pm 0.5°$–$\pm 2°$) do not produce large observable deviations in power, which motivates the use of alternative indicators (mechanical moments) in this region.

[Action]

[Reviewer] L130: "fro" instead of "from". L174: There is a "1" in the text. L176: Missing parenthesis for citation.

[Answer] All the indicated typos and formatting issues have been corrected.

[Action] The spelling error at L130 has been fixed, the extraneous character at L174 has been removed, and the missing parenthesis in the citation at L176 has been added.

[Reviewer] L186: It is recommended to turn the definition of the metrics into a table or formulas that use the more common definitions of False/True positives and False/True Negative (see e.g., https://www.osti.gov/biblio/1769877)

[Answer] We thank the reviewer for the suggestion. The definition of the performance metrics has been revised and reformulated using the standard notation based on True Positives, True Negatives, False Positives, and False Negatives.

[Action] The manuscript has been revised to explicitly report the definitions of the evaluation metrics in terms of true and false positives and negatives, following standard conventions.

[Reviewer] Fig. 4: These plots are a bit unclear: • Are gray dots all the cases? • Where are the blue dots on the left panel? • It is unclear when anomaly and anomaly predictions superimpose. It would be better to use a cross and an empty circle as markers or play with transparency, for example.

[Answer] The gray dots indeed represent all the analyzed cases. This has now been explicitly clarified in the figure caption. We agree that the superposition of anomaly labels and anomaly predictions could reduce readability.

[Action]The figures have been updated by using distinct marker styles and improved visual differentiation to clearly distinguish true anomalies from predicted ones.

[Reviewer] L215: "model'" is a typo.

[Answer] Corrected.

[Action] The typographical error at L215 has been fixed.

[Reviewer] L261: The concept that a 10-minute window leads to better classification performance was already introduced at L256, consider removing it

[Answer] Corrected.

[Action] The redundant statement at L261 has been removed to avoid repetition, as the concept was already introduced at L256.

We look forward to your kind reply, and in the meanwhile, we send our warmest regards.

Sincerely yours,

Sabrina Milani, on behalf of all Authors.