Review comments for Boeykens et al., Version 3

The authors have effectively and thoughtfully addressed previous comments and improved the manuscript to a more impactful level. The revisions in response to comment have improved the manuscript focus, readability, and clarity of results while maintaining robust detail of the extensive data processing and methodologies. I have suggested below only few, minor revisions to further improve the manuscript, and believe further review not necessary for publication.

Main Comments:

Comment 1:

The scale and color schemes of figures (particularly Figure 1, Figure 8, and the inset of Figure 3) make it difficult to see. Suggest increasing the size of Figure 1. Figure 3 it is difficult to see the purpose of the inset and yellow box - the inset can cover a smaller area. FSC errors are discussed, but I don't see how that relates to the figure – is it depicting an area of zero snow depth? Figure 8, the point estimates are difficult to see at this color and scale.

Comment 2:

In response to Reviewer 1, Comment 6, the authors state

> "*Rather, the purpose of this manuscript was to assess different input feature configurations within an XGBoost ML setup, to see 1) if PolSAR variables improve performance compared to backscatter intensity ones, 2) how much improvement can be achieved by including either meteorological forcing data or Snowclim SD estimates (with smaller side questions as to how the latter two compare and whether S1 observations are still be influential), and 3) to determine whether training XGBoost with spatially distributed SD training data is necessary for a good representation of topographical impacts."*

This could be more explicitly incorporated into the Introduction (line 96-107) to better establish the purpose and goals of the study.

Comment 3:

Section 3.4: Overall the spatiotemporal CV structure is rigorous. However, the construction of spatial folds raised questions. Fig. 1 shows that the point locations are quite dense in some areas. How many clusters result from the 5km threshold, and how does the distribution and number of sites vary within these clusters? Since each fold contains a similar amount of data, some folds must contain much denser data with fewer clusters (e.g., the folds containing clusters of sites in NW Italy) than others (e.g., central Austria). Does this impact the CV at all?

Comment 4:

Line 456 introduces another XGBoost based model built on the point based data. Is this the "baseline" model referenced in Table 2? This should be introduced earlier, outside of the performance metric section.

Comments by line number

L75: Section 3.3.2 states $P_{s,c}$ and Ua_c are summed from September 1st to the prediction date, while SWd_7 and MD_7 are the 7 preceding day summation. Why is wind speed not at the 7-day interval?

L289: How many locations have both ascending and descending S1 orbits? Is there any difference btween the XGBoost outputs for ascending vs descending observations?

Line 381: there is reference to an increased computation cost, though this is not addressed. Suggest

including a computational demand figure or table of each configuration in the appendices.

Line 440–448: this paragraph discusses correlation between LIA and other input variables, but the cited figures (C1b and 5) do not depict LIA. I think this could be clarified somewhat.

Line 542: “with evident improvements at unseen locations during uncovered time periods, whereas no significant improvements are observed at unseen locations” – this sentence seems to contradict itself. Which time periods is the second half of the sentence referring to? Are uncovered time periods those not included in the training data?

Line 552: *“XGBoost operates at the pixel level, limiting its ability to incorporate information from neighboring pixels”* This is a difficult sentence to introduce without further explanation in the final paragraph of the conclusion. In these feature sets, the incorporation of TPI is incorporating neighboring pixel information, and there are indexing or windowing techniques that may be applied to further spatially inform the model. I suggest removing the sentence.

Editorial comments:
Line 55: “Differently” is awkward; rephrase

Line 62: remove “also”

Line 65: “was” instead of “has been”

Line 90: “a data assimilation setup, who used”: “which” instead of “who”

Line 233: XGBoost and other tree-based methods are ML algorithms, not models. Change both instances of “model” to “algorithm”

Line 293: “since nested CV allows to characterize the generalization error of an ML model correctly”: awkward; rephrase

Line 403: change “The S1 PolSAR” to “the S1 PolSAR”

Line 466–478: Here and elsewhere the phrase “as such” is misused. “Such” is a pronoun that requires an antecedent noun. In this context, “thus,” “therefore,” or “consequently” would be more appropriate, but I’m not sure the logical flow of this paragraph requires this phrasing at all; the paragraph could be revised to flow more intuitively.