Referee comment #2:

This study presents methodology for global modelling of surface level UFP with the ECHAM/MESSy model. Calculations are performed for 2015 and the gridded annual average values are compared to measured annual values obtained from EBAS, GAW-WDCA and ACTRIS covering stations in Europe, North America, India, China and remote regions from 2015 and other years.

The coarse resolution model results are downscaled to high spatial resolution using medium and high spatial resolution emission data together with statistical parameters from the model evaluation with measurements to give estimated annual values on a high resolution grid for the Globe.

The manuscript is a highly relevant contribution to the science revolving around UFP modelling, and it is well-written and clearly structured. I have some comments that may require a bit more work for the authors, but I leave it to the Editor to decide if major revisions are needed (I would be interested in reviewing a revised manuscript, but it is not an ultimate demand).

I encourage the authors to also make the emission data set for UFP in the 7 classes available publicly, as this would surely foster many citations.

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

Concerning the publication of the emission dataset for UFP, we need to clarify that we did not develop any new emission dataset here. Rather, we used the median diameter of the size-segregated particle number emissions from Paasonen et al. (2016) to translate primary mass emissions from CEDS for the respective years into number emissions for each sector. This is just an adaptation of the emission size distribution dataset from Paasonen et al. (2016) that is required for our model input. For future use of UFP emission data, we recommend either using the size distributions provided by Paasonen et al. (2016) or the median diameters derived by us, depending on the model requirements.

Specific comments:

1) In my experience, health impact studies of long-term exposure, need two major components from the air quality data: interannual variability and rural-urban contrasts (urban increment). This present study is very thorough and highly interesting for the air pollution modelling community and the research of UFP distribution and dynamics, but promoting it for health effect studies at this stage is a little premature. Often epidemiologists conducting health impact studies have very little possibility to truly realise the issues associated with data sets of this kind. I suggest to rephrase with less emphasis on the possibility for immidiate use of these data for health studies.

We thank the reviewer for this comment. We acknowledge the uncertainties and issues that still cannot be addressed in this dataset. However, we believe that the UFP dataset is of immediate use for the public health community. To substantiate this, we extended the simulation to the years 2016 and 2017 and added additional text on the (still statistically limited) interannual

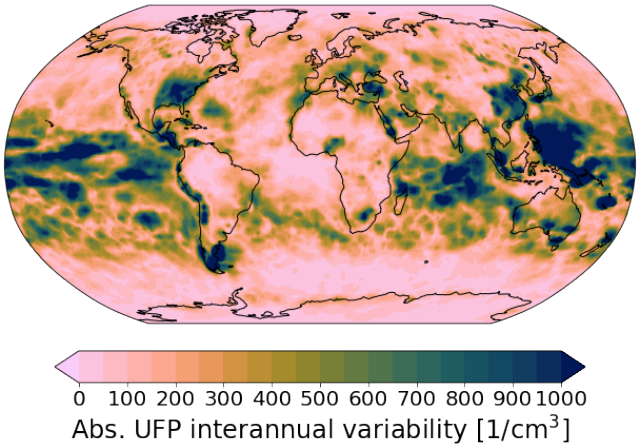

Figure 1: Absolute interannual variation of UFP concentrations calculated based on the years 2015-2017. It is defined by the maximum minus the minimum annual averaged UFP concentrations.

variability of ultrafine particle concentration. As we provide results for health applications of long-time exposure to UFP concentrations, annual averaged UFP concentrations over 3 years will be of great importance for the community.

We present the absolute interannual variability in UFP concentrations in Fig. 1. We see a strong interannual variation, exceeding 1000 per cm^3 and more than 50 % along the ITCZ, that is mostly caused by meteorological differences in the different years, as 2015 was a strong El Niño year, followed by weaker La Niña and El Niño events in 2016 and 2017. This meteorological influence on UFP concentrations could be the basis for future studies. We additionally see higher interannual variations over polluted regions, although below 20 %. This is mostly due to a decreasing trend in UFP concentrations over South and East Asia, Europe and the United States. However, we see an increasing trend over parts of Africa, especially Nigeria and South Africa. We added this figure and the respective discussion to the "Results" section in the manuscript.

Regarding the urban increment, the rural-urban contrast is implicitly represented by the "high-resolution" dataset of $0.1^\circ \times 0.1^\circ$, and explicitly in our downscaling approach, where we derive a direct relation between increasing (anthropogenic) particle emissions and increasing UFP concentrations. In the future we will work on approaches to derive UFP concentrations at even higher resolution, giving way to a more detailed analysis of the rural-urban contrast.

2) The horisontal resolution of the model is quite coarse in the simulations, and this is discussed and partly addressed in the downscaling. What about the vertical resolution/representation? A surface layer of 45-70 meters is also quite coarse for representing a measurement station at e.g. 4 meters above the surface. How is this handled? Also in connection with the down-scaling? Please share your considerations about this.

We agree with the reviewer that the vertical extent of the surface layer adds uncertainty to the dataset. While the surface layer is situated entirely in the boundary layer most of the time, there is still a profile observed in general, as reported by Du et al. (2017), Harrison et al. (2019), Heintzenberg et al. (2011), and Villa et al. (2017), with the strongest decrease

with altitude being observed close to UFP sources. We analysed the relative change in UFP concentrations from the surface layer to the next-highest model level in Fig. 2 of this document. We find a decrease of about 20-30% in UFP concentrations on average over polluted regions, with the reduction slightly exceeding 50% over Central Africa. At higher latitudes, we even find increasing UFP concentrations at the second lowest model level.

The decrease over polluted regions is in line with the observations from Du et al. (2017), Harrison et al. (2019), and Heintzenberg et al. (2011), who observed a decrease of up to 30 % at comparable altitudes. Villa et al. (2017) even observed a decrease of up to 80%, however these measurements were performed directly at a highway and at some distance to other sources. This setup is not representative for a grid box covering approx. 180 x 180 km and tends to overestimate the effect.

Considering a linear decrease respectively increase within the range of 30% within the model surface layer would add uncertainty of ±15% (as the surface emissions will be 15% higher or lower than the grid box average concentrations) to the UFP concentrations. We added a short summary of these findings in the section on "Limitations and Uncertainties" of the manuscript.

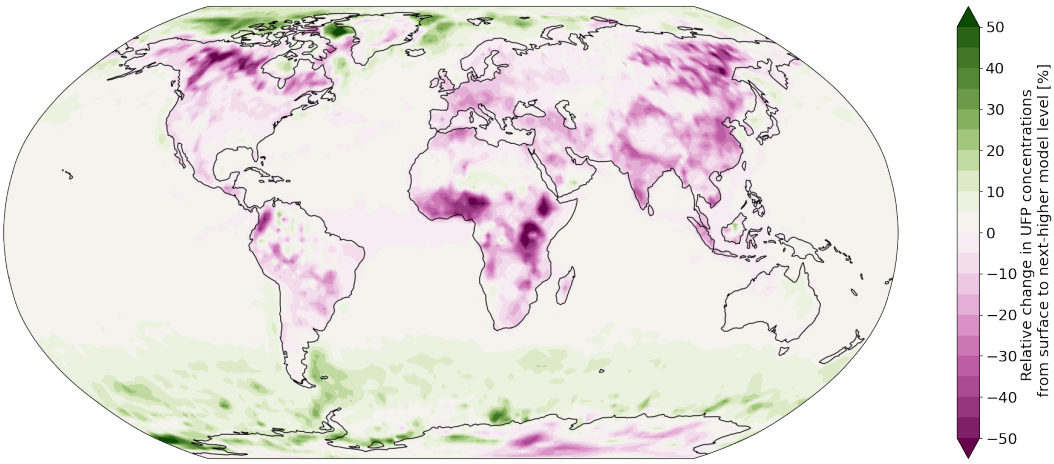

Figure 2: Relative change of simulated UFP concentrations from the surface layer to the next-highest model level in %.

3) Where does equation (1) originate from? Please either include more information on the derivation or add a reference.

We thank the referee for pointing out to a more detailed description of equation (1). There is actually a typo and the (+) should have been a (·). We adjusted the equation, and added more description and a reference to Seinfeld and Pandis, 2016:

"The number of emitted aerosols (N_{aer}) is calculated as

$$N_{aer} = \frac{6 \cdot M_{aer}}{\pi \cdot \rho_{aer} \cdot d_{med}^3} \cdot \exp\left(-4.5 \ln^2 \sigma_{ln}\right) \tag{1}$$

where the emitted aerosol mass M_{aer} is given by the respective emission dataset. ρ_{aer} is the density of the considered aerosol species and σ_{ln} is the width of the log-normal mode in the model. The fraction is just the geometrical derivation of the number of spherical particles from total

mass, given a diameter d_{med}, while the exponential function corrects for the lognormal volume distribution (compare equations (8.34) and (8.51) from Seinfeld and Pandis, 2016)."

4) In the methods description, references for the methodology for median emission diameter for SLV is missing.

There are no aerosol emissions for the sector SLV, but only gas phase emissions. Thus, no emission diameter is needed. We added this information in the description of the sectors.

5) I realise the global model is very complex and time consuming to run, but for the argumentation that it cannot be run for more than 1 year to stand more strongly, please add some numbers on complexity, e.g. number of chemical components included, how long does a simulation take on a state-of-the-art server etc.

As mentioned in comment 1, we extended the simulation up to 2017, covering 3 years of analysis. A typical simulation of 1 year takes roughly 2 days on 3 *AMD 7763 CPU* nodes with 128 cores each at the German Climate Computing Centre (DKRZ, 2021), and thus 8 days for the total simulation run (including spin-up). 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). We also added this additional information in the revised manuscript.

6) Line 266: there have been made attempts to model street-scale in individual countries/cities using other approaches, see e.g. Ketzel et al., 2021 (this could maybe also be relevant to include in the discussion)

Thank you for the suggestion. Firstly, we used the reference for a short comparison to our "Results": "Ketzel et al. (2021) simulated PNCs at street- and address-level in the same order of magnitude as our simulation up to urban background level, but with much higher peak values reaching more than $30,000$ per cm^3 in traffic hotspots, which we cannot resolve.". Secondly, we also extended the "Limitations and Uncertainties" section based on the same citation: "A next step could be to apply high-resolution dynamical downscaling of concentrations guided by comprehensive measurements of which data may become available in the future, either using machine learning methods or a combination of operational street pollution models and human exposure modelling as done by Ketzel et al. (2021).".

7) The authors discuss the issues related to comparing model results for 2015 with measurements from other years. It is an understandable approach, given the small number of measurements available, but it seems problematic when also the measurement-model relations from non-matching years are used to guide the downscaling of results for 2015 using emissions from 2015 (are the emissions from 2015?). Please either repeat the derivation of the linear fit in Figure 7 with only 2015 matches, or discuss the potential issues in depth.

We understand the concerns of the reviewer to include measurements of varying years to scale down UFP concentrations based on 2015 emissions. However, there is only one 2015 measurement from India and one from China. Combined with the request of the comment below of subdividing the observations into a training and validation dataset, this would not leave us with sufficient statistics. We would like to to emphasize that no complex model is trained using the observational data. Rather, the fit parameter c is just an estimate for the order of magnitude, by which the UFP concentrations are influenced by local emissions. This fit parameter is

optimised for the available observations.

It would be desirable to do the downscaling and evaluation based on only one year, but we would need more measurements for that. As pointed out by the reviewer, we included a discussion on the potential issues of this methodology in the limitations and uncertainties section:

"The downscaling was performed using evaluation results from all years, while the simulation results were only taken from 2015. Thus, there are potential interfering effects from the downscaling and general temporal trends as we can see from the study of interannual variability in Sect. 4."

8) Using measurements to guide downscaling and then the same measurements to validate the downscaled model results can be considered problematic. Again I understrand the approach, as not many measurements are there to use, but would it be possible to use one set of data to do carry out the downscaling, and then evaluate with another separate set (like it is done in data assimilation)?

We thank the referee for the suggestion. However, as mentioned in the previous comment, the observations are only used to estimate the order of magnitude of the influence of local anthropogenic emissions on UFP concentrations. For this purpose, we wanted to exploit the maximum number of the limited observations to get the best estimate.

As we are in favor of an independent validation as well, we added an analysis, randomly subdividing the observations into a training (25 stations) and validation (24 stations) dataset, performed the analysis 5000 times and show the improvement gained from this independent analysis. The results for the improvement in the validation dataset are shown in the revised manuscript and here (Fig. 3), supporting the claim that the parameter estimation is robust and largely independent from the chosen stations.

The fit parameters c_{CEDS} and c_{EDGAR} range between 0.07 and 0.23 in 90% of the runs, while the average values are very close to the values derived from the complete dataset, and all derived fit parameters are greater than 0. The RMSLE consistently decreases with rising order of downscaling, with an average improvement of 0.11 in total (95.6% of the time RMSLE decreases), while the logarithmic correlation increases on average by 0.07 (improves 95.1% of the time). The bias $\overline{\mathrm{M/O}}$ evolves from 0.78 on average in the model output resolution (2.4% of the time between 0.9 and 1.1) to 0.92 after the CEDS downscaling (55.3% between 0.9 and 1.1) to 1.04 (66.2% between 0.9 and 1.1) after the final downscaling, slightly overestimating the observed values. This analysis shows that the fit parameters can be transferred to stations that have not been seen by the training dataset. However, as we want to maximise the number of measurement stations to use for the global downscaling, we use the complete evaluation results for the downscaling as described before.

9) Introduction: One of the issues with UFP measurements is the lack of consolidated guidelines for methodology. Could be good to mention this issue in the introduction, and perhaps also discuss at a later stage.

We added a short discussion on this topic in the introduction and "Limitations and Uncertainties" section, based on work of Trechera et al. (2023): "Furthermore, there are no clear methodology guidelines for measuring PSDs or particle number concentrations (PNCs) (Trechera et al., 2023) with new recommendations developing recently (ACTRIS, 2021; CEN/TS 16976:2016, 2016; CEN/TS 17434:2020, 2020), and measurement size limits vary greatly."

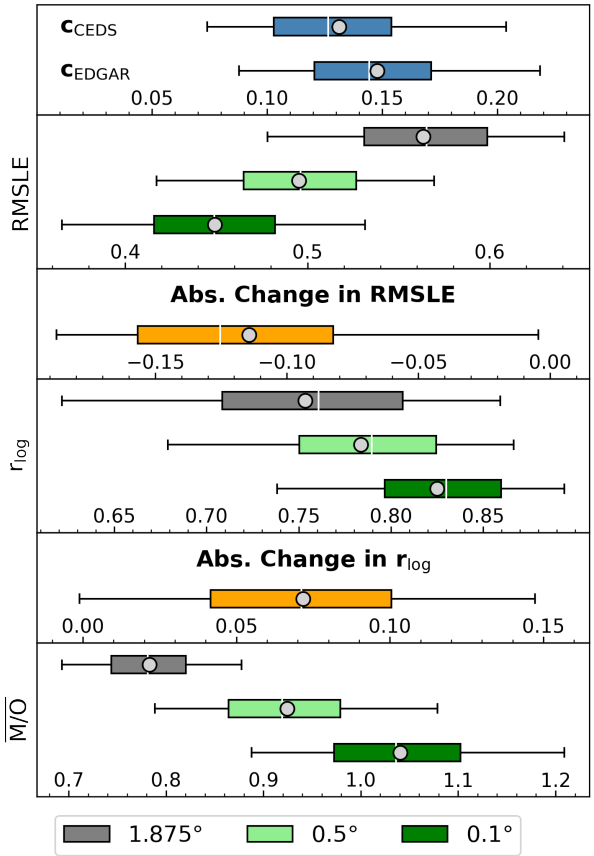

Figure 3: Results of 5000 independent training and validation runs for the fit parameter estimation (25 stations in training dataset, 24 in validation dataset). The fit parameters are displayed in the top panel, while the statistical measures are displayed below at different orders of downscaling, along with the absolute change of RMSLE and the logarithmic correlation. We show box-whisker plots, where the boxes illustrate the 25-75% interval and the whiskers the 5-95% interval. Median values are marked by the vertical white line and mean values by the light-grey circles.

10) Examples of other modelling studies for Europe that could be mentioned (to refer to multi-year modelling studies) are

- **Fountoukis, C., Riipinen, I., Denier van der Gon, H.A.C. et al, 2012. Atmos. Chem. Phys. 12, 8663–8677.**
- **Kukkonen, J., Karl, M., Keuken, M.P., et al 2016. Model Dev. (GMD) 9, 451–478.**
- **Frohn, L.M., Ketzel, M., Christensen, J.H. et al.,, 2021. Atm. Env. Vol. 264, 118631.**

We thank the reviewer for pointing out this additional literature. We refer to all these studies in the introduction, and briefly discuss differences with the results from Fountoukis et al. (2012) in the "Results" section: "Our simulated UFP concentrations are much lower than the modelled results from Fountoukis et al. (2012), that reach around 20,000 per cm^3, even for background regions in Europe in May 2008, and up to more than 100,000 per cm^3 in local hotspots in

Southeastern Europe. However, the general increasing trends from Western to Eastern Europe agree."

11) Line 328: how did you estimate surface level concentrations from a grid cell? Is this methodology also used for the other regions (than remote)?

We apologize for formulating this in a misleading way. We did not estimate surface level concentrations from the grid cell. Surface level referred to the lowest vertical model level and "the corresponding grid cell" just referred to the horizontal extent. We reformulated to: "... and compared them to UFP_M concentrations in the lowest vertical model level of the corresponding horizontal grid cell at the respective day of the year."

Technical comments:

12) Last sentence of the introduction – how does population density introduce inconsistencies? Surely it is the model representation that introduce inconsistensies?

Yes, this was improperly worded. We reformulated to: "... introduced by the inability of the coarse model resolution to capture population density and air pollution gradients"

13) What is shown in Figure 3: lowest model layer average, or surface layer (and at what height?)

We show the average of the lowest model level. The height of the surface level is varying between 45-70m depending on latitude and season, as described in lines 80-81. We added the following in the caption of Figure 3: "... and from the lowest vertical model level (surface level)."

14) Figure 5: the daily fluctuating concentrations are difficult to see in the figure, please consider replotting

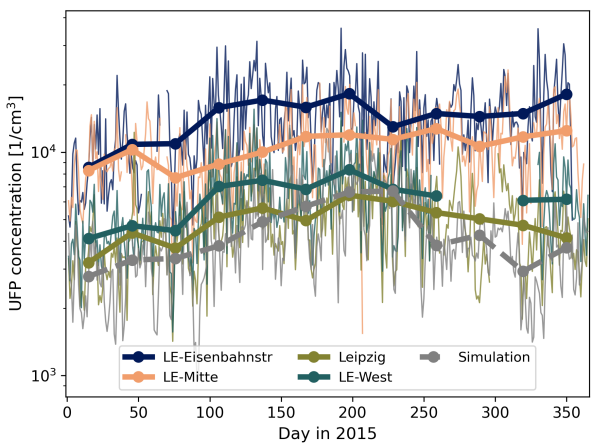

Figure 4: Monthly averaged UFP concentrations in the simulation (grey, dashed) and measured at different stations in Leipzig, Germany (thick solid lines). The daily fluctuating UFP concentrations are shown with thin, transparent lines in the same color.

We thank the reviewer for pointing this out. We made a new figure with some adjustments and a slightly modified color scheme (see Fig. 4).

15) Would be good to make the text consistent across tables, e.g. table 2 should use the formulation from table 4 considering modelling results from 2015 and measurements from other years.

Added in table 2, thank you for pointing out.

16) I am assuming that emission data are from 2015, but please specify the year of the emission data used.

Yes, the emission data was used for the respective simulated years. We added the years for the description of the nudging data, and subsequently referred to the emissions datasets for the "respective years".

17) Table 3: a typo, PNC_M and PNC_M should be PNC_O and PNC_M.

Fixed, thank you for pointing out.

18) Line 283: what is meant?

We are unsure what the referee is referring to, and assume it is the phrase "days with valid measurements". The observations contained days, where no measurements were specified, i.e. missing values were introduced. We excluded these days in the analysis.

19) Line 296: is there really only one grid cell covering Delhi? Or should it be grid cells

Yes, there is only one model grid cell covering Delhi and New Delhi, as can also be seen in Figure 8 on the downscaling method (the model grid box has a horizontal resolution of approximately 180×180 km). In the downscaled version, several grid cells represent Delhi, as apparent by the differing values of UFP_R in Table 4, and pointed out in lines 417-419.

20) Was the population data regridded to the MESSy grid? Or how did you determine the population density for each of the points in Figure 6

We used the population density dataset at a resolution of $0.1° \times 0.1°$ and evaluated it at the coordinates of the measurement station. Thus, the values of population density are taken from a dataset with finer resolution than the simulation. We did this to illustrate that there is a model underestimation at high local population density (urban regions) and overestimation at low local population density, i. e. UFP_M is dsrongly diluted by the surroundings, while the population density data is only to a lesser extent.

21) Line 335: "with a higher than model resolution" – what is meant?

We reformulated to: "sampled at the measurement stations with a resolution of $0.1° \times 0.1°$"

22) Line 352: indivuduals should be individuals

Fixed, thank you for pointing out.

23) Line 378: "referred to as PAE_{GB}"

We reformulated as PAE_{GB} was not used in the following anyways: "and the relation to the respective grid box average at coarser resolution"

24) Figure 9: the symbols for "Before DS" are tricky to see in the plot, please consider using another font or color.

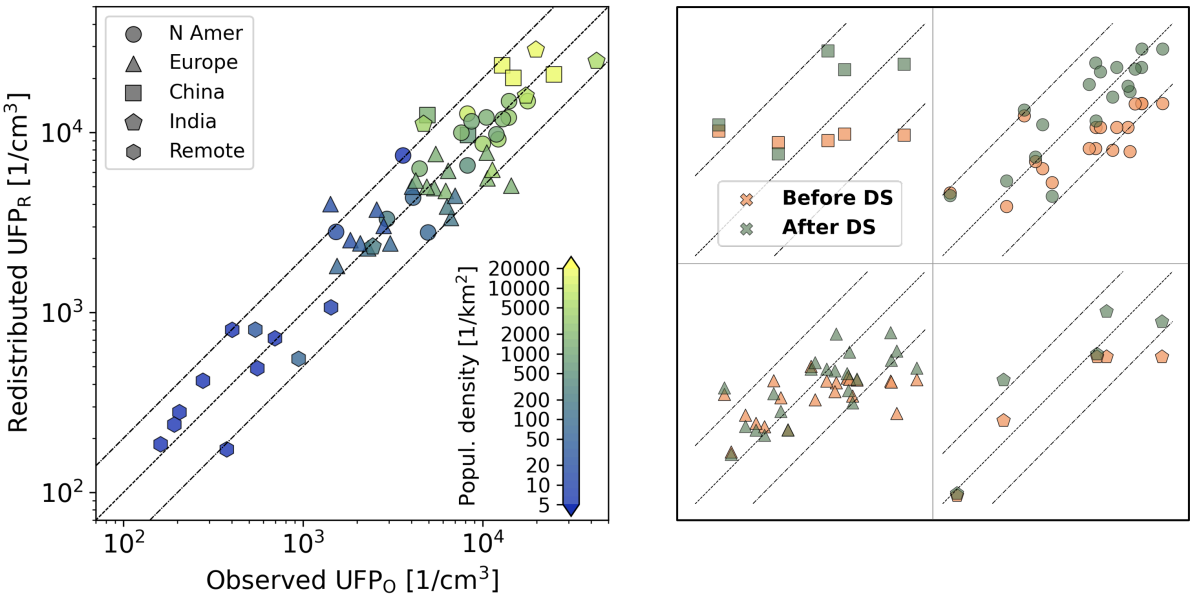

Figure 5: Improved Fig. 9

We thank the reviewer for pointing this out. We made a new figure with a slightly modified color scheme (see Fig. 5).

References

ACTRIS (2021). *ACTRIS Recommendation for mobility particle size spectrometer measurements.*

CEN/TC 264/WG 32 - Air quality - Determination of the particle number concentration (2016). *CEN/TS 16976:2016: Ambient air - Determination of the particle number concentration of atmospheric aerosol.*

CEN/TC 264/WG 32 - Air quality - Determination of the particle number concentration (2020). *CEN/TS 17434:2020: Ambient air - Determination of the particle number size distribution of atmospheric aerosol using a Mobility Particle Size Spectrometer (MPSS).*

Du, W., J. Zhao, Y. Wang, Y. Zhang, Q. Wang, W. Xu, C. Chen, T. Han, F. Zhang, Z. Li, P. Fu, J. Li, Z. Wang, and Y. Sun (2017). "Simultaneous measurements of particle number size distributions at ground level and 260 m on a meteorological tower in urban Beijing, China". In: *Atmospheric Chemistry and Physics* 17.11, pp. 6797–6811. DOI: 10.5194/acp-17-6797-2017.

Fountoukis, C., I. Riipinen, H. A. C. Denier van der Gon, P. E. Charalampidis, C. Pilinis, A. Wiedensohler, C. O'Dowd, J. P. Putaud, M. Moerman, and S. N. Pandis (2012). "Simulating ultrafine particle formation in Europe using a regional CTM: contribution of primary emissions versus secondary formation to aerosol number concentrations". In: *Atmospheric Chemistry and Physics* 12.18, pp. 8663–8677. DOI: 10.5194/acp-12-8663-2012.

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

Harrison, R. M., D. C. S. Beddows, M. S. Alam, A. Singh, J. Brean, R. Xu, S. Kotthaus, and S. Grimmond (2019). "Interpretation of particle number size distributions measured across an urban area during the FASTER campaign". In: *Atmospheric Chemistry and Physics* 19.1, pp. 39–55. DOI: 10.5194/acp-19-39-2019.

Heintzenberg, J., W. Birmili, R. Otto, M. O. Andreae, J.-C. Mayer, X. Chi, and A. Panov (2011). "Aerosol particle number size distributions and particulate light absorption at the ZOTTO tall tower (Siberia), 2006–2009". In: *Atmospheric Chemistry and Physics* 11.16, pp. 8703–8719. DOI: 10.5194/acp-11-8703-2011.

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.

Ketzel, M., L. M. Frohn, J. H. Christensen, J. Brandt, A. Massling, C. Andersen, U. Im, S. S. Jensen, J. Khan, O.-K. Nielsen, M. S. Plejdrup, A. Manders, H. D. van der Gon, P. Kumar, and O. Raaschou-Nielsen (2021). "Modelling ultrafine particle number concentrations at address resolution in Denmark from 1979 to 2018 - Part 2: Local and street scale modelling and evaluation". In: *Atmospheric Environment* 264, p. 118633. ISSN: 1352-2310. DOI: https://doi.org/10.1016/j.atmosenv.2021.118633.

Paasonen, P., K. Kupiainen, Z. Klimont, A. Visschedijk, H. A. C. Denier van der Gon, and M. Amann (2016). "Continental anthropogenic primary particle number emissions". In: *Atmospheric Chemistry and Physics* 16.11, pp. 6823–6840. DOI: 10.5194/acp-16-6823-2016.

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.

Seinfeld, J. H. and S. N. Pandis (2016). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons.

Trechera, P., M. Garcia-Marlès, X. Liu, C. Reche, N. Pérez, M. Savadkoohi, D. Beddows, I. Salma, M. Vörösmarty, A. Casans, J. A. Casquero-Vera, C. Hueglin, N. Marchand, B. Chazeau, G. Gille, P. Kalkavouras, N. Mihalopoulos, J. Ondracek, N. Zikova, J. V. Niemi, H. E. Manninen, D. C. Green, A. H. Tremper, M. Norman, S. Vratolis, K. Eleftheriadis, F. J. Gómez-Moreno, E. Alonso-Blanco, H. Gerwig, A. Wiedensohler, K. Weinhold, M. Merkel, S. Bastian, J.-E. Petit, O. Favez, S. Crumeyrolle, N. Ferlay, S. Martins Dos Santos, J.-P. Putaud, H. Timonen, J. Lampilahti, C. Asbach, C. Wolf, H. Kaminski, H. Altug, B. Hoffmann, D. Q. Rich, M. Pandolfi, R. M. Harrison, P. K. Hopke, T. Petäjä, A. Alastuey, and X. Querol (2023). "Phenomenology of ultrafine particle concentrations and size distribution across urban Europe". In: *Environment International* 172, p. 107744. ISSN: 0160-4120. DOI: https://doi.org/10.1016/j.envint.2023.107744.

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.

Villa, T., E. Jayaratne, L. Gonzalez, and L. Morawska (2017). "Determination of the vertical profile of particle number concentration adjacent to a motorway using an unmanned aerial

vehicle". In: *Environmental Pollution* 230, pp. 134–142. ISSN: 0269-7491. DOI: https://doi.org/10.1016/j.envpol.2017.06.033.