

Atmos. Chem. Phys. Discuss., referee comment RC1
<https://doi.org/10.5194/acp-2021-751-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2021-751

Anonymous Referee #1

Referee comment on "Technical note: Investigating sub-city gradients of air quality: lessons learned with low-cost PM_{2.5} and AOD monitors and machine learning" by Michael Cheeseman et al., Atmos. Chem. Phys. Discuss.,
<https://doi.org/10.5194/acp-2021-751-RC1>, 2022

General comments:

This paper uses the machine learning (ML) method to investigate the spatiotemporal variability of PM_{2.5} in winter over Denver. Although this is an interesting attempt, I found the methodology of this study has not been clearly stated, so that I can't confirm the results are scientifically sound under current condition. Following are my main concerns:

1) Section 2.1.3. Question on the spatial resolution of the meteorological inputs. The meteorological inputs for the RF model are derived from the GESO-FP data with a coarse resolution of 25 km. According to fig.1, almost 2/3 of the sites are located in one grid. You mention that the data were interpolated spatially to the CEAMS and EPA sites, but no detailed information is given. What method do you use to downscale the areal data into point data? How do you check the accuracy of the interpolation results?

2) Section 2.2.3 and figure 6. Question on the validation process of the ML model. As far as I understand, you use the whole dataset to tune the RF model with k-fold cross validation method, and then use the same dataset to validate the model performance with k-fold cross validation method and bootstrapping. In my opinion, to give unbiased evaluation on the robustness of the ML model, the validation dataset should never be used in the training process. Otherwise, the accuracy of the ML model is certain to be high since the model has already learned the pattern. Please clarify if my understanding is wrong.

Line 464-466: "However, even though we do not have confidence that our CEAMS model would have predictive skill for new time periods, we do have more confidence that our interpretation of the top meteorological and geographical relationships is valid under the conditions of the CEAMS campaign."

I do not agree with this sentence. A well-developed ML model should be able to work on new datasets. This is why we test the model's ability with new datasets in the model validation process. If the model can only work well on the training dataset, it may have an overfitting problem.

3) Question on the temporal resolution of the inputs and outputs. In the 24-hour RF model case, the model inputs and outputs are not of the same temporal resolution. The output/prediction is 24-hour $PM_{2.5}$. But the meteorology inputs, separated into daytime (11am-3pm) group and nighttime (11pm-3am) group, only cover the information of these 8 hours of a day. This method is valid if you can prove that the 8-hour data is enough to represent the whole day.

Since these comments are related to the fundamental methodology of the study, I cannot recommend this study for publication before these questions are explained.

Specific comments:

1) The target of this study is to "*investigate the potential drivers of fine-scale $PM_{2.5}$ spatiotemporal variability in wintertime Denver...*" (line 93). However, you pay a lot of attention on testing the importance of including co-located AOD measurements in the RF model. What is the reason of picking this specific variable out of all factors that could contribute to the spatiotemporal variation of $PM_{2.5}$? The motivation sounds weak especially when your conclusion is that adding co-located AOD data makes very little improvements to model prediction (line 508-509 and line 514-515).

2) Line 149-150: "*In this study, the Plantower $PM_{2.5}$ data were not corrected using the time-integrated filter measurements of $PM_{2.5}$ taken by the AMODs as in Ford et al., (2019)*".

Did you compare the real-time measurement with time-integrated filter measurements? Are they in good agreement? The word "corrected" sounds that the real-time measurement is not so reliable as the filter measurements. Please rewrite it.

3) Line 372-373: "*We also found that the RF models were better at capturing temporal variability than spatial variability during the CEAMS deployment.*"

Is Figure S13 the average results of all available monitoring sites? If so, I can only see the model's ability on temporal variability but not on spatial variability. Please give more explanation on this finding.