Referee comment

The authors present a nice analysis and dataset of ultrafine particle concentrations using a state-of-the-art model. They have mostly taken my previous comments on board, and the paper is almost ready.

Two key reasons why I still have "minor comments" and not just "technical corrections".

We thank the reviewer for the second round of reviewing our manuscript and for providing helpful comments. We report the comments (grey, bold) along with our replies (black).

1) First, I encourage the authors to discuss spatial as well as temporal representativeness uncertainties (as per Schutgens et al) in more detail as they are so critical to the model-observation comparisons in urban regions. This valuable prior work sets the context for much of the authors' study and is only acknowledged in a cursory comment in the discussion of uncertainties. A measurement at a single location within a large model gridbox containing exactly the spatial gradients in aerosol concentrations that the authors address with their downscaling approach is unlikely to represent the average concentration in that grid cell.

We thank the reviewer for the emphasis on the representativeness. In the first round of replies, the comment was understood to be mainly focused on the comparison to the ATom flights, as we cannot address the temporal representation error here and Schutgens et al. (2017) does not target in-situ ground measurements in great detail. However, after the more detailed study of Schutgens et al. (2016a), we agree that especially the spatial representation error is a perfect setting for the introduction for the downscaling. In fact, we try to address and reduce spatial representative errors by the downscaling without explicitly mentioning it in the manuscript.

In general, we believe that we have three different types of representation errors:

1. Purely spatial representation errors only due to the extent of the grid box at stations, for which timeline measurements of 2015 (mostly hourly) were available and we collocated our simulation with the observations according to Schutgens et al. (2016b), i. e. all observations from the EBAS database.

2. Spatio-temporal representation errors for measurement stations, for which only annual averages were available or the years were differing from the evaluation year (2015).

3. (Spatio-)temporal representation errors for the aircraft measurements from ATom, for which we used daily averaged model output and the measurement years are differing from the simulation year. The spatial representation error is reduced by the fact, that we have several measurements for each grid box.

The horizontal downscaling of UFP concentrations addresses and reduces the spatial representation error for types 1 and 2, while the temporal representation error of types 2 and 3 cannot be addressed.

We introduced and applied the representation error concept from Schutgens in more detail in the revised manuscript.

2) Second, thanks to the authors' helpful responses to my comments, it is now more obvious that their approach to comparing model to measurements is not quite correct, because they do not explicitly exclude particles which were too small to be seen by the counters used to make the measurements. Usually, the people who took the measurements are forthcoming with information about their lower size cut-offs, and most similar studies I am aware of do explicitly consider them (e.g. Ketzel et al 2021 as referenced in the paper, or Spracklen et al, ACP 2010: https://acp.copernicus.org/articles/10/4775/2010/). I think the authors need to be clearer about the short-cut they took here in the paper text, but I personally would not try to insist that the authors do the comparison more correctly with the lower size cut-offs included in their analysis, because they do present an analysis of the uncertainty introduced by not doing this.

We believe that we were not clear enough in our previous replies. In fact, in contrast to the referee comment, we do explicitly exclude particles which were too small to be seen by the measuring instruments in the model-observation comparisons. In Eq. (2) and the subsequent lines we outline how we calculate UFP numbers based on D_{up} and D_{low}:

"... where D_{up} is the fixed upper bound of 100 nm and D_{low} is the variable lower bound associated with the measurement device. For the final dataset we report the total number of UFPs, and thus D_{low} is set to 0 nm and the second error function in Eq. (2) evaluates to -1."

Thus, we set D_{low} to the lower cutoff diameter of the instrument ("Cut" in Tables 1, 2, 4 and 5) in the observation-simulation comparison. The only exception is the comparison to PNC measurements from the EBAS database, as no cutoff diameters were available (Table 3 and last three measurements presented in Table 6). We stated this in the caption of Table 3: "There is no particle size cutoff value given in the datasets, and thus none is applied on the simulation."

The exact procedure was also described in Section 4.1, lines 275-277: "The daily averaged number concentrations of the model aerosol modes were integrated for the same size region (from the lower measurement cut up to the highest measurement bin with a mean diameter below 100 nm) according to Eq. (2)."

The procedure how we calculate the resulting global dataset of UFP concentrations (setting $D_{low} = 0$ as outlined in detail in the replies to the first round of comments) should not be mistaken with the procedure we perform for the evaluation of the model results using observations (setting D_{low} to the lower cutoff of the measurement device - "Cut" in the evaluation tables).

We included the information additionally in the caption of Tables 1, 4 and 5, while it was already provided in Table 2. We removed Fig. 6 of the revised manuscript as we believe that it is not of great importance for the paper and can lead to misunderstandings, i. e. in Fig. 6 we present UFP concentrations of the simulation with $D_{low} = 0$ (in contrast to the actual evaluation presented in the table) along with measurements with differing D_{low}. Additionally, we slightly reformulated the description of our evaluation procedure to make it clearer.

Minor comments:

3) L60 methodology → methodological

Fixed, thank you for pointing out.

4) L258 "increased" compared to what? You probably mean "significant"?

We rephrased this to: "We additionally find absolute interannual variation exceeding 1000 per cm^3 over polluted regions, although below 20 % in relative terms."

5) "respectively" → "or"

Adjusted, thank you for pointing out.

6) "Additional uncertainties are introduced by the missing guidelines for PSD and PNC measurements along with different measurement size ranges (Trechera et al., 2023)." —this sentence is unclear, and needs to be rephrased to specify what these "guidelines" are. Also it's not obvious why different measurement size ranges introduce uncertainty, unless the comparison is done without matching the size range between measurements and simulations: in line with my comment above, I think the authors need to specify more explicitly that (if I understand correctly) they made no attempt to do this.

The referee is right, that the formulation is unclear. It should actually be a reference on missing guidelines in the measurements, that we shortly discussed in the introduction. In fact, it does not add uncertainties to the simulation and evaluation, especially because we are using the same size range for the simulation as in the measurements for the evaluation (compare reply on comment (2)). Therefore, we removed this sentence.

7) L533 the "free ocean" is not well-defined, at least not to me, and I think representativeness uncertainties are more relevant to the urban regions the authors mainly focus on, where there are strong urban -rural contrasts and intra-city inhomogenenities. See earlier comment.

Compare reply on comment (1): We removed this part and reworked the manuscript with respect to the inclusion of the representation error concept based on the work of Schutgens et al. (2016a,b, 2017).

References

Schutgens, N. A. J., E. Gryspeerdt, N. Weigum, S. Tsyro, D. Goto, M. Schulz, and P. Stier (2016a). "Will a perfect model agree with perfect observations? The impact of spatial sampling". In: *Atmospheric Chemistry and Physics* 16.10, pp. 6335–6353. DOI: 10.5194/acp-16-6335-2016.

Schutgens, N. A. J., D. G. Partridge, and P. Stier (2016b). "The importance of temporal collocation for the evaluation of aerosol models with observations". In: *Atmospheric Chemistry and Physics* 16.2, pp. 1065–1079. DOI: 10.5194/acp-16-1065-2016.

Schutgens, N. A. J., S. Tsyro, E. Gryspeerdt, D. Goto, N. Weigum, M. Schulz, and P. Stier (2017). "On the spatio-temporal representativeness of observations". In: *Atmospheric Chemistry and Physics* 17.16, pp. 9761–9780. DOI: 10.5194/acp-17-9761-2017.