Reviewer Report

Manuscript Title: *'The potential of drone observations to improve air quality predictions by 4D-var'*
Manuscript Number: egusphere-2024-517
Authors: Hassnae Erraji, Philipp Franke, Astrid Lampert, Tobias Schuldt, Ralf Tillmann, Andreas Wahner, and Anne Caroline Lange

The authors present an interesting and novel study outlining the use of drone-based vertical profile measurements of O_3 and NO (collected over a 2-day field campaign) in a 4D-Var data assimilation system coupled to the EURAD-IM CTM to optimize the vertical and spatial distribution several pollutants, as well as the related emission fields. These drones may serve to fill a gap in our observations, particularly within the PBL where few vertically-resolved measurements exist. In general, the manuscript is well written and easy to follow, although I did find there to be quite a lot of minor formatting and grammatical errors throughout.

While the authors do present some reasonable evidence to support the conclusions that the drone observations lead to an improvement in the simulation of these trace-gases and that optimized emissions can be obtained, my largest concern is with the robustness of the results. In particular, I am not entirely convinced with the optimization of the emission fields, given the fact that the assimilation was only performed over a very short time-period and with generally few observations.

Major Comments:

- P6, L150-151 – The authors mention they only used the ascent profiles from the drones in the assimilation exercise due to the "higher accuracy". What is exactly meant by this - why were the descent profiles less accurate? Shouldn't it be the same if the ascent/descent rate is fixed at ~1 m/s.

  Similarly, how much does the result change when assimilating both the ascent and descent profiles in the 4D-var system? This could give some indication of how sensitive the assimilation system is to additional measurements (particularly those that might be more uncertain). Due to the overall low number of observations being assimilated over a short window of less than 12 hours on each day, I suspect that this could have a notable influence on the results and is important to investigate.

  The authors even highlight this themselves later on when discussing the optimized emissions; P12, L231-233 "However, their generalization and significance should be rated carefully, mainly because of the limited number of drone profiles being available, the short assimilation windows selected, and the deficiency to perform a long-term statistical analysis".

- From Figure 3, there appears to be a very significant bias in the modeled vs. measured NO profiles on the order of ~30 ppbv, and if this difference was placed in relative terms (in %) it would appear even more extreme. This is only very briefly addressed in Section 4.1, L215: "On both days, the reference simulations underestimate the NO vertical distribution at all heights, with the strongest discrepancies at ground level."**.** I believe a greater investigation of the source of the bias should be performed or at least explained here, as it may indicate a broader issue in the model (e.g., incorrect NOx chemistry or partitioning between NO and NO_2). I think this is particularly crucial if you are going to assume that the partitioning and chemistry of NO_x is correct in the model for the purposes of estimating the NO_2 emission corrections based solely on the optimized NO fields.

- P13, L263-264, the authors write "However, it is unfortunately not possible to directly obtain information about the NO_2 emissions from the TROPOMI data.". While this is somewhat true, in theory information on the NO_2 emissions could be derived from assimilating the S5P observations in the 4D-var system. I'm not necessarily saying this must be done in this paper as it may fall slightly outside of the scope, but it could provide an interesting comparison with those optimized emissions obtained solely by assimilating the drone observations of NO. I think it would also provide increased confidence in the results if separately assimilating both datasets provided a similar result in terms of the optimized emission fields (at least for NO_2) for these days.

Minor Comments:

- P1, L6: "4D-var takes advantage of the inverse technique…", I think it would be better here to say something along the lines of "4D-var is an inverse modelling technique". There is no single inverse technique, but rather a slew of inverse modelling approaches and methods.
- P2, L25-29, the authors state that very few ground-based monitoring networks exist that provide vertically resolved measurements of these pollutant species. They mention LIDAR and sonde networks, but fail to mention ground-based spectrometer networks such as the Network for the Detection of Atmospheric Composition Change (NDACC) or the Pandonia Global Network (PGN; global network of Pandora UV-Vis spectrometers). These networks provide vertically resolved measurements of O_3 and NO_2. I suggest that the authors revise this section of the text to include mention of these other networks as well.
- P3, L69: "The aim is to investigate the ability of the 4D-var to adjust". Minor grammatical comment, but I suggest inserting "system" here so it reads "The aim is to investigate the ability of the 4D-var system to adjust".
- General comment on the text on P10 and Table 3 – In the text here, you provide some absolute differences (in ppbv) while later on you start providing relative differences

(in %) for the biases, however in Table 3 you only provide absolute biases (in ppbv). I find this makes it harder to follow. For clarity, I suggest either providing both simultaneously in the text, or choose one and be consistent.

- P12, L226: "The 4d-var data assimilation…" the 'D' should be capitalized here.
- P13, L267-268: "…especially for emissions that are emitted at high altitudes, such as power plants and industries.", I am not really sure the top of a smoke-stack is considered "high altitude" as implied here. Maybe change the text to "higher altitudes", or something similar.
- P14, Figure 4 – In the titles of the left column figures "Nox" should be changed to "NO_x". Also, "Septembre" -> "September" in the top titles of the middle and right-most column of figures. I also suggest using a consistent range for the color-bars at least for the middle and right-most column of figures. As it currently stands, upon a quick look it appears as though the road-transport sector has the largest difference, but this is only because the color bar scale is much lower.
- P15, Figure 5 – In the color-bar label, "molec cm-2" should be written as "molec. cm^{-2}".
- P15, L285-286, the improvement in the Pearson correlation coefficient of 0.15 here does not coincide with the values listed in Table 4. Only a difference of 0.04 is listed in the table.
- P15, L288-291; the authors state here that a "remarkable improvement" in the O_3 concentrations is seen for the beginning of the day on September 23, but neglect to discuss the fact that the optimized simulation agrees more poorly with the observations for the remainder of the day, particularly at the end of the day. I think this should be highlighted more in the text here, as the overall simulation has not necessarily been "improved" with respect to the ground-based observations, only for a short period in which there was originally a large discrepancy. This is discussed to some extent later on P20, but should probably be mentioned here first.
- P16, Table 4; I think it should be made clearer (in the table caption), what values are being provided in the brackets. It does not seem to be mentioned anywhere, are these standard deviations/standard errors?
- General comment on P15-16; similar to my earlier comment about the absolute/relative differences when discussing the comparisons, in the text here on these two pages the authors make statements such as "the assimilation of drone observations results in a strong reduction of the bias by 87% and the RMSE by 20% with an amelioration in the Pearson correlation of 0.15". I believe important context is lost by the authors only providing the differences, but not mentioning in the text what the initial and final values were. For example, if the original bias was 2 ug m^{-3} and the new bias is 1 ug m^{-3}, then you can say there was a 50% decrease in the bias, but this would not be as significant as it sounds. I think the text should be amended here to include this.
- P17, Figure 6 – O3 and NO2 in the subplot titles should be subscripted as O_3 and NO_2.

- P19, Figure 7 – This figure looks a bit too simplistic and like it was made hastily, and I do not think is of publication quality, particularly in relation to the other figures in the manuscript. "O3" in the title should also be subscripted, this also applies to Figure A2, A3, and A4.
- P20, L338-339; the text currently reads "Moreover, the assimilation process allows to obtain optimised emissions rates", this should read "allows *one* to obtain…" or "the assimilation process provides optimized emission rates".
- P20, L348-350 "This is supported by the findings of Wu et al. (2022), affirming that observation at high altitudes can be advantageous for optimising emissions under suitable wind 350 conditions", what exactly is supported by the findings of Wu et al. (2022) here? The authors need to elaborate on this point.
- P20, L361: "Secondly, Some…", "some" should not be capitalized here.
- P20, L362: "(Fig.6, Fig.A3, and Fig.A4)" should have spaces between "Fig." and the figure numbers.
- P21, Figure 8: "NO2" should be subscripted in the y-axis label of the bottom left figure panel.
- P22: The section heading for 'conclusion' should be capitalized.
- P22, L395; Minor comment but Observing System Simulation Experiments (OSSEs) is a more commonly used term for this, not OSE. I am also not quite sure what the authors mean here by " …to assess the advantages and limitations of integrating drone observations into CTMs through the application of a variational data assimilation technique", is that not exactly what was done in this work?