Responses to referee 1:

We would like to thank the referee for the useful comments and constructive suggestions. In the following, we address the referee's comments and describe corresponding changes we have made to the manuscript. The referee's comments are listed in *italics*, followed by our response in blue. New/modified text in the manuscript is in **bold**.

The authors have addressed most of my concerns, and the effort they put into the response is appreciated. Just a couple things need to be addressed, which may not have described as well as necessary in the initial review.

Figure 4 is still troubling me. Normally when the testing errors are higher than the training errors and there is no overlap at all between training and testing errors, this suggests overfitting. Overfitting is a modeling error that introduces bias into the model because the model is too closely related to the training data set. In fact, it is so closely related that it captures the noise in the training dataset and becomes less relevant to any other data set (i.e., the testing set, or eventually to using the method to interpolate between locations with AMDAR data). If the evolution of training and testing errors are examined, the training error will always decrease with more trees, depth, etc. The testing error will also decrease, but the testing error will eventually reach a minimum and then increase again. If the number of trees, depth, and other hyperparameters suggest that this is really the optimal configuration and not overfitted, then there could issues coming from something else. In any event, please include the number of leaves that are being used.

We thank the referee for this important insight. We set the max number of leaves to the default, which does not limit the number of leaves. We acknowledge this limitation by revising the sentence at lines 266–267 of the manuscript:

"We note that the interquartile ranges of the metrics' distributions on the training and testing datasets do not overlap. This indicates a certain level of overfitting still exists, and the model may be further improved by tuning more hyperparameters other than the number of trees and the maximum tree depth. For example, the max number of leaves is at default value which is unlimited."

The last concern is again with throwing out the uncorrelated AMDAR-ERA5 data pairs. It is clear that including these will add uncertainty to the model, so the model is developed without these uncorrelated pairs. However, if the intent is to then use this model to fill the spatial gaps where there is no AMDAR data, it is not possible to throw out pairs that would be uncorrelated since there is no AMDAR data to perform the initial step of removing those uncorrelated data pairs. So, a model would be applied to a subset of data that is not accounted for in the initial construction of the model. These issues should at least be clearly stated so a reader is aware of the limitations.

When we use the trained model to make predictions, only the predictors will be needed, so the fact that "*it is not possible to throw out pairs that would be uncorrelated since there is no AMDAR data to perform the initial step of removing those uncorrelated data pairs*"

does not cause a technical problem. We do acknowledge that the model is constructed on a subset of AMDAR data, whereas when making predictions, all conditions will be included. The sentence at line 391 is revised to:

"A preprocessing step that filters out these data clusters had to be implemented to mitigate its impacts on the model training. Consequently, the model is trained and tested on a subset of AMDAR data and does not fully represent the entire AMDAR dataset."

This limitation is also acknowledged in the abstract by adding the following to line 4 of the manuscript:

"A preprocessing step was implemented to exclude AMDAR data points that were unexplainable by the predictors, mostly under stable conditions."