



EGUsphere, referee comment RC2
<https://doi.org/10.5194/egusphere-2022-578-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-578

Anonymous Referee #2

Referee comment on "A deep learning approach to increase the value of satellite data for PM2.5 monitoring in China" by Bo Li et al., EGU sphere,
<https://doi.org/10.5194/egusphere-2022-578-RC2>, 2022

General comments

This work concerns the development of a deep learning model to estimate PM2.5 concentrations from satellite data also under cloudy conditions. Although the topic addressed is interesting and quite original, the manuscript is not in a good shape for being published. I share most of the comments from Anonymous Referee #1 in the lack of clarity of some sections, especially as for the methods and the many missing details. In the Introduction section, lack of knowledge of the main physicochemical processes affecting PM production and removal seem to emerge. This is not negligible, since these can be connected with the discussion of the results obtained. Indeed, also the discussion section right now seems more like a sort of celebration of the results achieved, while a clear discussion on the results obtained, including reasons (processes?) and references, is totally missing. Clear linkages between the different variables that are obtained from ground based stations (PM concentrations, in $\mu\text{g m}^{-3}$) and from satellites (AOT/AOD data) are also completely missing. Again, this is very useful for understanding the differences between the use and meaning of the variables, and why it is in any case not trivial to use satellite data for estimating PM concentrations at ground.

Specific comments

Line 30: It is not clear which attempts you are referring to. Please rephrase.

Line 31: "constrained" does not seem the most appropriate term in this context. Rephrase.

Line 32: Add "In this work," before "We introduce".

Lines 33-34: Absolutely unclear what you mean by "We use sensitivity analysis and visualization technology to open the neural black box data model": rephrase.

Lines 35-37: Details on the accuracy and errors of the method need also to be given here.

Line 40: What do you mean by "impediments to human health"? Revise.

Lines 41-42: Please explain better, quite obscure the meaning of this sentence.

Lines 44-45: Not totally true, considering that PM_{2.5} (and PM₁₀) precursors are in some cases not regulated (consider for instance ammonium deriving from ammonia) or limitations are due to the fact that precursors themselves are pollutants/toxic. Also, secondary aerosols are transported from long-range. So, this sentence contains many technical faults which need to be addressed and better clarified as the formation of secondary aerosols is key and one of the main reasons we have difficulties in addressing limit PM concentration values worldwide.

Lines 46-47: The reason of this long residence time is well known and must be added.

Lines 47-51: Not well explained, needs revision.

Line 69-70: Unclear for people not using those techniques.

Lines 75-76: It would be interesting to know the reason of such low temporal resolution.

Line 78. Rephrase "...Yu et al., 2017). In addition, errors in this method can derive from the uncertainties..."

Line 80: Change "offered" to "obtained". What do you mean by "hourly predictions of daytime PM_{2.5}"?

Lines 87-89: Statistical analyses on what? And why this result?

Line 90: Add "often impacted by" after "regions" (remove "with").

Line 93: Change "elements" to "variables". And explain better whether "geographical information" is just position or also other details such as topography, land use, or other..

Lines 93-95: This type of statements are not appropriate for the Introduction section. Better to replace it with information on the article structure.

Lines 97-99: Please add more details on the regions (e.g., better description, maps). This sentence is not that useful in the current version. Also you can introduce the different subsections that the reader will encounter in this section.

Lines 101-104: Please provide references on the data sources (e.g.. websites).

Line 108: Until now, you have always referred to AOD. Please explain what is aerosol optical thickness (AOT) and its relation with AOD.

Lines 107-110: References needed.

Line 119: Which meteorological fields? Of which variables?

Lines 121-122: Validation against what? Please provide a discussion on the validation results.

Line 122: Here and throughout the text, I don't understand the need to have a point between "Figure" and the figure number.

Lines 127-128: Not clear.

Lines 134-135: Not clear.

Lines 135-137: How did you classify the severity of pollution?

Lines 125-159: This section is very unclear. Several methods are listed, but I cannot understand how most of them were used.

Lines 161-163: So you used just 10% of data for testing: isn't this test period too short?

Lines 169-170: References to the libraries are needed.

Lines 161-170: Also this section is quite unclear, without references and details that can help the reader to repeat the process if needed.

Lines 172-204: The aim of this section is quite obscure.

Lines 209-210: Well, not really, as the methodology section lacks many details for instance on the kind of meteorological data used.

Line 211: Change "that" to "when".

Line 216: What about the R2 or R value?

Lines 225: Change "pleasant" to "good".

Lines 222-226: What about the other parameters (RMSE, MAE, ...)?

Lines 210-212: If you do not have sampling sites and satellite retrievals, how can you train the model and how can you test (I mean, which data can you use as "measured PM2.5" in Figure 1)?

Line 227: Not all aerosols survive that long in the atmosphere! And also, the residence time is essentially driven by wet deposition processes. I assume that under monsoons period aerosols do not last long in the atmosphere..

Lines 231-232: This value is not that low..

Line 232: Please avoid the use of these non scientific terms ("delightful").

Lines 242-247: Please discuss also other metrics apart from R2.

Lines 281-282: Wet removal is not connected with presence of clouds but of precipitation.

Lines 311-315: This seems a sort of repetition of the Introduction.

Lines 306-361: The Discussion section seems just like a list of advantages of the methods, rather than a true discussion of the results. For instance, I failed to understand what are the variables that finally enter the model, and why the other variables are probably not affecting PM. Also, reasons for limitations, issues are discussed with small details. References against which to compare the results are given, but references on how to interpret the results are instead not given. Finally, a conclusion section is missing.