Response to comments of Anonymous Referee #1

Lines 49-51: There are also studies on the combined effects of buildings and vegetation, e.g., https://doi.org/10.1016/j.buildenv.2018.09.014 and https://doi.org/10.1016/j.ufug.2016.03.006. Please add some introduction on the impacts of street vegetation on dispersion here.

It is true that urban vegetation has an impact on the wind field too, even though a much smaller one than buildings have. We added two studies in the introduction, one which directly investigates such effects, and another one, which focuses on the sensitivity on how urban trees are represented (directly, by the roughness approach, or not at all). In our case, trees and other urban vegetation are simply represented by the surface-roughness approach, which is for sure not the most accurate one. But any more complex approach (e.g. represented by diffuse obstacles) would demand a careful evaluation, preferably using wind-tunnel data. This is beyond the scope of this study but can be addressed in a future model version.

Lines 65-67: The authors might also need to mention the potential buoyancy induced by heating between buildings, e.g., https://doi.org/10.1016/j.buildenv.2012.08.029.

The effects of buoyancy from heating of building walls and ground surfaces were described in more detail and the suggested study was gratefully added.

Line 130 and Fig. 1d-f: Some modeling details should be mentioned here.

The air-quality simulations shown at this point are described in detail in Section 2.3 'Mesoscale air-quality modeling'. What was missing at this point was a reference to Section 2.3. This is now included in the text, as well as in the caption of the figure. This avoids duplication of information.

Lines 212-218: Please elaborate on the emission datasets used in model simulations. What are the spatial and temporal resolutions of these datasets? How did the authors reconcile the resolution mismatch?

This was in fact not so well elaborated. The requested information was added in Section 2.3 for the COSMO-MUSCAT model and Section 2.4.1 for the CAIRDIO L0 simulation. The resolution mismatch between the original dataset of area pollution sources (500m x 500m) and the CAIRDIO simulation (40m) was reconciled by a special downscaling method using the building geometries, as described in Section 2.4.1. Moreover, for traffic and railway emission there is no such mismatch because of the representation as line-sources. Remaining area emissions (e.g. from farming and other mobile activities) are simply interpolated conservatively to the corresponding cell area..

Lines 225-250: The relationship between M domains and L domains is unclear. Although further details can be found in the authors' previous work, a concise summary of the model is still necessary. For example, how are the buildings "effectively represented as diffuse obstacles"?

The relationship between M domains and L domains follows clearly from the different models applied (M: mesoscale, L: local). M domains are simulated with the mesoscale model COSMO-MUSCAT. The final M domain (M3) delivers the boundary conditions for the first L domain (L0) simulated with the CAIRDIO model. The coupling is described in detail in Section 2.4.2. To respond to the Reviewer's request we added a concise summary of CAIRDIO at the beginning of Section 2.4.1, which also includes a description of "diffuse obstacles boundaries (DOB)". For more details about DOB, we refer to the companion paper Weger et al., (2021).

Line 266: I noticed that the authors used different interpolation methods for these 3D variables. Is this based on some sensitivity analysis?

It would make little difference using linear interpolation for all variables. Cubic interpolation preserves the spatial details a bit better than linear interpolation, but may overemphasize larger gradients. Thus it is not suitable for air pollution fields, which should remain positive-valued after interpolation. This explanation is now included.

Equation 2: This equation is unclear. Did the authors use this simple fitting method as a substitute for a land surface model? In addition, what are the forcings of the mesoscale model (domains Ms)?

This equation is in fact a linear regression model for the surface variables. It includes the different land-cover fractions as predictor variables. Yes, the results of this fitting method are used as a substitute for the land surface model. We definitely see this as a point of improvement, but we are convinced that this approach already meets the requirements of the study reasonably well. Now, we elaborate on the limitations of the approach a bit more, and explain the method, including the equation, better. We hope that it is now more easily understood.
To the second question, the mesoscale domains are simulated with the meteorological model COSMO (a reference to the model description was still missing in 2.3.) This is a full-fledged meteorological model with an own surface model, which is still in operational use in many countries. The coarsest domains (M0, and M1) are driven with different re-analysis data from the German meteorological office Deutscher Wetterdienst (DWD). This is all explained in Section 2.3.

Lines 354-379: It is unclear how well the mesoscale model (and the CAIRDIO model) performs during the selected period. I suggest the authors add some model evaluation results (probably in supplementary).

Although we trusted the re-analysis data, we agree that this was still a big missing part. We additionally evaluated the mesoscale simulations with a rich set of remote-sensing data (observations of horizontal wind, mixed-layer height), radiosonde and lidar data (thermal stratification), and ground-based meteorological data (for the evaluation of the urban wind field). These additional observational data are now included in a subsection of Section 2.2. Part of the model evaluation referred to the planetary boundary layer structure is now provided in Section 3.1 (and also in Figures 9 - 10). To evaluate the urban wind field, we added a supplementary section.

Lines 406-407: What are the data sources of building geometries (and land use)? Is this “30 m” here the building height averaged across the domain?

This is only roughly the average building height plus some additional margin (accounting for the roof vortex). Basically it included the layer that is most directly influenced by buildings. Note that buildings are described by a detailed building-geometry dataset with individual roof heights. The data source of the building geometries is GeoSN and for the land use data it is Pflugmacher (2018) and Banzhaf (2018). Detailed references and access information are found in Section 2.4.1 and also in the references or acknowledgments, depending on the source.

Lines 449-450: The authors attributed the underestimated spikes to observational noises. Could this be due to uncertainties in the forcing data/emission data?

This was not very clearly formulated in the text. Correct, the observed noise indicates that there are more complex sources not represented in our emissions.

Figure 13: The authors compared the results of the LES-based dispersion model and mesoscale model in this figure. If my understanding is correct, the CAIRDIO case here is the nested LES, instead of the LES with domain L0. I am curious about the performance of the LES model at L0 level, because this will be critical to demonstrating the necessity of a finer-scale dispersion model. Did the authors check the result of the online coupled LES model?

This is in fact the LES simulation on domain L0, which was compared in this figure. We added the domain labels in the caption as it was probably not clear without them. All the nested simulations (which are only offline-nested into domain L0, see Section. 2.5) are only used for the grid-spacing sensitivity study under Section 3.3.

Lines 527-530: This comparison is interesting. Is there any possible explanation in terms of the scale-dependence here?

It is not entirely clear to us what exactly your comment refers to. However, we assume that your comment refers to the slightly higher average concentrations in the 10- and 20-m runs compared to the 5- and 40-m runs. This indeed has to do with different scales. Vertical mixing in the 40m run is predominantly at the subgrid scale, while it is mostly gridscale for the 5-m run. It seems like these two runs compare very well, given the different numerical representation of mixing. However more issues are with the 10-m and 20-m runs, where subgrid-scale and grid-scale mixing is blended together. Here, it seems that the mixing length could be increased further, more towards 40 m. This was discussed in the paper, but a few lines after line 530. We changed the order in the text, so that this discussion is directly connected to the observation.

Minor comments:

Abstract: Please shorten the abstract to make it concise.

Abstract is now few lines shorter as many formulations could be written in a more compact way.

Table 1: “station Central” is not in this table.

It is, but under the label LC. “Station central” should be “Leipzig Center, LC“. This was corrected.

Line 319: QvS is undefined. Not sure if this is a typo (QvS).
This were in fact some typos. Should always be a capital “V”.

Line 604: Please explain “diffuse buildings”.

The term “diffuse buildings” is not very precise in the given context.
We replaced it with “buildings represented as physical obstacles”.