

## Referee comment on AMT-2023-15

The manuscript entitled “Long-term airborne measurements of pollutants over the UK, including during the COVID-19 pandemic, to support air quality model development and evaluation” by Angela Mynard et al., introduces a dataset of airborne observations taken with the MOASA measurement platform during 63 flights over UK in the period June 2019 to April 2022. These are useful and interesting measurements that should be shared with the scientific community. However, while the measurements likely provide important new information, the manuscript has the character of a technical report and the analysis of the few data examples presented is poor and needs to be extended.

The manuscript can be published in AMT after a major revision of the structure and the scientific contribution of the work.

### General comments

Apart from the potential usefulness of the dataset, the conclusions of the paper are unfortunately lost in the extended technical description of the dataset. The interpretation of the examples presented is generally too simplistic. The manuscript would benefit from moving most of the technical details of the database to the supplement while extending the analysis of the examples given to provide a concrete picture of the actual improvement achieved through this observations in air quality model development and evaluation, in particular for the case of the AQUM regional model, as indicated by the title and announced in the introduction.

In addition, the interpretation of the differences in ground based observations during the pre and post COVID19-lockdown is poor and simplistic, and is not linked to the airborne data obtained with the MOASA platform.

### Specific comments

Abstract: The abstract does not include any findings. Please highlight them.

Introduction: The introduction is too long and has rather the character of a measurement report than of a scientific work. Most of the details and acronyms are neither relevant for the data analysis presented or further mentioned in the manuscript. Please move them to the supplement. In particular, the information about COVID19 can be reduced to a paragraph and does not require a separate section.

Line 94. Please include a more specific publication for GOME than Molina and Molina, 2004.

Line 122: Please specify what is meant by “*an introduction to the vertical structure of pollutants during COVID 19 period*”. Which are the findings? Please include them in the conclusions and the abstract.

Line 145 Section 2: This section should be shortened and most of the information moved to the supplement. Please include in the main text a table with a summary of the instruments and their most important features, sensitivity, detection limit etc and refer to the supplement for details for the measurement techniques, calibration procedures, flagging of data, etc.

Line 167: It is confusing to mention an Appendix A that actually is a Figure A1. This happens with the rest of appendices all over the text, which are difficult to be identified. Please include them in the supplement and name them accordingly (e.g. Figure S1 in the supplementary information).

Line 203: Why is specifically mentioned that the instrument measures up to 3000 ppbv? This NO<sub>2</sub> mixing ratio is not expected to be frequent in airborne measurements in the BL if not directly flying inside industrial plumes. How is the accuracy at the lower end and detection limit? Please clarify also in relation to the sensitivity given in Line 242.

Line 234: Figure 1 and Figure 2 do not seem to be mentioned in the text. Please correct.

Line 241: It would be more informative to show the data of this separate experiment instead of the data that were corrupted by the impact of the filter in the pump performance.

Line 259: Does it mean that the first 7 flights do not have any valid O<sub>3</sub> data?

Line 335: Please revise the figure caption of the Appendix B, i.e., figure B1 to make it more understandable: "*The vertical bars represent the error in response for each bead size and is the mean standard error of the mean for 15 second segments of each bead response.*"

Line 359: Is Appendix 5 the in Line 369 mentioned Appendix C, which in reality are the Figures C1 and C2? Please clarify and come to a systematic naming of the so-called appendices to avoid confusion.

Line 370: "... *it is subject to potential uncertainties...*". Are these uncertainties the size distribution uncertainties described in the following section or additional uncertainties for other reasons? Please clarify in the text and revise the necessity of a separate 2.7.3 section for this. As recommended above, part of these details should anyway move to the supplement.

Line 415: Is the Appendix D the table D1?

Line 422 "were" instead of "are"

Line 431: Why is Figure 8 mentioned before Figure 7? Please change the numbering.

Line 444 to 505: All these subsections are not necessary and can be removed. To the main manuscript belongs the content of the summary (3.7) and the figures 7-8. The table 1 and the rest of the information should move to the supplement.

Line 517: I guess that Appendix 7 is the Table F1. See comments above related to the naming and numbering of Appendices.

Line 554: It is not clear how the so-called "ground distances" are classified. What is the reference used? Are they distances over the same geographical area during different flights, distances to a selected source, distances flown during the same time interval at different velocities of the aircraft? Please clarify.

Line 558: What is the interest and usefulness of having a flight mean variability in the case of different tracks or transects over areas of different chemical and meteorological complexity? In a rough analysis, without getting into the details of the sampling area, it is somehow obvious that in a non-remote atmosphere, the greater the sampling scale the more difficult becomes to observe the effect of individual pollution sources. The suitability of the scale will depend on the characteristics and pollution complexity of the area studied.

Line 580: "*For NO<sub>2</sub> this absolute variability is below 7.35 ppbv and for particulate counts below 2412.830 counts/second for 90% of data points*" What are the implications of these results for the analysis of the area sampled when using the regional model?

Line 592: "... *no statistical post processing has been applied*". How will this affect the results obtained? Please clarify what you expect.

Line 598: From here the structure of the section 4.3 is a bit erratic: the subsection of O<sub>3</sub> mentions both flights but only discusses one, the subsection of NO<sub>2</sub> and the BL height discusses only M270 and the following section seems to focus on the NO<sub>2</sub> of the M296. I would recommend to discuss the flights individually and to summarise the findings at the end.

Line 604: In Flight M270 the model data seem not to reproduce the variability of the observations at any altitude (the model results vary in each level a few ppb while the observations vary around 20 ppb). Do you have any reason for this? The M296 shows quite a different pattern. Please discuss these differences.

Line 609: "*It may be possible to use the aircraft observations to help identify sources....*" It would be very useful to see if any correction based on the MOASA data (such of this used by Savage et al. ) introduces any improvement in the case studies presented. Similarly, it should be shown the effect of replacing the modelled by the observed BL height.

Line 613: "...shows "the" comparison"

Line 615: Actually, the largest difference between model and observations seems to be below 600 m (in red). Please change the way of plotting the difference (a colour scale is not clear enough) and explain the differences more accurately.

Line 617: How solid is this interpretation? Has this pattern been observed on other days? Is the whole M296 within the observed BL?

Line 620: Why the error in the altitude of the BL can lead to any conclusion about the agreement at the surface (within the BL of the model)?

Line 625: Please change the colour scale for the difference plot. It is impossible to see any difference by this large range.

Line 627: It would be informative to see the Circuits 2 and 3 in the supplement. Please include them.

Line 630: There is about 1 h difference between the first and the last circuit. It is realistic to talk about the same plume? As stated in Line 633 the plume aloft has greater NO<sub>2</sub> concentrations. How do you explain this if both circuits are within the BL?

Line 636: How is the comparison between O<sub>3</sub> modelled and observed in those circuits?

Line 640: As the concentrations at the ground level are also as low as in the model and so different from the airborne measurements, these results indicate that you have a real gradient in the pollutant concentrations within 423 m (altitude of the first circuit), that I guess is still within the BL on that day. Please comment on this. Is that gradient also visible in other species measured during this flight?

Line 646-647: Is that the case for M296?

Line 651: Please change the scale of NO<sub>2</sub> in Figure 16 to enable a more accurate comparison with the airborne data (should not be larger than 50 mg/m<sup>3</sup>). The range of the London AURN data can be specified in the figure caption. Generally, a table with the values

used for the comparison would facilitate the interpretation. In the figure caption please correct “corresponding”

Line 654: I do not understand what is the meaning of 6.6 or 4.6 sampling sites. How can you have a fraction of a sampling site? Please clarify.

Line 663: Please rephrase. “reform” does seem to indicate a null cycle

Line 664: “*As such, the increase in O<sub>3</sub> is coincident with a reduction of the observed NO<sub>2</sub> aloft, which, in addition to being reduced by chemical reaction, is also further away from sources (fossil fuel burning, traffic (Jones et al., 2021, Lee et al., 2020)).*” It is not clear what this sentence tries to say. It seems to be quite a simplistic analysis. Please clarify.

Line 666: “*Here, the impact of external factors (meteorology, boundary layer height, seasonal changes, complex chemistry) are not discussed and is beyond the scope of this paper. However, the persistent difference between the surface-based observations and airborne observations aloft demonstrates the importance in quantifying the vertical structure of pollutants.....*” This is quite well known, what is then the scope of this work?

Line 679: If the Appendix E is the table E, the difference between the pre and post COVID19-lockdown averages is well within the standard deviation of the averages in all the sites. That implies that taking into account the large variability of these hourly averages they are not significantly different.

Line 681 to 690: The ozone production is known to be a non-linear and complex process and is not surprising that changes during the lock down cannot be explained by a simple comparison of NO<sub>2</sub> and O<sub>3</sub> hourly averages of ground-based measurements. As cited in the paper there are a multitude of studies on this subject in the literature. It is not clear why this manuscript includes here such a simplistic interpretation of the data from ground-based stations and then recommend others to make further work in interpreting the data. As these statements do not seem to complement in any form the cited comprehensive analysis of COVID data published by Lee et al, 2020, I recommend either discussing more in detail the relation between ground based and airborne measurements or otherwise removing this part from the manuscript.

Line 723: “*Specifically, we show lower concentrations of NO<sub>2</sub> and higher concentrations of O<sub>3</sub> aloft.*” Please revise this statement; it does not reflect accurately the results shown.

Line 725: “*Analysis of long-term surface-level trends in the Greater London region show a decrease in NO<sub>2</sub> and an increase in O<sub>3</sub> following the mandated COVID-19 restrictions.*” Please revise carefully the accuracy of this statement based on the analysis and the interpretation presented in the manuscript.

Line 838: Please correct the title of the reference