Referee comment #1:

A global simulation of ultrafine particle concentrations for the year 2015 is presented and evaluated, with a technique to infer concentrations at high spatial grid resolution from a coarser resolution global model based on a high resolution emission inventory. The primary motivation is health effects.

The chemical transport model is state-of-the-art, as is the emissions data: EDGAR data for 2015 were only recently released at the high resolution used. However, the grid resolution of the model is surprisingly low (see comment below). An impressive number of evaluation datasets are used. The model performs well in the evaluation, though many discrepancies and details are surely lost in the annual averages presented. The paper is well written.

I recommend the paper for minor revisions, though the editor will want to consider that my suggestions are on the boundary between minor and major: calculations will take time and I suggest a new figure be added.

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

1) Introduction: a few more state-of-the-art papers on global aerosol number concentration, for example by Liu and Matsui, https://doi.org/10.1029/2022GL100543, and Chen et al, https://acp.copernicus.org/articles/21/9343/2021/, should be cited and discussed.

We thank the reviewer for the additional literature. We added the references in the Introduction (along with supplemetary references, especially on vertical distributions of particle number concentrations and particle size distributions). We also discussed the results of Chen et al. (2021) in more detail in the results section, pointing out good agreement in most continental regions, while they simulated considerably lower PNCs in India compared to our simulated UFP concentrations (which are supported by observations). We did not focus on the work of Liu and Matsui (2022), as it mostly concentrates on (and evaluates the results against) vertical distributions, which is not the scope of our manuscript.

2) I think the duration of the simulation or the time period for which the data are applicable should be mentioned in the abstract or early in the introduction, and maybe again when nudging is mentioned. It would also be useful to specify the time resolution of the model output that was used in the evaluation, and the time resolution of the dataset that will be published, in the abstract or introduction, as this will help potential users of the dataset (some might want diurnal or weekly cycles, for example).

We added the duration of the simulation (which was extended to the years covering 2014-2017 with 2014 as a spin-up year) in the abstract, the introduction and also when introducing the nudging. We mentioned the time resolution of the output in the abstract ("annual averaged") and in the Data availability statement. We additionally added the information in the conclusions. The time resolution (daily averages) of the model output that was used for the evaluation was also added in the evaluation part. We do not plan to publish diurnal or weekly cycles, or

any other higher time-resolved data here, as this study focuses on long-term concentrations, and we use monthly averaged emissions (we are aware of the uncertainties introduced by that - see reply to comment 7). We plan to study seasonality and diurnal patterns in more detail in the future, by using emission time factors to distribute the emissions more realistically.

3) The grid resolution is coarser, in both vertical and horizontal, even than that of a good number of CMIP6 climate models (that were run for hundreds of years tens of times). This seems odd since only two years were simulated. Surely, despite the high complexity of the chemistry and aerosol model, simulating two years at this resolution did not take more than a few days using a modest HPC resource at DKRZ? And only one simulation is shown – there are no sensitivity studies (unlike in the study of Gordon et al 2017 the authors cite, which also had low resolution). 1 degree resolution and 60 levels is typical of CMIP6 Earth System models, which surely are not more than a factor ten cheaper per simulated year unless EMAC is very inefficient or unnecessarily complex (e.g. > 10 volatility bins, or thousands of chemical species). Something to discuss? I should point out that the authors' work to infer higher resolution output using the emissions datasets is still clearly necessary and valuable despite this comment.

Climate models used for CMIP6 are optimized to run on longer timespans, and do not include extensive and computationally demanding chemistry and aerosol schemes. Our chemical mechanism comprises more than 100 gas phase species and more than 250 reactions (Jöckel et al., 2006). Additionally, we include aerosol microphysics and explicit aerosol-gas equilibrium calculations (Pringle et al., 2010; Tsimpidi et al., 2014). A typical simulation of 2 years takes 4 days on 3 *AMD 7763 CPU* nodes with 128 cores each at the German Climate Computing Centre (DKRZ, 2021). An increase of horizontal and vertical resolution of a factor of 2 would additionally require an increase of the model timestep by roughly a factor of 2 due to the Courant-Friedrichs-Lewy condition (Courant et al., 1967). This would lead to a model runtime of approximately 2 months for the two simulated years. As we decided to increase the duration of the model run to 4 years (2014-2017) based on the comments of the referees, this would even increase to four months.

On the other hand, we do not expect a drastic increase in agreement with the model, as the resolution of $1.0° \times 1.0°$ is still too coarse to resolve urban to rural contrasts and would require further downscaling (as the reviewer also pointed out), and the vertical extent of the surface layer is identical for the higher resolution vertical resolution setups available for our model (with minor improvements in resolution for the following model levels). Sensitivity studies for different emission sectors were performed for a follow-up study on the health effects and are not the scope of this manuscript, but would additionally drastically increase the computational costs with an increased resolution.

We therefore decided not to increase the resolution at that point, but might do so in future studies. We also added this reasoning to the manuscript.

4) Table 1 and 6: for mountain sites, did the authors compare the lowest model level with observations, or calculate the level that matches the altitude of the site relative to the average surface altitude in the 180x180km grid box?

We did the latter, sampling the UFP concentrations at the level that matches the site altitude. We state this in line 234/235: "UFP_M concentrations were sampled at the vertical grid box

covering the measurement site altitude."

5) Table 6: how was the comparison with aircraft measurements done? By interpolating, or matching grid cells, with daily or monthly averaged model output, to the four days (in the case of ATom) in which any given grid cell was sampled? A discussion of representativeness uncertainties would be appropriate here (see papers by Schutgens et al, e.g. https://acp.copernicus.org/articles/17/9761/2017/, which could also nicely set the scene for the downscaling work).

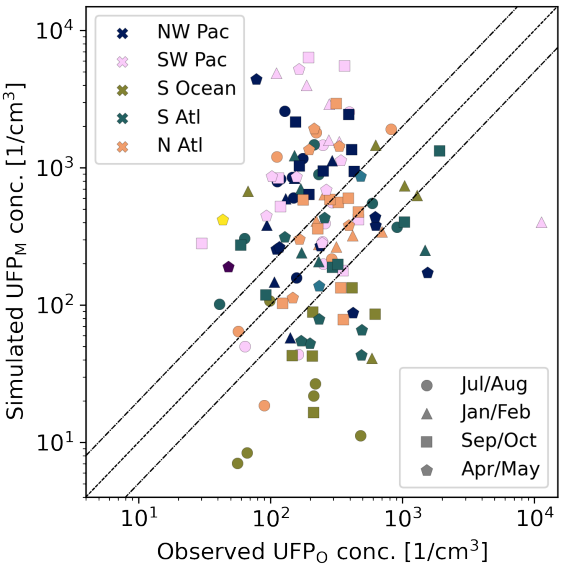

Figure 1: Summary of ATom measurements below 200m asl, compared to simulated values at the same day of the year in the relevant model grid box. Regions (Northwest pacific - NW Pac, Southwest Pacific - SW Pac, Southern Ocean - S Ocean, South Atlantic - S Atl, North Atlantic - N Atl) and seasons were distinguished by color scale and shape. Note that the timing of measurements and model calculations does not coincide (differing years), implying the meteorological variability in UFP concentrations is poorly captured.

We did the comparison by matching grid cells and the respective days when the data was sampled. We included a more detailed analysis in Fig. 1 of this document. We acknowledge two main uncertainties: Firstly, measurements and observations are from different years, and thus meteorological conditions do not coincide. Secondly, we compare daily averaged model output against the observations, which leads to representation errors according to Schutgens et al. (2017). A more extensive evaluation of our model against aircraft campaigns, including ATom, will be subject of future work, after finalising the model setup for the free and upper troposphere. We added a brief summary of the mentioned uncertainties to the "Limitations and Uncertainties" section.

6) Excluding free-troposphere stations from the evaluation on the grounds that the model wouldn't perform well there because it is optimized for the boundary layer seems odd. What optimization was done? Do the authors have reason to believe the model will perform badly in the FT, and does this also have implications for the boundary layer?

We apologise for formulating this in a misleading way. The simulation was not specifically "optimized" for the boundary layer, but rather the focus of the study is the boundary layer. We are currently working on extending the setup and also evaluating it against measurements in the free troposphere and the stratosphere. High altitude measurements at mountain peaks are strongly influenced by the steep orography, which our model cannot represent, and at the same time is not the scope of the manuscript. Potential issues in the free and upper troposphere can influence boundary layer processes as well. However, the evaluation focuses on the boundary layer, and we show good agreement there especially after the downscaling. We reformulated this description.

7) What diurnal cycles for emissions were assumed? Weekly cycles? The nonlinearities in chemistry and microphysics probably make this quite important even if only monthly averages are considered.

For this simulation we used monthly averages of emissions. We agree with the reviewer that this introduces additional uncertainties due to non-linearities, which we cannot reduce at the moment, and thank the reviewer for pointing that out. Thus, we added these uncertainties in the "Limitations and Uncertainties" section ("Anthropogenic emissions are distributed in each time step based on monthly averages from the emission datasets. This potentially introduces bias, due to the non-linearity of chemical and microphysical processes. Future studies will include emission time factors, to improve the time resolution of emissions."). We are working on an extension to the model setup to additionally include emission time factors (manuscript on that in preparation) for daily and diurnal variations, and plan to look into seasonal and diurnal cycles of UFP concentrations in more detail in the future.

8) L105 may as well state the actual number of volatility bins.

Added, alongside with the range of saturation concentration values: "ORACLE distinguishes between primary and secondary organic aerosols from different sources and volatilities (in up to 5 logarithmically spaced saturation concentration bins, ranging from 10^{-2} to 10^{6} ug/m^3, depending on the emission sector)."

9) Figure 2: The nucleation mode geometric mean diameter is between 1 and 1.5nm in this case, so half the particles in the mode are smaller than 1.5nm, and a big fraction are smaller than 1nm. This is really small! The CLOUD NPF parameterizations that are used produce particles at 1.7nm diameter. So does it make sense to produce a dataset with nucleation-mode particles smaller than that? Since the authors use CLOUD NPF parameterizations, I would suggest the authors exclude anything smaller than the CLOUD collaboration's favourite cut-off diameter of 1.7nm from their model output and their evaluation tables and think about tweaking their model for future studies.

We thank the reviewer for pointing out this issue. We agree that particles below 1.7 nm in diameter seem unreasonable to include, and being molecular clusters, as defined e. g. by the CLOUD experiment. Freshly formed particles are always added to the model at a diameter of 1.7 nm. However, adding particles to a modal distribution automatically produces particles at a lower size as part of the distribution, going down to infinitesimal small diameters. The redistribution between the different modes (more detailed outline in Pringle et al. (2010)) can then actually lead to a median diameter smaller than 1.7 nm, when larger nucleation mode particles are moved to the Aitken mode. Thus, this is only a computational artefact as a result

of the modal redistribution, while we actually do not explicitly add particles with a diameter smaller than 1.7 nm. We therefore decided to keep the definition of the UFP concentrations as it is, as otherwise (setting a lower cut to 1.7 nm) we would exclude actually nucleated particles. Additionally, we added size-resolved distributions as described in the following reply, also including the cut at 1.7 nm.

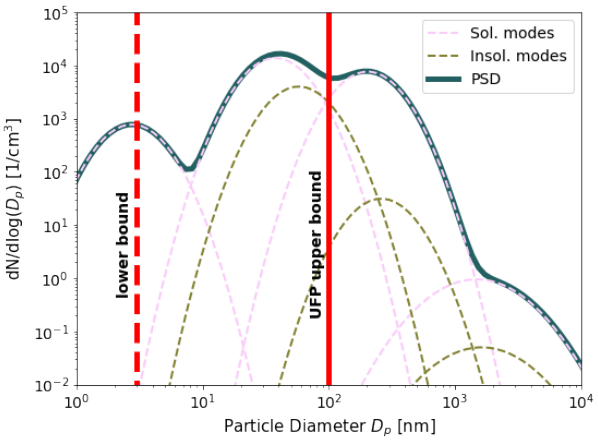

Figure 2: Typical particle size distribution (PSD) taken from the simulation in an urban region. The dashed lines represent the soluble and insoluble aerosol modes. The PSD is the sum of all these modes (blue line), typically dominated by the soluble modes. UFPs are defined as all particles with diameter below 100 nm (right red line), while the total particle number concentration is the full integral over the PSD. For comparison between observed and simulated concentrations, the lower bound (red dashed line) is considered as a cutoff, which depends on the measurement device in practice. The final dataset includes all UFPs without lower bound.

The size distribution in Fig. 2 in the manuscript was actually taken from a remote region over the ocean, at which the described computational artefact is most clearly apparent, which is not representative for the globe. We replaced the figure with Fig. 2 of this document taken from an urban region (Beijing).

We additionally compared the UFP concentrations with and without cut at 1.7 nm. Over continental regions, the differences are well below 10 %, with differences below 2 % in polluted regions, while over the ocean differences can increase to maximum values of 20 %. We added a short discussion on this to the "Limitations and Uncertainties" section.

10) If the authors don't want to change this, they really need to make the number concentration between 1.7nm and 100nm public so users could exclude particles that are, at face value, molecular clusters, from their analysis. In fact, I strongly encourage the authors to make public at least UFP concentrations in a couple of size ranges, irrespective of the lower cut-off. Say 50-100nm particles and 10-100nm particles. And monthly rather than annual averages. The paper will surely attract more citations that way.

We thank the reviewer for this suggestion to improve our manuscript. We added three size-resolved datasets to the published datasets (1.7-20 nm, 20-50 nm, 50-100 nm) and additionally discussed and showed these results in the manuscript and here in Fig. 3. We see a clear

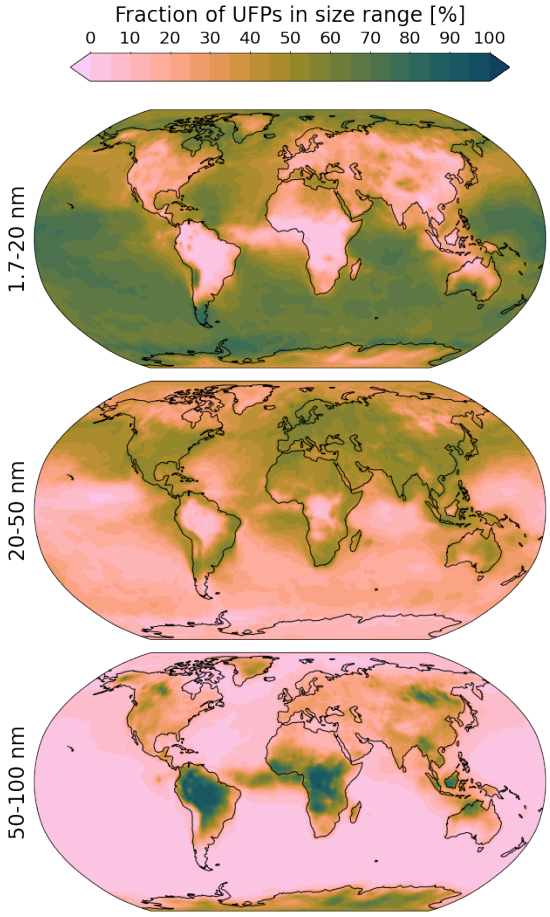

Figure 3: Fraction of UFP concentrations in different size ranges

distinction for the three size ranges. Ocean environments are mostly dominated by freshly nucleated particles between 1.7 and 20 nm, biomass burning regions by larger particles between 50 and 100 nm, and polluted regions by Aitken mode size particles between 20 and 50 nm, that are either freshly emitted fossil fuel primary particles or grown secondary particles. As outlined in the reply to comment 2, we do not plan to publish datasets with higher temporal resolution in this scope. We focused here on the evaluation of long-term (annual averaged) concentrations. A more detailed look into seasonality will follow in future publications.

11) The evaluation section could generally benefit from more time-resolved data. Seasonal cycles are lacking. I suggest adding another figure with seasonal cycles at many sites around the world, to complement Figure 5, which only shows sites in Leipzig.

We refer the referee to the replies to previous comments: Studying the seasonality of UFP concentrations and size distributions is not within the scope of this manuscript and will be subject of future studies.

12) L354 individuals (typo)

Fixed, thanks for pointing out.

References

Chen, X., F. Yu, W. Yang, Y. Sun, H. Chen, W. Du, J. Zhao, Y. Wei, L. Wei, H. Du, Z. Wang, Q. Wu, J. Li, J. An, and Z. Wang (2021). "Global–regional nested simulation of particle number concentration by combing microphysical processes with an evolving organic aerosol module". In: *Atmospheric Chemistry and Physics* 21.12, pp. 9343–9366. DOI: 10.5194/acp-21-9343-2021.

Courant, R., K. Friedrichs, and H. Lewy (1967). "On the Partial Difference Equations of Mathematical Physics". In: *IBM Journal of Research and Development* 11.2, pp. 215–234. DOI: 10.1147/rd.112.0215.

German Climate Computing Centre (DKRZ) (2021). *Introduction to Levante*.

Jöckel, P., H. Tost, A. Pozzer, C. Brühl, J. Buchholz, L. Ganzeveld, P. Hoor, A. Kerkweg, M. G. Lawrence, R. Sander, B. Steil, G. Stiller, M. Tanarhte, D. Taraborrelli, J. van Aardenne, and J. Lelieveld (2006). "The atmospheric chemistry general circulation model ECHAM5/MESSy1: consistent simulation of ozone from the surface to the mesosphere". In: *Atmospheric Chemistry and Physics* 6.12, pp. 5067–5104. DOI: 10.5194/acp-6-5067-2006.

Liu, M. and H. Matsui (2022). "Secondary Organic Aerosol Formation Regulates Cloud Condensation Nuclei in the Global Remote Troposphere". In: *Geophysical Research Letters* 49.18. e2022GL100543 2022GL100543, e2022GL100543. DOI: https://doi.org/10.1029/2022GL100543. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022GL100543.

Pringle, K. J., H. Tost, S. Metzger, B. Steil, D. Giannadaki, A. Nenes, C. Fountoukis, P. Stier, E. Vignati, and J. Lelieveld (2010). "Description and evaluation of GMXe: a new aerosol submodel for global simulations (v1)". In: *Geosci. Model Dev.* 3.2, pp. 391–412. DOI: 10.5194/gmd-3-391-2010.

Schutgens, N., 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.

Tsimpidi, A. P., V. A. Karydis, A. Pozzer, S. N. Pandis, and J. Lelieveld (2014). "ORACLE (v1.0): module to simulate the organic aerosol composition and evolution in the atmosphere". In: *Geoscientific Model Development* 7.6, pp. 3153–3172. DOI: 10.5194/gmd-7-3153-2014.