Atmospheric Chemistry and Physics
Discussions

Atmos. Chem. Phys. Discuss., referee comment RC2
https://doi.org/10.5194/acp-2021-427-RC2, 2021

Comment on acp-2021-427

Anonymous Referee #2

Referee comment on "Input-adaptive linear mixed-effects model for estimating alveolar lung-deposited surface area (LDSA) using multipollutant datasets" by Pak Lun Fung et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-427-RC2, 2021

This research looked how to model missing lung-deposited surface area data from both street canyons and urban background environments. This work showed that more research is needed in this area to better predict these gaps in data, but provides correlations between their revised model and real-world data.

Comments to the Authors

Line 45 – in particular to respiratory systems. “System” needs an “s”

Is there a reference for the particle deposition assumptions from line 48 – 49 of 5-30 um particles?

Lines 50 – 58 – the discussion of COVID in the introduction does seem directly relevant to the study, at least in the context discussed here. It could be said that the surface area of the particles could act as transport vectors for viruses and bacteria, and therefore, the commonly monitored particle matter is number concentration and mass concentration, …” picking up on line 59.

Methods section: What are the instruments’ allowed variance/uncertainty (+/- 5%, 2%?)

Is there a reference for quantifying LDSA from derivations of particle size distribution?

The introduction currently focuses on what LDSA is, how they move through the respiratory system, how they are currently measured, and other models that have tried to do this similar modelling. Although the introduction is already quite lengthy, it does not explain why the gaps in data are so critical to understand. Line 106/107 only mentions that these instruments sometimes "lose" data and it should be accounted for, but the authors need to address *why* the data needs to be accounted for. This can then be used as a central talking point in the conclusion as to how this model is helpful to the community.

To help concentrate the introduction, the discussion of the different types of LDSA measurement techniques (Lines 82 -103) could be summarised in two or three sentences.

Line 108 - Possibly changing the end of that sentence to "… under certain circumstances, such as traffic activities."

Are there other circumstances that correlate well? Are there areas that do not correlate well that data correction cannot be used for?

Line 205 – The LDSA study that showed large accumulation mode particles should be similar to the street canyon area of the study. What does this mean for this study?

Lines 203- 209, due to these findings, are there further limitations on this work? What does the India study contribute to this work? What fraction of the particles measured are assumed to be above 400 nm at these locations? Presumably, the street canyon site would have more near traffic particles above 400 nm, whereas the urban background would be more influenced by long range transport particles. How can you discern these artifacts measured in this campaign?

Line 206 – "environment" needs an "s" to make it "environments".

Line 231 – is there any indication of what caused the outliers, or how many were deleted from the dataset?

Line 383-391 – can the correlations be quantified here (i.e. R^2 = …), for at least a few of them?

Figure 6 – Could the authors give a 1-2 sentence (further) explanation of the Taylor diagram. It is an interesting way to summarize statistical correlations. A simple solution would be to put in brackets the color of the lines that the correlate to within the text so:

Line 407 would become – "The five mostly used sub-models are shown in Figure 6 where *r* (Blue contours) is 0.85–0.87, ð□□□ð□□ ð□□□ð□□□ð□□· (Green contours) is 5.67−5.77 µm2 cm−3 407 and ð□□□ð□□□ð□□· (Black axis) is 0.75−0.79, and also shown in Table 4"

Also, the figure captions has the two panels are reversed in order (scatter plots are top panel and the Taylor diagrams are the lower)

Line 437 – Is there a way to correct for the over/underestimation for sharp peaks? How important are these peaks for contributing to the motive of the study? If these peaks are used to determine 1 hour- exposure levels, they would need to be fairly accurate, but if they are to close the gap for monthly averages, the inaccuracy is less important.

Is the amount of work needed to model the missing data worth the inaccuracy of it? If the model is over or underestimating by up to 100%, what is the contribution of this modelled data to anyone using the real-world data? This goes back to the rational behind the project and its contribution to the scientific community

Line 504 - In the street canyon scenario, IAME is less likely to accurately model instantaneous peaks, meaning that using this for determining the least polluted route to take in an urban area might not be the best application for this model, as it would not reliably be showing what is happening in real time.