Responses to 2nd review

“Tracking the Impact of Urban Air Masses on Convective Precipitation: A Multi-Member Modeling Study”

by Keil et al.

We thank the editor and both reviewers for their continued constructive feedback. Below, we provide detailed point-by-point responses to each remaining comment. Our responses are shown in blue text, and corresponding changes or additions to the revised manuscript are presented in *gray italic*.

Reviewer #1

I want to thank the authors for their efforts to address the reviewers' comments. They have provided detailed responses to the comments that have improved my understanding of the manuscript. Many parts of the manuscript have improved in this iteration. Nevertheless, I still have a few concerns that the authors should address before publication. I list them below separated as general and specific comments. Please note that line numbers are based on the tracked changes version of the resubmitted manuscript.

We thank the reviewer for the positive assessment of the revisions and for the continued constructive feedback. We have carefully addressed all remaining concerns, as detailed below.

General comments

1. Response to general comments by both reviewers: I think the authors still need to more clearly communicate the motivation of the manuscript. Their responses have helped me better understand it, but I feel important parts of these responses are missing in the manuscript. For example, 'Whether realistic urban perturbations remain detectable above model noise in high-background aerosol environments was previously unknown and represents an important finding.', was not readily apparent to me even in the text of the revised manuscript. I encourage the authors to read their responses to the reviewers' general comments and make sure that all the points mentioned in the responses are also clearly communicated in the manuscript. In my opinion that would make the manuscript much stronger.

We agree with the reviewer and have revised the following text passages to better communicate our motivation:

Abstract: “Urban emissions impact aerosol-cloud interactions and thereby modify precipitation patterns, yet whether realistic emission perturbations from a mid-sized European city produce detectable effects above internal meteorological variability in a high-background aerosol environment remains an open question. This study investigates this by quantifying the influence of urban aerosol fields on convective precipitation through explicit chemistry-cloud coupling, using a trajectory-based ensemble approach designed to isolate weak aerosol signals from natural variability.”

Introduction: “We trace the transport pathways of polluted urban air masses from their emission sources into regions of cloud and precipitation formation by analyzing Lagrangian back trajectories, combined with an ensemble modeling framework to capture the range of possible atmospheric responses. The detection of weak urban aerosol signals above the level of internal *meteorological variability in a high-background aerosol environment is not straightforward, and it remains an open question to what extent realistic emission perturbations from a mid-sized European city are sufficient to produce discernible effects. Addressing this requires isolating these small signals from natural variability, which motivates the trajectory-based ensemble approach applied here.“*

Conclusion: “The five-member ensemble showed that internal variability is comparable in magnitude to urban emission-induced changes, yet statistically significant signals remain detectable despite the relatively small urban emission perturbation in a high-background aerosol environment.”

2. I would like to see some more discussion on why the temporal evolution of the precipitation remains basically unchanged between BASE and NONURBAN (Fig. 6c and 9c). It is quite interesting that there is a spatial redistribution of precipitation but essentially no temporal redistribution. Actually, I do not really understand this result. How can there be significant microphysical changes in the cloud and spatial changes in the precipitation amounts, but temporally precipitation rates are virtually the same? I would like the authors to discuss potential reasons for this and potentially do further analysis to explore this (as much as the scope of this study allows). For reference, in Fan et al. (2018) (Fig. 3c) aerosols significantly change the temporal evolution of precipitation rates.

We thank the reviewer for this insightful comment. The box-averaged precipitation rate represents a spatial mean over the entire analysis domain. A redistribution of precipitation within the box, where some areas receive more and others less, does not alter this domain mean and therefore does not appear as a temporal change in Figs. 6c and 9c. The urban emission signal primarily acts to redistribute precipitation spatially rather than to alter the total precipitation amount or its timing, which is consistent with the spatial difference patterns shown in Figs. 4 and 5.

For case I, this interpretation is straightforward: the difference patterns show alternating positive and negative anomalies of similar magnitude at the system edges, consistent with a spatial shift of the precipitation core rather than an overall intensification. For case II, a clearer precipitation enhancement is evident in the spatial difference plots, but this signal is spatially confined to one intensity core of the system. When averaged over the full analysis box, this localized enhancement is diluted by the surrounding areas where differences are small, such that no coherent temporal signal emerges in the domain mean.

While minor temporal shifts are present in individual ensemble members, these are not coherent across the ensemble and therefore average out in the ensemble mean. This distinguishes the present cases from scenarios with substantially larger aerosol perturbations, such as Fan et al. (2018), where aerosol loadings were sufficient to produce a coherent temporal signal in the domain-mean precipitation.

3. I would also like to see further discussion on the implications that such little aerosol differences can cause the results found here. As discussed before, the aerosol changes between the simulations are minimal compared to most other previous studies, yet some of the results are comparable. This could be a very important result, and potential reasons and implications for future studies should be discussed in more detail.

We thank the reviewer for highlighting this result. We have added the following discussion to the Discussion and Conclusion section:

"A notable aspect of this study is that the urban emission contribution to the total aerosol loading amounts to less than 3 % in both cases, as the regional background aerosol concentrations dominate. Despite this small relative perturbation, statistically significant microphysical responses and spatial redistribution of precipitation are detected, with magnitudes comparable to studies that examined substantially larger aerosol perturbations. However, domain-mean surface precipitation and total condensate remain within the ensemble spread, indicating no statistically significant net effect on total precipitation amounts. This suggests that even modest urban emission perturbations can produce detectable impacts on convective clouds when the aerosol-cloud coupling is represented explicitly and the analysis is designed to isolate the urban signal from natural variability. Nevertheless, a systematic and robust alteration of convective precipitation patterns likely requires substantially larger emission perturbations, as indicated by sensitivity experiments across a wider range of emission scaling factors in a companion study. These findings have implications for projections of urban aerosol effects under changing air quality conditions: as background aerosol concentrations decline due to emission controls, the relative contribution of urban emissions to total CCN may increase, potentially lowering the perturbation threshold required to produce systematic effects on convection."

4. 492 – 493: In my opinion, the results from Fig. 9 do not support the statement of an early decay of this system. The temporal evolution of these bulk characteristics remains virtually unchanged.

We agree with the reviewer. The statement of an earlier decay of the system is not directly supported by the bulk time series in Fig.9 and has been replaced by a more precise description of the dynamical changes, namely the suppression of updrafts by enhanced downdrafts, which is supported by the vertical velocity analysis in Fig.11.

Specific Comments

1. 36: this sentence still reads wrong, please consider revising E.g, '[...] cloud droplet size spectrum tends to consist of more numerous and smaller cloud droplets, [...]'

Changed.

2. 274-275: You can consider citing Sharma et al. (2026) (https://doi.org/10.1029/2025GL118530), who found that 10 ensembles are a good number to cover effects of internal variability.
We added the reference.

3. 338: 'diagonally'
Changed.

4. Figure 6 and 9: Please add a color legend to the figure itself.

We added color legends to both Figures.

5. Please mention the Figures S2 – S5 at appropriate locations in the manuscript.

We included references to the supplement figures in the manuscript.

Reviewer #2

Summary

In this study, the impact of urban aerosols on deep convection is investigated using the coupled atmosphere-chemistry model (COSMO-MUSCAT) with an improved urban parameterization, microphysics scheme, and a novel methodology. Two convective cases on the evening of 13th and 20th July 2019 are simulated with and without urban aerosol emissions (BASE vs NONURBAN) from Leipzig, Germany, using an ensemble of simulations. The results show that < 3% of total aerosols are from urban sources, and the rest are background. In both case studies, the differences in the impact of urban aerosols on the volume-averaged total condensate and the mean areal precipitation rate lie within the ensemble spread. However, vertical profiles shows varying impacts depending on the storm mode. For the July 13th case with moderate instability and a small convective systems, the urban aerosol effects are limited to the periphery, with increased precipitation on the northwest flank, closer to Leipzig, and decreased precipitation on the southeast flank during the mature stage. For the July 20th case with high instability and organized convection to the northeast of Leipzig, urban aerosol effects produced intense precipitation at the center of the system. In both cases, the increased aerosols from urban emissions showed signatures of increased cloud droplet activation and ice-phase processes leading to higher precipitation rates. In the July 20th case study, urban aerosols produced a higher precipitation rate, leading to a stronger downdraft and early system decay with lower total rainfall than simulations without urban aerosols.

This study is timely and relevant to understanding the effects of urban aerosol emission on deep convection. The introduction strongly motivates this. The aerosol-cloud interactions have large uncertainty not only in climate projections but also in its effects on deep convection dynamics and microphysics. The methodology and results from this study will benefit our scientific community but require a major change involving significance test and clear description and interpretation of the novel methodology. I recommend to the editor that the manuscript be published after major revisions listed below.

Things I liked about the manuscript

• This study uses a novel methodology. Past studies focused on a single convective system and apply aerosol forcing uniformly throughout the domain. Here, the authors use all convective systems (precipitation areas) that intersect with the analysis box centered on the urban plume trajectory.

• The model experiments are well designed to answer the intended research questions with an improved microphysics scheme, a coupled atmosphere-chemistry model, and urban parameterization

• The manuscript is well written and easy to follow, except for the methodology. Though the model experiments are complex, they are clearly explained. The readers unfamiliar with COSMOS-MUSCAT models can still understand it.

We thank the reviewer for the positive assessment and the constructive and detailed comments. We are pleased that the novelty of the moving box approach, the model design, and the overall clarity of the manuscript were recognized. The major comments regarding the significance test and the description of the moving box methodology have been addressed as described in detail below.

<u>Major comments</u>

1. Moving box analysis method: Use of terminologies such as "system", "early stage", and "late stage" are inappropriate

After re-reading multiple times- the methodology sections, figures 6 through 11, and the appendix, I now better understand the analysis method used in this study. I don't want other readers to go through the same.

Past studies typically analyze aerosol-cloud interactions of a single convective system or the entire domain if there are multiple systems in the domain. In this article, only columns within the analysis box centered on the plume trajectory points are analyzed. This implies that the **analysis includes only portions/chunks of multiple convective systems within the analysis box**. The choice of words throughout the manuscript created a lot of confusion. For example, singular nouns such as "the convective system" or "a precipitation system", cause confusion, because the analysis box may have multiple systems within it. Also, the use of the terminologies "early stage of a system" or "late stage of the system" is difficult to interpret. If there are multiple systems or cells inside the box, they are likely to be in different lifecycle stages- initiating, growing, dissipating, merging, or splitting.

Please provide a more detailed explanation of the moving box analysis method, describing how it differs from the traditional method and explicitly stating that only chunks or portions of convective systems within the analysis box are examined. Throughout the manuscript, please use the language that reflects this. The term "convective event or convection clusters" may be used to describe all convective systems and cells that initiated on a case study day. This is done at some places in the manuscript (e.g., lines 90,95, and other lines); please do the same throughout the manuscript. Also, instead of using "early stage or late stage of a convective system", I recommend using "initial hours of the convective event or final hours of the event".

Results are described from a system's perspective, which bothers me because there are multiple convective systems or cells in the box, and they may be in different stages. The microphysical process varies with the lifecycle stage of a convective system. So, as a reader, it begs the question "within the analysis box, is it the same part of the same system that had a high cloud droplet concentration developed ice at high altitude in the later time?"

If all the systems in the analysis box are growing or decaying at the same time, please explicitly mention that case by case; or, if a single system contributes to the major percentage of columns examined within the analysis box, then explicitly mention that and then use the system's perspective. Similarly, downdrafts produced a cold pool that affects convection in the area, not just a single system. This may help interpret the results from a convection event perspective, not just a system perspective.

An example of using singular nouns:
Line# 241 (precipitation system), and line#265 (convective system)

An example of text with system perspective and lifecycle stage:
Line# 298 (early stage of the system), and line#311 (later stage of the convective system)

These are just examples. Please look for the words "system" and "stage" through the manuscript and modify them appropriately as mentioned above.

We thank the reviewer for this detailed and helpful comment. We agree that the choice of words was misleading and have revised the manuscript accordingly.

We have expanded the description of the moving box analysis in Sect. 2.4.2 to clarify that the analysis box is centered on the mean urban aerosol plume trajectory and that the extracted profiles represent horizontal averages over all grid points within the box at each time step. This includes portions of one or more convective cells that happen to be located within the box at a given time and does not necessarily capture a single coherent convective system in its entirety. This distinction from traditional single-system analyses has been made explicit.

"For analyzing cloud dynamics and microphysics along the trajectories, a moving box analysis was implemented. At each time step, a rectangular box is centered on the mean urban trajectory point of each ensemble member, and vertical profiles are extracted by horizontally averaging all grid point values within the box, including grid points with zero values (unconditional average). Each box is defined in the COSMO rotated geographical coordinate system as a rectangular area with the mean trajectory point at its center, extending symmetrically in both horizontal directions (east–west and north–south) and including the full vertical column of model layers. The grid point values extracted within each box define the analysis volume at that time step. Although the trajectories are calculated backward in time, the moving box analysis follows the trajectory points in forward temporal order, tracing the evolution of the analysis volume from the urban emission region toward the precipitation area. As the trajectory point advances in time, the box moves accordingly through the model domain. This approach differs from conventional methods that either analyze a single convective system or average over the entire model domain: rather than tracking an individual convective system, the moving box isolates the microphysical state of convection within the aerosol-influenced region at each time step. The box may therefore contain portions of one or more convective cells at any given time, and the resulting profiles reflect the aggregate microphysical state of convection along the plume trajectory."

Throughout the manuscript, singular references to "the convective system" or "a precipitation system" have been replaced with "the convective event" where the broader context is meant. The terms "early stage" and "late stage" have been replaced with "initial hours" and "final hours" throughout.

Regarding the content of the analysis box, the two cases differ. In case I, the convective event is localized and the analysis box is dominated by a single convective cell for most of the analysis period. In case II, the event is part of a larger squall line, and the box typically contains portions of multiple convective cells at different lifecycle stages. This distinction is now stated explicitly in the results sections for each case, and the interpretation of the microphysical evolution is framed accordingly from a convective event perspective rather than a single system perspective.

2. Significance test

The study uses the Student's t-test to determine whether differences between BASE and NONURBAN simulations are significant for surface precipitation rate, condensate mixing ratio/density, and vertical velocity. However, these variables, along with many other meteorological variables, don't follow a normal distribution. They are typically modeled as gamma, exponential, or log-normal distributions. Observations sometimes show that these variables don't fit any of the above parametric distributions. The Student's t-test assumes that a variable follows a normal distribution; hence, this test cannot be applied. **I recommend using the bootstrapping method or another non-parametric significance test that doesn't assume a parametric distribution for the variable.**

The author notes that the sample size is small and that additional simulations cannot be run due to high computational demands. I agree that it's not practical to run so many simulations. With small sample size, many tests are not reliable and defeat the purpose of having the significance test. I recommend a couple of options below.

1. **Non-Parametric significance test:** use a different but non-parametric significance testing method (Mann-Whitney U test or others). Even with different testing method, the sample size is a big problem. Without sufficient samples, the significance test is not reliable and can be misleading. So use the sampling method below on any of the non-parametric significance test.

2. **Bootstrapping method with more samples:** The sample size can be increased without running more simulations. For figures 7, 8, 9, 10, and supplemental figures, I recommended using all the values on the horizontal plane at the specified elevation and time step from all ensemble members for the significance test of the averaged values. For example, when comparing the horizontally averaged cloud droplet concentration, QNC, in Fig 7b (a single x in the figure)**, I recommend using all QNC values on a horizontal plane at a specified elevation and time. Step in the analysis box from five ensemble members** (for each of BASE and NONURBAN) **for the significance test to see if horizontally averaged QNC values from two groups are significantly different.** All the horizontal data will provide the variability in QNC and help determine whether the averages from two emission scenarios are significantly different.

Similarly, for figures 4 and 5, to test the significance of the ensemble mean rain rate between experiments, please use the rain rate from neighboring pixels limited by an arbitrary distance centered on a pixel, (say 10 x 10 km2) from all ensemble members. For an example, 100 pixels centered on a pixel per member x 5 ensemble member = 500 pixels of rain rate values for each experiment.

As authors are aware, the significance tests are used to determine if the observed difference is due to internal variability or not. Rather than using only mean values from five ensemble members, actual values from them can be used to increase the sample size and represent the internal variability. **I recommend using bootstrapping method with more samples to test the significance of difference between experiments.**

Please clarify in the text (section 2.4.2) how averages are computed. Is it a conditional or unconditional average. A conditional average is based on conditions such as 1) grid cells with non-zero values, 2) above a threshold value, or 3) cloudy grid cells. Unconditional average uses all the grid cells and include zero values when computing mean.

We thank the reviewer for this detailed and constructive suggestion. We agree that the paired t-test is not ideal given the small ensemble size and the non-normal distribution of the tested variables. Following the reviewer's recommendation, we have replaced the paired t-test with a bootstrap significance test throughout.

For the vertical profile figures (Figs. 7, 8, 10, 11), the bootstrap test uses all grid point values within the analysis box at each height level and time step from all five ensemble members, for each experiment separately. For the precipitation difference maps (Figs. 4 and 5), grid point values within a 0.1° x 0.1° neighborhood centered on each pixel are pooled across all five ensemble members per experiment. In both cases, significance is assessed at the 90 %

confidence level: the null hypothesis of no difference is rejected when the bootstrap confidence interval of 1000 resampled mean differences excludes zero. Spatial averages are unconditional, i.e., they include all grid cells within the analysis box, including those with zero values.

To demonstrate that the two methods yield consistent results, we show comparison figures below with both t-test (x) and bootstrap (o) significance markers. The revised manuscript uses bootstrap significance throughout.

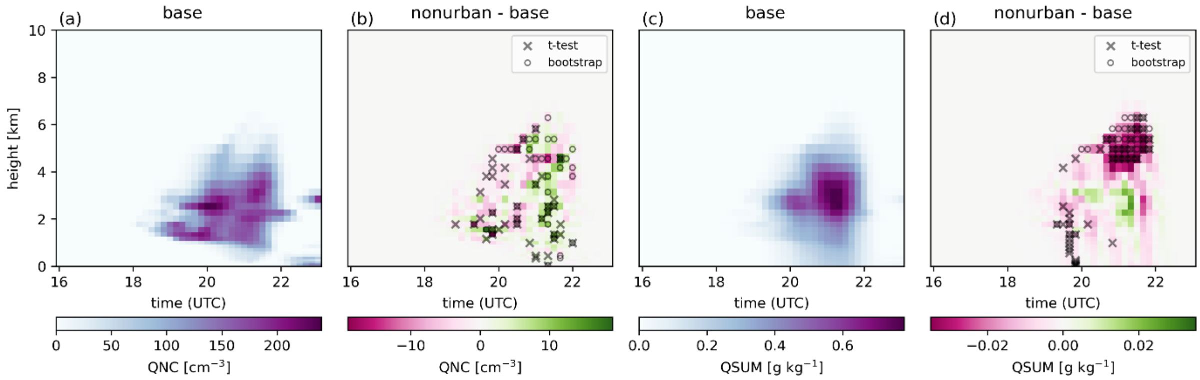

Figure 1 Horizontally averaged vertical profiles of number of activated cloud droplets QNC (a) and total cloud condensate QSUM (c), showing ensemble means for the BASE experiment for case I. (b) and (d) show the corresponding differences between the NONURBAN and BASE experiments for QNC and QSUM, respectively. Black crosses and circles indicate levels where the differences are statistically significant at the 90~% confidence level based on a t-test and bootstrap resampling, respectively.

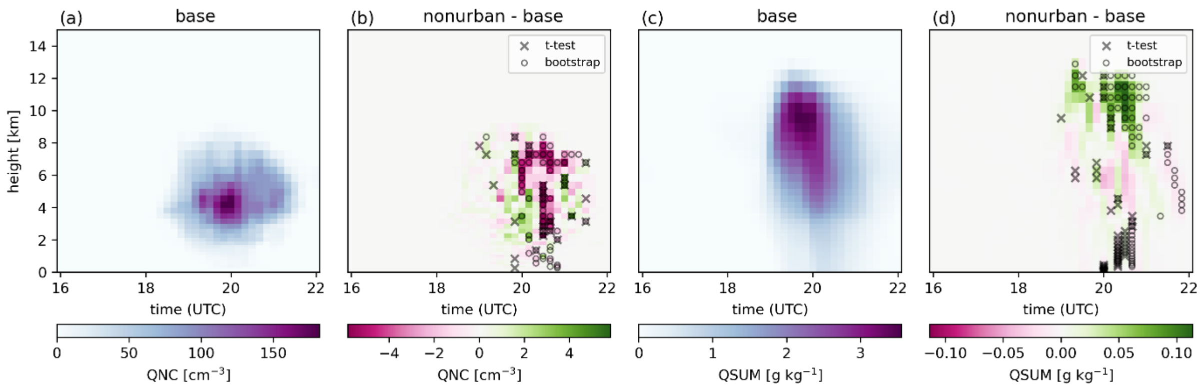

Figure 2 Horizontally averaged vertical profiles of number of activated cloud droplets QNC (a) and total cloud condensate QSUM (c), showing ensemble means for the BASE experiment for case II. (b) and (d) show the corresponding differences between the NONURBAN and BASE experiments for QNC and QSUM, respectively. Black crosses and circles indicate levels where the differences are statistically significant at the 90~% confidence level based on a t-test and bootstrap resampling, respectively.

The description in Section 2.4.3 has been updated accordingly:

"We are aware that five members are likely insufficient to capture the full range of variability, Sharma et al. (2026) found that an ensemble of ten members is required to sufficiently reduce internal variability in aerosol-convection interaction studies. Increasing the ensemble size beyond five members was, however, not computationally feasible within this project. To partially compensate for the limited ensemble size, a bootstrap significance test was applied. Rather than comparing only the five ensemble-mean values per experiment, all grid point values within the moving analysis box (Section 2.4.2) at a given height level and time step were pooled across all five ensemble members. This pooling approach increases the effective sample size without

requiring additional simulations. For each combination of height level and time step, 1000 bootstrap resamples were drawn with replacement from each pooled distribution, and the difference of the resample means was computed. A difference was considered statistically significant at the 90 % confidence level if the resulting bootstrap confidence interval did not include zero. The same approach was applied to the precipitation fields, where grid point values within a 0.1° x 0.1° neighborhood centered on each pixel were pooled across all ensemble members prior to bootstrapping."

The averages shown in the figures are unconditional, i.e., all grid cells within the analysis box are included regardless of whether values are zero or non-zero. This has been clarified in Section 2.4.2:

"For analyzing cloud dynamics and microphysics along the trajectories, a moving box analysis was implemented. At each time step, a rectangular box is centered on the mean urban trajectory point of each ensemble member, and vertical profiles are extracted by horizontally averaging all grid point values within the box, including grid points with zero values (unconditional average)."

Minor edits

Lines 36 to 41 - These lines lack clarity; please rewrite. They convey that both scenarios lead to invigoration from additional condensation and latent heat release. Both have more condensation and latent heat release compared to which aerosol scenario?

Both pathways described in these lines lead to invigoration through enhanced condensation and latent heat release, but via distinct microphysical mechanisms: suppression of warm-rain formation in the case of large particles, and activation of ultrafine particles following rapid warm-rain formation in the UFP-dominated case. The reference baseline in each case differs accordingly. We have revised lines 36-41 to clarify the reference conditions for each pathway.

"In deep convective systems, warm-rain suppression is an integral component of aerosol invigoration mechanisms. Condensational invigoration occurs through enhanced condensation and latent heat release relative to clean-air conditions (Cotton and Walko, 2021). When large particles are abundant, they activate at cloud base and suppress warm-rain formation, prolonging the condensation phase and releasing additional latent heat compared to a scenario with fewer large particles. In contrast, when ultrafine aerosol particles (UFPs), a significant component of urban emissions (Kumar et al., 2014), are abundant, the low concentration of large particles allows rapid warm-rain formation, generating high supersaturation that activates the UFPs and producing substantial additional condensation and latent heat release compared to conditions without UFPs (Fan et al., 2018)."

Line 50 - Replace "warm-based storm" with "storms with warm cloud base"

Changed.

Line 67 - Please provide specific aerosol effects on deep convection rather than a general statement that aerosols significantly affected the convection and precipitation patterns.

We added the following sentence:

"Kawecki et al. (2016) studied a mesoscale convective event passing Kansas City, Missouri, using scaled urban emission scenarios. They found that increased aerosol loading redistributed

precipitation, with higher accumulation in the downwind direction and reduced accumulation upwind, driven by aerosol-induced modifications to cloud microphysics and cold pool strength."

Line 75 – replace "The study" with "It also"; because it creates confusion whether it's referring to the current study or Van Den Heever and Cotton (2007)

Changed.

Line 90- replace "highly resolved" with "high resolution"

We thank the reviewer for this suggestion. However, we deliberately chose the phrasing "highly resolved" as a relative term, acknowledging that what constitutes high resolution depends on the perspective of the reader. From an urban modeling standpoint, the resolution employed here would not typically be considered high, whereas from a climate modeling perspective it may well be.

Line 112 – replace "Aerosol dynamics" with "Aerosol processes"

We thank the reviewer for this suggestion. However, we intentionally used "aerosol dynamics" at this location, as the sentence specifically refers to the representation of aerosol population evolution through the hybrid bulk–bin scheme within MUSCAT. Formation and transformation processes such as secondary organic aerosol formation and multiphase chemistry are described separately in the same paragraph, making the distinction between dynamics and processes explicit in context.

Line 156 – replace "maximum" with "final" or "resulting"

The term "maximum supersaturation" is retained here, as it follows the terminology of the original parameterization by Abdul-Razzak and Ghan, where it refers specifically to the maximum supersaturation reached during cloud droplet activation. Replacing it with "final" or "resulting" would deviate from the established definition without adding clarity.

Line 158 – replace "in-cloud conditions" with "in-cloud activations"

Changed.

Line 166 – replace "westerly inflow" with "westerly flow"

Changed.

Line 168 – please rewrite "3-step one-way nesting strategy. The outer domain covers" as "3-step one-way nested domains – D0, D1, D2. The outer domain, D0, covers" for clarity

We changed the sentence to:

"To simulate both cases, we employ a 3-step one-way nesting strategy (D0, D1, D2). The outer domain, D0, covers Europe..."

Lines 205 & 206 – these lines about the creation of the ensemble based on spin-up time are not clear. Please explain more. On line 184, the manuscript states that the simulation for domain D1 is started 24 hours before each case study day using input from COSMO-D2 reanalysis. Are the ensemble member simulations started at 24h, 21h, 18h, 15h, 12h before 00 hours of the case day, with corresponding initial conditions from COSMO-D2; hence leading to different spin-up times of 24h, 21h, 18h, 15h, 12h?

We thank the reviewer for this comment. The ensemble construction is based on a single D1 simulation, not on multiple independent COSMO-D2 initializations. All ensemble members are initialized from this single D1 run, with each member extracting boundary and initial conditions at its respective start time (24h, 21h, 18h, 15h, and 12h before 00 UTC of the case day). This ensures that large-scale forcing remains consistent across the ensemble while the differing start times introduce variation in spin-up duration, which serves as the source of ensemble spread. We have revised lines 205–206 accordingly to clarify this.

"For each experiment, we created an ensemble with five members, respectively, by varying the length of the meteorological spin-up run, while the 24 h coupled COSMO-MUSCAT simulations remained unchanged. The spin-up lengths are 24 h, 21 h, 18 h, 15 h, 12 h. All ensemble members are initialized from the same D1 simulation, with each member extracting boundary conditions at its respective start time, ensuring that the large-scale forcing remains consistent across the entire ensemble. This approach allows to assess the impact of slightly varying initial meteorological conditions on the results. By comparing multiple ensemble members, variability is captured more effectively, and consistent, non-stochastic patterns can be identified more clearly by statistical analysis. This increases the robustness of the findings in this study and helps to separate systematic signals from noise, providing greater confidence in the conclusions drawn from our simulations."

Line 219 – replace "chapter" with "the Section"

Changed.

Line 223: rewrite "region of highest precipitation" as "region centered on pixel with highest precipitation". **Also, it is not clear what "initial time" means.** Please rewrite

The manuscript has been revised accordingly. As the starting region was defined as a spatial box rather than a single pixel, we use "region centered on the area of highest precipitation" instead of the suggested "region centered on pixel with highest precipitation". The start times are specified in the following sentences and have therefore been removed from this sentence to avoid redundancy.

Changed to: "*We calculated backward trajectories starting from a region centered on the area of highest precipitation downwind of Leipzig for each ensemble member, using meteorological fields from the D2 output.*"

Line 225 – rewrite "The horizontal start box cover" as "The starting region of trajectories horizontally covers"

Changed.

Line 228 – is it km2?

Corrected.

Figures 3, 4, and 5 - Please increase the font size. They are difficult to read. Also, please have two or three latitude and longitude grid lines to give some idea of the scale

Changed.

Figure 3 – Caption missing information about the grey regions in the plot. Please add at the end of the caption - "City boundaries of Leipzig marked with a grey and of Chemnitz with light grey outlines."

Added.

Line 236: replace “meteorological conditions” with “cloud dynamics and microphysics”; I don’t see any meteorological variables (T, Qv, P, U) analyzed in the manuscript.

Changed.

Line 237 – delete “atmospheric”

Deleted.

Line 238, 245 – replace “trajectory point” with “mean trajectory point”

Changed.

Line 249 to 251 – the term “air mass” is loosely used throughout the text and in the title. I would suggest using “air parcels” instead of “air mass” if referring to grid cells or single trajectory. Please use the term “urban air mass” or “urban aerosol plume” if referring to the entire plume.

We agree that "air mass" is imprecise in this context, as the term carries a specific meteorological meaning that does not apply here. However, we did not adopt "air parcel", as this term is typically associated with Lagrangian frameworks tracking individual infinitesimal fluid elements, whereas our analysis is based on an Eulerian model and considers finite three-dimensional volumes of air. We therefore replaced "air mass" throughout the manuscript, including the title, with "air volume", which better reflects the nature of the sampled region. Where referring to the collective urban emission signal, we use "urban aerosol plume" as suggested.

Titel changed to: *“Tracking Air Volumes for Assessing the Effect of Urban Aerosols on Convective Precipitation: A Multi-Member Modeling Study”*

Line 249 to 250 – please rewrite as “While the analysis box along the mean filtered trajectories show the effects of urban aerosols on cloud dynamics and microphysics, the analysis box along the mean of all trajectories (shown in appendix) represent total convective cloud properties from grid cells irrespective of its interaction with urban aerosols”.

Changed to: *“While the analysis box along the filtered mean trajectories captures the effects of urban aerosols on cloud dynamics and microphysics, the analysis box along the overall mean trajectory (Appendix A) represents convective cloud properties from grid cells irrespective of their interaction with urban aerosols.”*

Line 250 – replace “air masses” with “air parcels”

We refer to our response to the previous comment, where we explain our choice of "air volume" over both "air mass" and "air parcel".

Lines 283 to 290 and lines 314 to 320 – Please include reference to synoptic charts in supplement material

Added.

Line 303 – delete “pattern” from “the difference pattern”

Deleted.

Lines 310 – please reference the figure, subpanel, and cardinal direction of the edges of the noted aerosol effects

We thank the reviewer for this suggestion. The specific figure references, subpanels, and cardinal directions are provided in the preceding sentences, where each stage of the precipitation response is described individually. The final sentence at line 310 summarizes the overall range and spatial characteristics of the urban emission influence and adding full references there would be redundant with the detail already provided above.

Line 315-317 – If there is warm air over cold air, it implies increased stability not instability. Please rewrite.

Changed to: *In the second case, a shortwave trough moved northeastward over Benelux on the forward flank of a broad Atlantic upper-level low. Meanwhile, a cold front reached western Germany during the day. Upper-level flow from the southwest advected warm, moist subtropical air into the lower troposphere, while colder air aloft associated with the approaching trough increased lapse rates and conditional instability.*

Line 329 – rewrite " a clear negative precipitation anomaly appeared in the NONURBAN experiment (Fig. 5f)," as "a negative precipitation difference between experiments is seen in Fig. 5f,"

Changed.

Figures 6 and 9 captions list hail as part of total condensate, but the microphysics scheme described in the methods section (2.1.1) has only five hydrometer species (cloud water, ice, snow, graupel, and rain). Can you please clarify how hail category data is obtained? Or delete hail!

We thank the reviewer for identifying this inconsistency. Hail is calculated by the two-moment microphysics scheme but was accidentally left out of the description in Section 2.1.1. We have added it to the list of hydrometeor species in that section.

Figure captions and in the manuscript: please replace "spatially" with "horizontally"; replace “box spatially’ with “horizontally”; replace “box averaged" with “horizontally averaged”; because box may refer to 2D or 3D box and spatial averaged may be horizontal, vertical, or both. So, it is confusing

Changed.

Line 379 – replace "pristine" with "background aerosol"

Changed.

Line 401 – for inference on increase in precipitation due to urban aerosols; please reference the supplemental figure or Figure 7D

We added reference Fig. 7d.

Line 409 – replace “The subsequent slower decay” with “The subsequent loss of cloud droplets at a faster rate”

Changed.

Figure 9 – delete “shows box-averaged”; because all the variables are box averaged which is mentioned at the start of the caption

We thank the reviewer for this suggestion. However, "horizontally averaged" is retained for the rain rate to explicitly distinguish it from the other variables in the figure, which are additionally averaged vertically. Removing the qualifier would imply that the rain rate is processed identically to the other variables, which is not the case.

Line 434 – up to 5% in "NONURBAN"?

Added.

Lines 446 to 449 – For the July 20th case, how much of the convective event duration is reduced due to strong downdrafts causing the early decay of systems in the BASE emission scenario? I don't see a difference in duration in Figure 9. Can you please clarify in the text?

We thank the reviewer for this comment. This issue was also raised by Reviewer 1 and has been addressed in the revised manuscript. The statement regarding early decay has been replaced by a more precise description of the dynamical changes, namely the suppression of updrafts by enhanced downdrafts, which is supported by the vertical velocity analysis in Fig. 11.

Line 456- rewrite "passing through" as "interacting with. Since the simulated convective system on July 13th didn't directly pass through Leipzig, it interacted with the plume southeast of Leipzig.

Changed.

Line 464 – Related to major comment, if the parts of multiple convective systems within the analysis box are examined; how can the downdraft be tied to a single system?

We thank the reviewer for this comment. Following the major comment on the moving box analysis method, the manuscript now explicitly states that for case II the analysis box captures a portion of the squall line, which may contain multiple convective cells. The downdraft signal is therefore described as a characteristic of the convective activity within the analysis box during that period, rather than being attributed to a single system. The relevant passage has been revised accordingly.

Discussion and conclusions section – Contribution of urban emissions to total aerosol is < 3% in both cases. I think this is an intriguing result. Is the small contribution significant to see microphysical effects? Can you please add this result and provide some discussion?

We thank the reviewer for highlighting this result. This point was also raised by Reviewer 1 and has been addressed in the revised manuscript. A discussion of the implications of the small urban aerosol contribution has been added to the Discussion and Conclusion section.

" A notable aspect of this study is that the urban emission contribution to the total aerosol loading amounts to less than 3 % in both cases, as the regional background aerosol concentrations dominate. Despite this small relative perturbation, statistically significant microphysical responses and spatial redistribution of precipitation are detected, with magnitudes comparable to studies that examined substantially larger aerosol perturbations. However, domain-mean surface precipitation and total condensate remain within the ensemble spread, indicating no statistically significant net effect on total precipitation amounts. This suggests that even modest urban emission perturbations can produce detectable impacts on convective clouds when the aerosol-cloud coupling is represented explicitly and the analysis is designed to isolate the urban signal from natural variability. Nevertheless, a systematic and robust alteration of convective precipitation patterns likely requires substantially larger emission perturbations, as indicated by sensitivity experiments across a wider range of emission scaling factors in a companion study. These findings have implications for projections of urban aerosol

effects under changing air quality conditions: as background aerosol concentrations decline due to emission controls, the relative contribution of urban emissions to total CCN may increase, potentially lowering the perturbation threshold required to produce systematic effects on convection."

Discussion and conclusions section – Also, the difference in contribution of urban aerosols to the timeseries of volume-averaged total condensate or mean areal rainfall lies within the ensemble spread. I think this is also an important result to include in the conclusion and the abstract. Though there are differences in vertical profiles, the net effect on surface precipitation is not significant.

We have added this finding to the abstract and conclusion to explicitly state that urban emissions redistribute precipitation spatially without producing a statistically significant net effect on total precipitation amounts, despite detectable microphysical responses.

"However, domain-mean surface precipitation and total condensate remain within the ensemble spread, indicating no statistically significant net effect on total precipitation amounts. This suggests that even modest urban emission perturbations can produce detectable impacts on convective clouds when the aerosol-cloud coupling is represented explicitly and the analysis is designed to isolate the urban signal from natural variability."