

## ***Interactive comment on “Using wavelet transform and dynamic time warping to identify the limitations of the CNN model as an air quality forecasting system” by Ebrahim Eslami et al.***

### **Anonymous Referee #1**

Received and published: 25 March 2020

The paper provides two case studies critiquing CNN models trained for AQF applications. The first ML model is directly an estimator, the second is used as a corrector for a CMAQ model. The authors use wavelet modal decomposition and a shape-invariant distance metric as analysis tools to find discrepancies in the model predictions and trace them back to environmental factors. This analysis is valuable and interesting in itself. Both positive and negative results are provided.

I encourage the authors to rethink the vision of this paper. What is the central thesis of the paper? Does CNNs work better as post-processing tools rather than raw predictors? Are model biases inevitable in these applications no matter the configuration?

C1

The authors diagnose important limitations of CNN models trained on their data but very few thoughts are offered for interested researches as to how to fix these issues. (except maybe in the conclusions). For example if there is a significant difference in the accuracy of the model in nighttime vs daytime, how does a single model compare to two models trained separately on subsets of data (day/night). If your analysis shows hidden correlations between the error and RH% how can you incorporate that into the input data?

When the model is not performing well, insufficient training (as suspected by authors) is only one possible cause. Another possibility may be under parametrization, such that the model is not complex enough to capture the details of special cases. I think providing error measures on the training data and comparing them with test data can illuminate the source of underperformance.

The authors state on line 46 : "Inevitably, a consequence of such enthusiasm in the field is the risk of exaggerated expectations, fueled by results focusing on the general performance of ML models compared to that of conventional statistical models" and give their previous works as examples. At the very least this assertion needs a more detailed explanation.

---

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2019-346>, 2020.

C2