We would like to thank both reviewers for their thorough work and helpful feedback on our manuscript. The feedback helped to significantly improve the manuscript.
In this document, all remarks by the reviewers are listed in black text, and our corresponding replies are given in blue. In addition to revising the manuscript following the reviewers' comments, we also gave it an editorial review.

Responses to Anonymous Referee#1

Drones equipped with low-cost sensors is a new approach for measuring trace gases within the planetary boundary layer. They enable vertical measurements of such gases, complimenting gaps in our current observational system. This manuscript explores a new application of drone observations. The authors use 2-day drone observations of O_3 and NO at a polluted site in Germany to perform data assimilation. They found improved performance in simulated levels of O_3 and NO_x after assimilation, showing enhancements not only in local vertical profiles but also across a broader spatial region at ground level. This topic is interesting, and the results show some potential for future applications in improving environmental forecasts. The manuscript is well organized and easy to follow.

Thank you very much for your thoughtful review, seeing the potential of our work, and your helpful comments!

My major concern is about the representation error and the correction of emissions. The authors use observations from a single location near the emission sources, to correct a model grid cell averaging 5km x 5km, which introduces representation error.

We agree with referee#1 that a representation error arises due to the mismatch between the spatial scale of the measurement and the spatial scale of the model grid cell used in simulations. However, a representativeness error has been taken into account for our assimilation runs. We included a related discussion in the newest version of the manuscript.
The representation error is calculated considering the model grid cell size dx and the specific characteristics of the observation locations by:

$$\epsilon_{rep} = \sqrt{\frac{dx}{L_x}} \times \epsilon_{abs}.$$

This equation is taken from Elbern et al. (2007), where L_x is the characteristic representativeness length and ϵ_{abs} is the absolute error.
As suggested, we added the information and assessment of the representativeness error in Section 3.3 (See Lines 170-181).

Indeed, it seems logical for the model to initially underestimate drone-observed NO (due to the model's coarse resolution being unable to resolve fine-scale mobile emissions) and overestimate O_3 (because the model's resolution cannot resolve the close-to-source NO titration of O_3). In such cases, directly aligning the model with single-spot drone observations and attributing this difference to emissions might result in a significant upward correction of NO emissions, as shown in the authors' results (with correction factors up to 15 on the second day, and the locations of these corrections are suspicious). However, these corrections could be false. Therefore, their findings regarding emission corrections need cautious interpretation and thorough discussion.

Thank you for his comment. We elaborated the discussion of the analysis with respect to the emission correction factors in the manuscript to better explain our results.
Firstly, we noted that the way we presented the correction factors in our analysis could potentially lead to false conclusions. While the emission factors are high for some grid cells, the actual emission rates

are not correspondingly high. In other words, the minimum of the cost function might lead to strong emission correction factors with little impact due to small a priori emissions. To address this, we now present an analysis of the absolute emissions changes by source sector in new Figure 4. This analysis allows us to demonstrate that the emissions on the two days of the study are comparable in terms of the quantity of emissions added and more reasonable when compared to the original emissions.
Secondly, we incorporate a comparison to TROPOMI (TROPOspheric Monitoring Instrument) NO_2 tropospheric columns. This satellite data offers valuable insights into the different pollution situations on both days. This explains the discrepancies observed in the emission optimisation results. In addition, We finally discuss the results in light of the wind conditions that mainly determine the location of the emission corrections (Section 4.2). These are different for the analysed days. However, it remains difficult to validate emissions rates in air quality studies. Thus, we refer to the validation of the model results against ground-based observations in order to evaluate the overall impact of the drone data assimilation. This evaluation still allows us to discuss the potential contribution of emissions optimisation to the observed improvements (Section 4.4). Based on these points, we made significant changes, especially in the Sections 4.2 and 4.4 to address your helpful comment.
Section 4.2 has been reworded, and more analysis added.

The old Figure 5 :

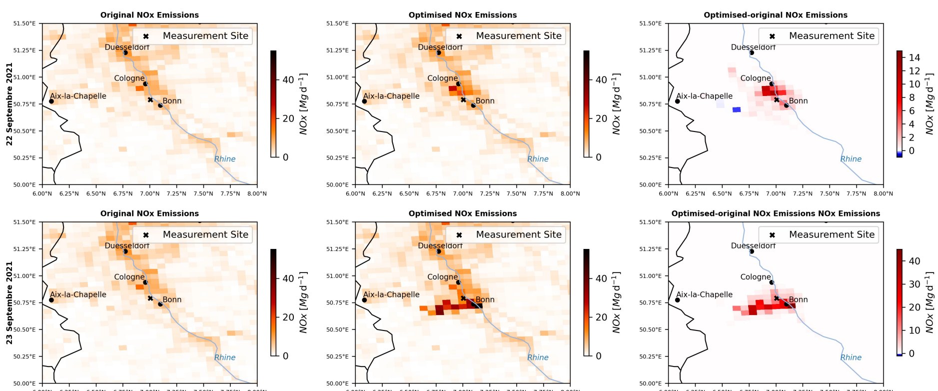

Figure 1: NOx emissions from the original emissions inventory (left), the optimised emissions resulting from the data assimilation simulations (middle), and the difference between the original and optimised emissions (right).

The new Figure 4 :

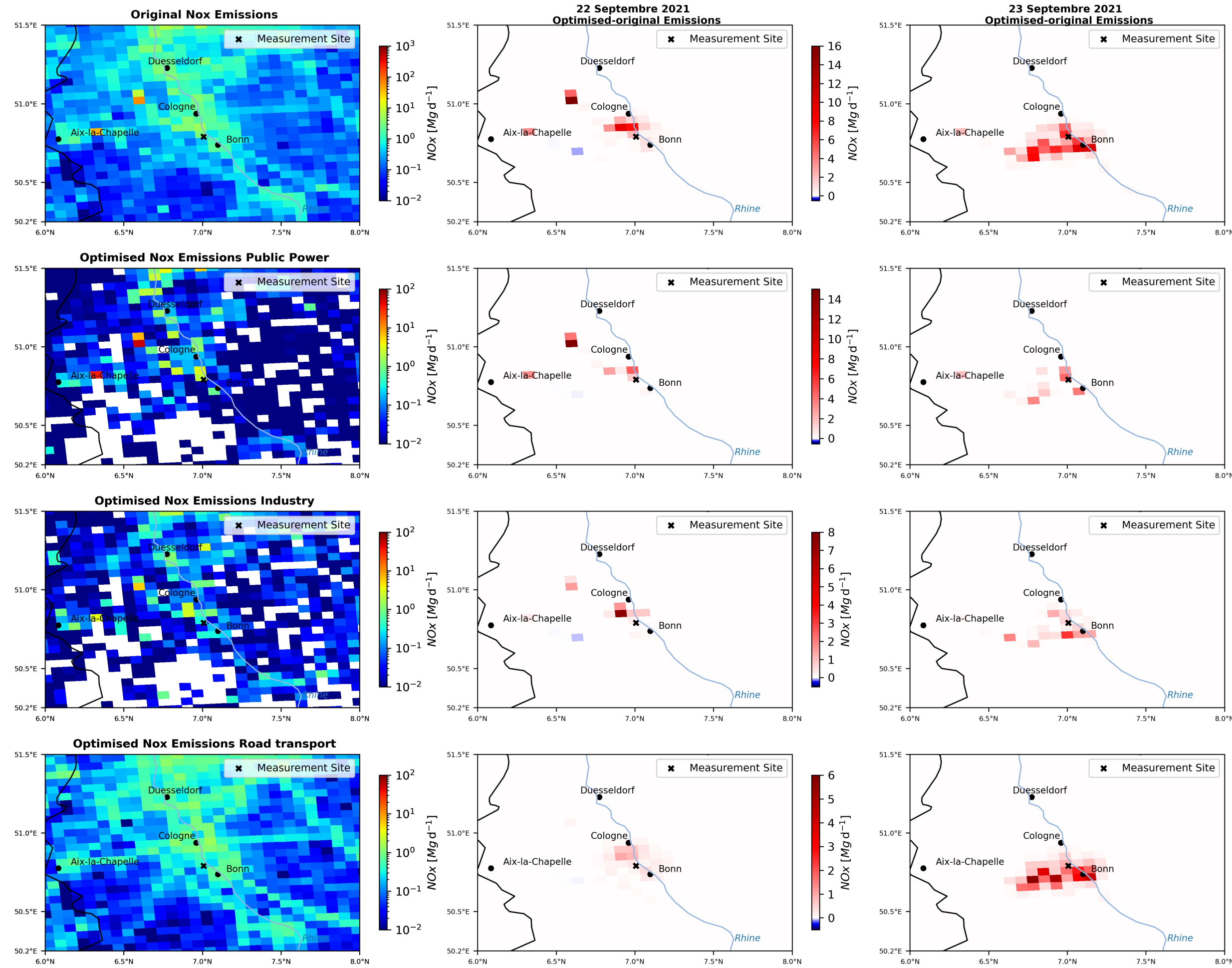

Figure 2: Daily NO_x emissions within the analysed domain (first column) and the analysed NO_x emission changes on 22 September (middle column) and 23 September (last column) 2021. The rows (from top to bottom) display the total NO_x emissions, and the emissions from the public power production, industry, and road transport, respectively.

Other comments are outlined below.

Major comments:

- 1. In Section 2 and 3, could the authors consider adding an introduction on the accuracy and uncertainty of drone observations, and on how the observation error is specifically treated in their 4D-var system?

  We agree that a discussion of the observation errors and uncertainties is of special importance to this study. Following on the referee comments, valuable information regarding the observation error of the drone measurements in the 4D-var system has been incorporated in the manuscript. You will find the major changes in the lines 170-181:

- 2. Line 234-240, what is the authors' definition of "optimal" conditions for efficient emission optimization? How do these conditions affect emission optimization, and are the two cases in this study considered optimal? Clarifications are needed.

By "optimal wind conditions", we refer to scenarios where the observation site is located on the downwind side of emission sources. In such cases, the observations are directly impacted by the emissions if the diffusion is sufficiently strong to disperse the pollutants effectively. Under these conditions, emissions corrections are more accurate and robust. This is because the wind direction and strength ensure that the observed pollutant concentrations accurately represent the emissions from the sources, as supported by Wu et al. (2020).
Regarding the two cases in this study, the wind conditions differed (see Figure 2 of the manuscript). We have added explanations in Section 4.2, where we discuss the differences between the two days and how they impacted the emissions optimisation results.

- 3. Table 6, what is the definition of "partial cost" and how is it calculated? Could the authors consider adding this information?

  Thank you for raising this point. The partial costs refer to the cost term for the initial values (J_b) and for the emissions (J_e), as the 4D-var cost function is composed of J_b, the background term, J_o, the observation term, and J_e, the emission term, with ($J = J_b + J_o + J_e$). This explanation has been added to line 110. In Table 6, we represent the percentage of these partial costs relative to the total costs J at the analysis iteration. This allows to evaluate if the minimisation algorithm attributed the model-observation discrepancies to uncertainties in the initial values or the emissions.
  As suggested, this information has been added to the revised manuscript on lines 340-342 as follows: "..., we present in Table 6 the cost reduction for O_3 and NO, as well as the partial costs attributed to the optimisation of the initial values (IV) $\left(\frac{J_b(\mathbf{x_0})}{J(\mathbf{x_0},\mathbf{e})}\right)$ and the emissions correction factors (EF) $\left(\frac{J_e(\mathbf{e})}{J(\mathbf{x_0},\mathbf{e})}\right)$."

- 4. Please change "NOx" to "NO_x" throughout the text.

  Thanks for spotting these typos. We replaced 'NOx' by 'NO_x' in the entire revised manuscript.

We would like to thank both reviewers for their thorough work and helpful feedback on our manuscript. The feedback helped to significantly improve the manuscript.
In this document, all remarks by the reviewers are listed in black text, and our corresponding replies are given in blue. In addition to revising the manuscript following the reviewers' comments, we also gave it an editorial review.

Responses to Anonymous Referee#2

Drones bring new opportunities to improve air pollution monitoring vertically within the Planetary Boundary Layer due to their portability, flexibility, and affordability. The authors apply drone profile measurement of O_3 and NO to optimize the anthropogenic emissions using the 4D-var data assimilation system of EURAD-IM. As the first application of drone data assimilation within a CTM, research is interesting, which also offers new insights and implications for future studies on emission assimilation. The authors' effort in conducting drone measurements, analyzing the data, and presenting their works are greatly appreciated. However, I still have some concerns in terms of their methodology.

Thank you very much for seeing the value of your work to the community and for the helpful comments you provided. We revised the manuscript significantly taking into account all comments from both referees.

The authors use of drone measurements at a single point to infer NO emissions across a 5 km $\times$ 5 km grid box, which raises concerns regarding data adequacy, potentially introducing significant bias in estimating emission correction factors.

We agree that this setup is not optimal. However at the current stage, the single point profile observations are the only available data set we have for analysing the drones impact on model analyses performing data assimilation. We see the discrepancy regarding the representativeness of the drone data compared to the model resolution. This is nevertheless taken into account by applying a representativeness error to the observation data. This is now comprehensively discussed in the manuscript (see lines 170-181). Moreover, we have extensively worked on the discussion on estimating emission correction factors (Section 4.2). For really proposing emission correction, the statistical evidence is missing. However, the emission corrections are a component of the data assimilation technique being applied and we aim at discussing its full performance.

Particularly, the NO and NO_2 emission correction factors derived from DA_23SEP exhibit a 4 to 5-fold increase compared to DA_22SEP. The results seem to be counterintuitive as anthropogenic emissions typically exhibit small changes over two consecutive days, unless extreme events occur that lead to significant changes in NO and NO_2 emissions. Even though the authors get 'improved' simulations after the assimilation, I assume the observed large differences in NO_x concentrations would be more related to the daily variations in transport, either horizontally or vertically.
In such cases, improving the representation of the meteorology of the model would likely be more beneficial than merely adjusting pollutant emissions. The authors might consider comparing meteorological parameters in their simulations with observations, particularly focusing on variables like winds, to identify potential discrepancies.

Thank you very much for this comments, which helped us to correct and improve the manuscript. We have identified that the old Figure 5 presenting the change in emissions was misleading. Thanks to your remarks, we decided to replace it with a Figure (Figure 4) illustrating the absolute changes of the daily emission changes. For the correction factors, even if they are large (in some grid boxes), they must be interpreted with caution. The model grid cells with high correction factors are often connected to emission rates that are very low. This is clearly illustrated in the corrected Figure 4 and

Figure A1. In addition, the emissions changes of the two days are comparable and reasonable when compared to the original emissions although different grid cells are affected by the corrections.
The horizontal transport and wind conditions primarily contribute to the observed differences, particularly in terms of location. In fact, during the first day, the wind in the region was low, causing an accumulation of pollutants, while on the second day, the winds were slightly stronger, allowing better transport and dispersion of emissions. We have added to the manuscript a new figure (Fig. 5) representing the tropospheric NO_2 column of the TROPOMI satellite, which clearly shows the difference between the pollution situation during the two days. Taking this knowledge into account helps us to explain the difference between the two days in terms of the location and to interpret the amplitude of the emission correction.
The old Figure 5 :

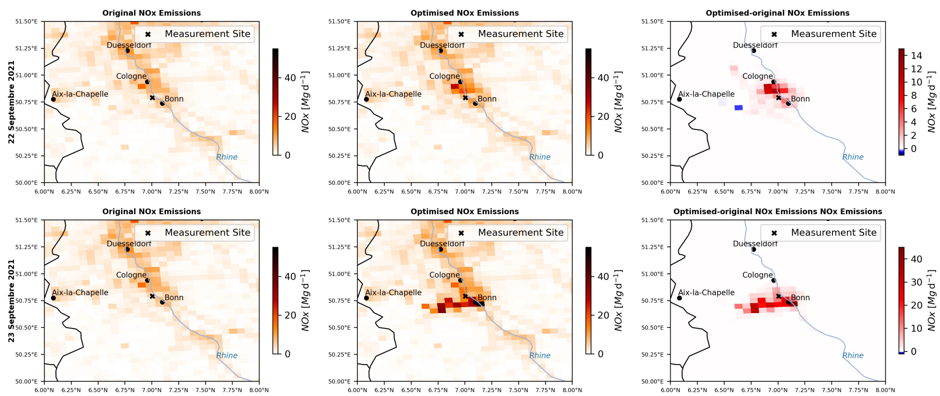

Figure 3: NOx emissions from the original emissions inventory (left), the optimised emissions resulting from the data assimilation simulations (middle), and the difference between the original and optimised emissions (right).

The new Figure 4 :

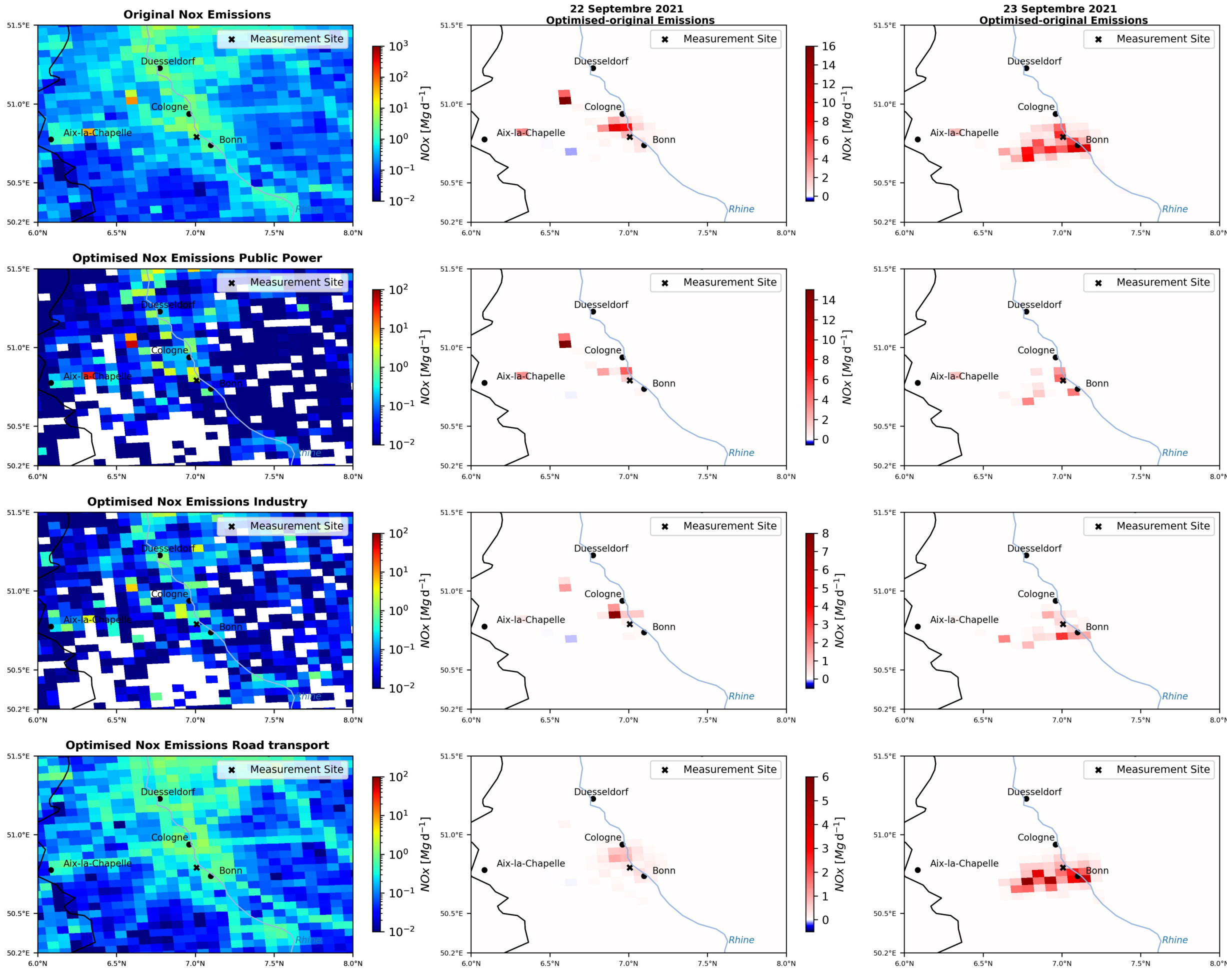

Figure 4: Daily NO_x emissions within the analysed domain (first column) and the analysed NO_x emission changes on 22 September (middle column) and 23 September (last column) 2021. The rows (from top to bottom) display the total NO_x emissions, and the emissions from the public power production, industry, and road transport, respectively.

The calculated emission correction factors for NO and NO_2 can be even larger than 15, suggesting the regional emission inventory they use in their simulation have an uncertainty of over 1400%, which might not be a reasonable value for emission correction. I suggest the authors checking their anthropogenic emission inventory and previous emission assimilation studies to identify a scientifically reasonable range for their emission correction.

We appreciate this comment. We realise that way we presented the correction factor in our initial analysis could potentially lead to false conclusions. Although the emission factors are high for some grid boxes, the absolute emission rates are not correspondingly high and thus did not result an unrealistic emissions change. To address this issue, we have provided an analysis of absolute emissions changes by sector in Figure 4. From previous emission correction studies, we know this behaviour, which appears extreme at the first sight. To propose general emission correction factors to the emission inventory used, it would need a more extensive analysis (more observations for a longer time frame), as the emission input to the chemistry transport model bases on annual emission values per grid cell. The emission corrections analysed here are rather a correction of the time profiles applied to derive daily emission corrections per grid cell. Section 4.2 has gone through a major transformation to clarify the interpretation of the analysed emission factors.

Additionally, I am not sure whether the authors' observational results would affect by the wind conditions. Both horizontal and vertical wind speeds can exceed the ascent rate (i.e., 1m/s) of their instrument, potentially affecting data collection. How did they keep their instrument ascending at a constant rate in a fixed location? NO is a highly reactive air pollutant which can be converted to NO_2 quickly upon emission. Consequently, NO usually presents significant decreasing gradient in the vertical direction. However, the observed NO concentration at 350 m can double the surface NO concentration in the manuscript (e.g., F10, F11). The result is confusing. The authors may want to check the credibility of their data and present relevant explanations.

We understand that a detailed description of the observational setup is missing in our model analysis manuscript. We solely refered to Brettschneider et al. (2022) to find all details about the drone. But we see the necessity to discuss the points addressed in the manuscript. The drone is operated by an autopilot system that uses an inertial navigation solution with an Earth related position based on GNSS data (Global Navigation Satellite System, e.g. GPS, Galileo). During the measurements, the autopilot controls a constant lateral position and a constant vertical climb rate. Wind affects only the attitude of the drone. For low wind situations (like in this study), the effect on the attitude can be neglected, the tilt angle value is low and the assignment of the inflow tube/electrochemical sensors are tilted in wind direction, so the effect on the sensor flow is even less. This information has been added to lines 145-148.
For NO observations, six electrochemical sensors are installed on the drone. The modeling experiment utilizes a dataset from the most efficient sensor. However, all datasets exhibit the same profile shape for flights F10 and F11. This allows us to attribute the higher concentration at altitude, compared to ground level, to actual emissions rather than observational uncertainties. The high NO concentration observed near the ground during the F7/F8/F9 flights is attributed to emissions from the transport sector (Figure 6 shows concentrations exceeding 120 $\mu g/m^3$ between 4 and 6 UTC). As vertical mixing increases after sunrise, the NO concentration decreases near the ground and rises at higher altitudes. Additionally, emissions from power plants and the industrial sector are released at higher altitudes in the studied area.

I recommend that the authors incorporate more observations, if possible, rather than relying solely on data from a single location, in their assimilation process. By doing so, they can enhance the robustness and credibility of their results, mitigating any potential suspicions regarding the validity of their findings.

We fully agree with your suggestion. Unfortunately, at the moment, we only have data from one measurement campaign available, which was conducted at a single location in September 2021. Because of this, we are presenting our work as a case study to highlight the potential of drone observations for air quality analysis and to encourage more measurement campaigns in the future. Including observational data from other measurement platforms would surely enhance the robustness for example in terms of the emission optimisation, however, would not fit the focus of this manuscript, which is the analysis of the benefits of drone observations. As a future perspective, we hope to collect more observations with such a drone. It is planned to perform similar flights in the Wesseling region on a regular basis for a certain period to enable a statistical evaluation. But also additional campaigns in polluted regions are foreseen.

Specific comments:

- L162: I assume the authors may want to say 'with data assimilation' here or above (i.e., L158)?

Otherwise, all the four experiments are conducted 'without' data assimilation, which is not consistent with what is shown in Table 2.

We are grateful for helping us to avoid the confusion with this phrase. We replaced the sentence: "covering the same period as reference simulations without data assimilation" by "for 22 and 23 September 2021." for a better clarity.

- L218-219 I am not sure whether anthropogenic emissions can have such large differences in two consecutive days.

  Many thanks for raising this point. We fully revised Section 4.2 on the emission optimisation to clarify the aim, the quality, and significance of the analysed emission factors.

- L227-228 As this is a model study, more quantitative analysis is expected.

  Based on your suggestions, we included an additional analysis on the emission corrections (the new Fig. 4), which highlights the sectors subject to significant changes in emissions over the two-day study period. Furthermore, we revised the whole manuscript to enhance the analysis accordingly.