We show our gratitude to Anonymous Referee #2 for his constructive comments. We have revised the manuscript accordingly. Please find our point-to-point responses below.

Response to Anonymous Referee #2's comments

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.

Response: Thank you for the good summary of the manuscript.

Comments to the Authors

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

Response: It has been revised accordingly.

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

Response: Thank you for the question. In response with a suggestion by another referee, we decided to remove the exact values of the particle size, but to keep the terms 'coarser' and 'finer' to avoid misleading. The reference for these two sentences is Gupta and Xie (2018).

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.

Response: Thank you for the suggestion. In order not to divert the focus of the article, I agree that we should remove the discussion of COVID in the introduction. The text is updated according to your suggestion.

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

Response: According to the manufacturer, the internal precision of the AQ Urban is ± 3%, but this was not tested prior the campaign. This sentence is now inserted in the method section.

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

Response: A reference by Hinds (1999) has been inserted for quantifying LDSA from particle size distribution and lung deposition curve model.

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.

Response: Thank you for the suggestion. It is very important to address why the data needs to be accounted for. The paragraph is improved by explaining that another aim of the paper is to use the statistical models as virtual sensors after they are validated. They are useful when the actual instruments are under long-term maintenance. Due to the health effects LDSA has demonstrated, it is useful for researchers to have continuous measurements for research purposes (See Ln 98-102).

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.

Response: The paragraph is now trimmed off into a few sentences (See Ln 77-88).

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

Response: We changed the wording '…under certain circumstances' to '…in traffic environments.' to avoid misunderstanding.

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

Response: Based on the paper we cited, the correlation of LDSA with BC and NOx is high (R^2>0.7) in street canyon, but not as high in urban background and detached housing. Therefore, we changed the wording '…under certain circumstances' to '…in traffic environments' to avoid misunderstanding. In my understanding, the correlation of LDSA with BC and NOx is not a factor of data correction.

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?

Response: Accumulation mode particles refer to particles sized 500nm to 2000nm. This echoes to the beginning of the sentence that we aimed to stress that the impact of larger particles (>400 nm) to alveolar LDSA could be significant but the Pegasor AQ Urban can only measures LDSA up to 400nm.

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?

Response: Thank you for raising the concern. It is rather difficult to estimate the fraction of particles >400nm. It depends on whether $PM_{2.5}$ episode is ongoing or when there are many particles from very low-quality residential burning in detached housing areas. In response to the artefacts, we also calculated LDSA by deposition curve ($LDSA_{ICRP}$) and compared it with the measured LDSA ($LDSA_{Pegasor}$, see Fig. 5). $LDSA_{Pegasor}$ correlates fairly with $LDSA_{400-800}$ (r=0.62) but very well with $LDSA_{6-800}$ (r=0.91). It indicates that although LDSA measured by Pegasor could not detect particles > 400nm, it did not have significant effects on this whole dataset. We included the India study because the study compared the LDSA concentration in India and Helsinki. Also, it showed that larger particles might have a significant impact on LDSA concentration. We improved the paragraph by reorganising the text in Ln 196-201. We added one more limitation on this work.

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

Response: Thank you for the comment. It has been revised accordingly.

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

Response: Thank you for the question. The outliers were detected by commonly used interquartile range (IQR) method. The underlying reasons can be potential measurement errors due to extreme weathers or instrument defects. Overall, there were 0.73% and 0.99% data points identified as outliers in street canyon SC and urban background UB site, respectively. This information is now updated in the manuscript (See Ln 223).

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

Response: Thank you for the comment. We agree that it is necessary to include some of the correlations in the paragraph. It has now been done.

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”

Response: Thank you for the suggestion. The sentence is now revised accordingly.

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

Response: We apologize for this mistake. The caption is now fixed.

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.

Response: To improve the accuracy of the model, I would suggest to include more input variables. One could be the number of road traffic because most of the over/underestimation took place in the morning peak hours. However, the information of traffic is not commonly available next to the measurement station. Therefore, it would sound a bit impractical.

One of the aims of the model is to utilize the model as virtual sensors. The results of the individual data points might not be that promising; however, if we use the model as virtual sensor in a long term, the overall accuracy is fairly high (SC: R^2=0.80; UB: R^2=0.77).

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

Response: Apart from filling up missing data, another important aim is to use the model as virtual sensors. The results of the individual data points might not be that promising; however, if we use the model as virtual sensor in a long term, the overall accuracy is fairly high (SC: R2=0.80; UB: R2=0.77). The underestimation of almost 100% is only for one data point. Our intention is to show some of the limitations of the model, but not to focus on just one individual data point. In order to get the focus back to the aim, we decide to remove that misleading sentence (See Ln. 432-441).

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.

Response: Thank you for the suggestion. IMAE could capture the diurnal pattern and most of the peaks. You are also right that it could over/underestimate some of the individual sharp peaks. Every model has uncertainties. The intention to deploy this model to GreenPaths is that currently GreenPaths do not take LDSA in account. Therefore, it would at least give a fairly good estimate by using IMAE proxy, not to mention the good long term accuracy. But for sure, we would love to investigate more to improve the model, for example making use of land use regression to give spatio-temporal estimation in the future. To avoid confusion, we remove this example as one of the possible applications of IMAE.

Reference:

Gupta, R., and Xie, H.: Nanoparticles in daily life: applications, toxicity and regulations, J. Environ. Pathol. Tox., 37, https://doi.org/10.1615/JEnvironPatholToxicolOncol.2018026009, 2018.

Hinds, W. C.: Aerosol technology: properties, behavior, and measurement of airborne particles, John Wiley & Sons, 1999.