

Response to Reviewer 1

We thank the reviewer for their time and the evaluations of our paper. We have carefully read the comments (shown in blue font) and provide point-by-point response here (in black font).

General comments:

The manuscript titled “A decision tree-based measure-correlate-predict approach for peak wind gust estimation from a global reanalysis dataset” by S. Kartal et al. submitted for publication in Wind Energy Science adopts a new decision tree (DT)-based Measure-Correlate-Predict (MCP) approach called INTRIGUE that utilizes several meteorological variables from a public-domain reanalysis dataset to generate long-term, site-specific peak wind gust (Wp) estimation. The introduced four different models perform very well for nominal (non-extreme) conditions but the full potential for extreme conditions is yet to be seen. In particular, the application of this approach in the context of wind energy is very interesting, as wind gusts are critical for the highly fluctuating power production. The work presented here may serve as yet another approach to address this open issue.

We thank the reviewer for the summary and encouraging remarks.

I am happy to recommend the manuscript for publication in Wind Energy Science, as the subject and the approach used in the paper will be of broad international interest. I also strongly encourage publishing the code on github. My main remarks are listed below.

We thank the reviewer for his positive recommendation. We have already published the codes at:

<https://github.com/serkankartal/PeakWindGustEstimation>.

Specific comments:

Section 7.1 Self-Prediction: Although the main ideas have been introduced, the overall presentation of this section still needs improvement. In particular, please clarify which data are used as training data and which are used for prediction. In particular, reconcile this section with Section 6.1 “Strategy for Splitting of Available Data”.

We agree that the original text was a bit confusing. We have expanded on the description in the revised manuscript (see the text marked in blue).

Table 3-5, 7: Highlight the essential items in the tables, such as the higher cor-

relation of the means, or choose another more illustrative form of presentation.

Following the suggestions of both the reviewers, we have replaced the tables by bar plots in the revised manuscript.

Figure 6: In addition to or instead of the confusion matrix, present the respective receiver operating characteristic curve.

The ROC curves are useful when one has multiple thresholds in a binary classification problem. In this work, we only use one threshold value of 20 m s^{-1} to distinguish between nominal and extreme wind speeds. Hence, we decided not to add any ROC curves.

Section 8 Limitations of the INTRIGUE Approach: Add a discussion about which input features would most likely enable the INTRIGUE approach to predict wind gusts for extreme conditions.

We added the following sentence:

“We speculate that parameters derived from vertical profiles of the ERA5 reanalysis (e.g., deep-layer wind shear, storm relative helicity, integrated Scorer parameter, Sangster parameter) as input features might improve the predictions.”

Technical comments:

Line 3: Define short & long-term in the manuscript.

We replaced ‘long-term’ with ‘multi-year’.

Line 3: Define nominal & extreme conditions in the manuscript.

We have rephrased this sentence as follows:

“The INTRIGUE approach outperforms the baseline predictions for all wind gust conditions. However, the performance of this proposed approach and the baselines for extreme conditions (i.e., $W_p > 20 \text{ m s}^{-1}$) is less than satisfactory.”

Line 27: The sentence: “The focus of the current... approach.” might be better linked to the sentence: “In this paper, we propose a... reanalysis dataset.” (Line 79).

We decided not to make this change. The line 27 briefly introduces the reader to the overall focus. Then, after we introduce various technical details (e.g.,

Eqs. 1 and 2), we give a more technical description.

Line 55 - 69: This section could better follow directly after the section that ends at line 27.

Moving this paragraph after line 27 will likely disrupt the flow. We want to describe the importance of nominal (lines 48-54) and extreme (lines 55-69) wind gusts following one another. For this reason, we did not revise the text.

Table 1: Highlight most important input features for the approach.

We have printed the important variables in bold. These variables were identified via permutation feature importance analysis.

Figure 2, 7: Add a liner relationship to each of the graphs to illustrate the correlation.

Following the reviewer's suggestion, we have included linear regression fits in these plots.

Figure 5: Add Definition of Wp10m in the caption.

We have added the definition.

Figure 6: Are the Confusion Matrix actually the exact same for the XG Boost and RF models?

We thank the reviewer for catching this plotting error. We have revised the plots.

Line 355: Add the note that for the cross-predictions the different sites across the world should experience similar regional conditions.

We added 'comparable climatic conditions' in the revised manuscript.