Author response on AMT-2023-15

We would like to thank reviewer #2 for taking the time to review this manuscript and for providing valuable, constructive feedback and suggestions that helped us to further improve the manuscript. We have carefully considered all the comments and revised the manuscript accordingly.

Below are reviewer #2's Specific Comments, in black, with an in-line corresponding reply from the authors in blue.

Referee Specific comments:

1. A- The introductory section is too long and should be improved. It contains detailed descriptions of the AQUM model and the Automatic Urban and Rural Network (AURN), an automatic ground monitoring network, which should be moved to Sect. 2. The authors feel the AQUM description is applicable to the introductory section but has been shortened. The AURN description has been moved to section 4.2, a new section: "Ground-based and airborne observation comparison using long term observations over London" (refer to comment #3).
   B- Describe in Sect. 2 also how vertical mixing is implemented into the model, since this seems to be a crucial parameter for the intercomparison with airborne measurements. Vertical mixing in the AQUM has been summarised in section 1. Given the non-diagnostic intentions of the paper (see following comment), the authors feel the following is sufficient detail: *"There are 8 vertical levels up to a model top height of 39 km and mixing is parameterised throughout the full depth of the troposphere using a non-local, first order closure, multi-regime scheme (Lock et al., 2000)."*
   C- To improve the large disagreements between the model and the measurements, it is recommended to implement other schemes to test the influence of the vertical mixing. This paper introduces the MOASA measurement platform, flight strategies and instrumentation and is not intended to be an in-depth diagnostic analysis, but rather a comprehensive technical reference for future users of these data, including illustrations of the potential uses of these upper air observations for regional-scale model evaluation. Both the abstract and introduction have been amended to emphasise this. Thus implementation of other schemes to test the influence of vertical mixing is beyond the scope of this paper but is hoped to follow in future work.
   D - In addition, the introductory section contains only few references to previous studies on this topic, e.g. Savage et al. (2013). Include more of such studies and results (as given in the Savage paragraph) and instead shorten some general information at the beginning.
   The paragraph has been amended as follows: "Comparisons of AQUM to AURN observations (Savage et al. (2013), Neal et al., (2017)), found that AQUM generally performed well, in particular for large air quality events, but had a number of systematic biases. For example, a positive bias in ozone at urban sites, a positive/negative nitrogen oxide (NO_2) bias at rural/urban sites and small negative biases in $PM_{2.5}$. These findings are generally conducive to similar air quality model evaluations that employ AURN observations, such as Williams et al., 2018 (10 km CMAQ-Urban model) and Neal et al., 2017 (HadGEM3-RA 50 km regional composition-climate model), where the latter showed a small positive bias in modelled $PM_{2.5}$. For AQUM, ground based observations are used to bias-correct the model data and minimise some of these systematic biases at the surface (Neal et al., 2014). These biases have the potential to introduce bias into any future predictions (Williams et al., 2018).
2. In Sect. 1.1 (Impact of COVID-19) incorporate and discuss results from other studies in Europe related to O3 and NO2 during COVID-19. In order to improve the manuscript (following additional feedback on this section) the description and subsequent analysis

based on the COVID lockdown has been removed and replaced with a more relevant section which compares airborne and ground-based observations over greater London. The only remaining COVID reference is to inform the reader that the observation period encompasses the COVID-affected period.

3. In Sect. 2.1 (Instrumentation – general setup) add a table listing the instrumentation, technique, precision, and references. Agreed, this has been added.
4. The results of the study (model evaluation) are rather sparse described on only 3 pages (page 16-19) compared the rather extended manuscript. Include a few more intercomparisons focusing on problems in the model (e.g. boundary layer height, vertical mixing). Add some more examples from other flights comparing the modelled and measured BL height. Discuss ways to improve the BL height in the model (add a new section "Discussion" ahead of the "Conclusions"). What about the influence of inversion layers located below the BL? Have such cases been observed in the winter flights and how does the model behave? Please refer to author response to comment #1C. These suggestions are somewhat out of scope of this technical-focus paper but are hoped to be address in future work.
5. On Page 19 (line 676-678) you write: "We define the pre-lockdown period as 26th March 2018 to 25th March 2020 and the post-lockdown period as 26th March 2020 to 25th March 2022, where comparing like-for-like months pre- and post- lockdown minimises the impact of seasonality on the comparison." □ would it not make more sense to define three periods (prelockdown, main lockdown, post-lockdown? Please refer to author response to comment #2, which advises that this section has been removed.
6. In general, it is recommended to give mean NO2 and O3 values for all flights in a table for a better overview of the airborne results. Please refer to author response to comment #2.
7. After Sect. 4 and before Sect. 5, a section on discussion of the results is missing. This will be added. A discussion/conclusion has been added to each sub-section of section 4.

Minor comments and technical corrections:

8. Page 10, line 359: Appendix 5 not available. Appendices have been reviewed and revised accordingly.
9. Page 12, line 413: gcm3 > g/cm3. Changed.
10. Page 14, line 506-508: "Datasets obtained during the MOASA Clean Air project are openly available from the Centre for Environmental Data Archive (CEDA) "Collection of airborne atmospheric measurements for the MOASA Clean Air project" repository (DOI: 10.5285/0aa1ec0cf18e4065bdae8ae39260fe7d)." Add this also at the end of the manuscript at the appropriate place. Added
11. Page 14, line 517: Appendix 5 not available. Appendices have been reviewed and revised accordingly.
12. Page 15, line 530: Add from who. Added *"…. by contacting the author."*
13. Page 15, line 547: Why data only used until July 2021 (44 flights) and not all available data until April 2022 (63 flights)? The analysis was originally completed whilst the flying campaign was in progress. This analysis has now been revised using all available flights. Figure 11 and the figure description has been updated accordingly.
14. Page 16, line 582: Sect. 4.2 is missing. Well spotted - corrected.
15. Page 17, line 607: "Savage et al. (2013) also reported biases during a ground-site AQUM comparison." discuss in the manuscript. This now refers back to section 1 (see comment 1D).
16. Page 17, line 633: "The AQUM model shows little variation and comparatively low NO2 concentration in all circuits above the city" This result gives little confidence in the model. Can you add other examples, where a city plume is well simulated by the model? Please refer to comment 1C. Other examples are beyond the scope of this non-diagnostic, technically-focused paper.
17. Page 20, line 732: Add from who. Replaced with alternative text indicating new archive.

18. Page 21-22: Some of the references are too sparse. Add more information how to find them:
    - Air Quality Expert Group: Fine Particulate Matter (PM 2 . 5) in the United Kingdom, 2012 Updated
    - DEFRA: Clean air strategy 2019., 2019 Updated
    - Ecotech: Aurora3000 Integrating Nephelometer with backscatter, 2009. Updated
19. Page 25: Give Appendix text first after Figs. 1-16. Done.
20. Page 26, line 986: Subtitle “Case study” somehow misplaced. Amended.
21. Page 29, Fig. 3: Legend covers data, move it. Blue straight line better visible in red. The figure has been updated as follows:

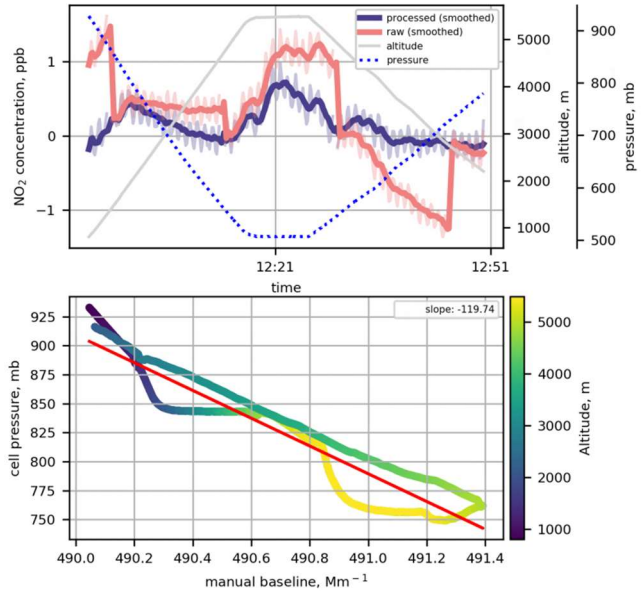
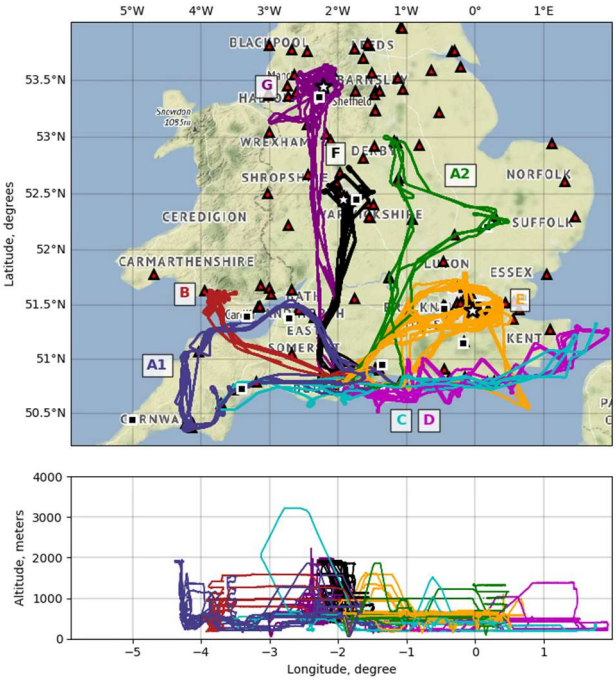


Figure 3: Top: timeseries of raw (uncorrected) and processed (corrected) NO_2 concentration. Oscillations seen in the raw and processed data during the filter test are an artefact of the filter, which impacted performance of the instrument pump. These oscillations have been minimised by arbitrarily smoothing (60 second rolling) the data, for visualisation purposes only. Bottom: baseline against cell pressure, coloured by altitude, with a linear fit shown as a red line. All data from 11:55:00 to 12:50:00 during flight M304 on 4th November 2021, averaged over 10 second intervals.

22. Page 32, Fig. 8: Here only 45 flights shown, why not all 63 flights shown as described in the abstract? This figure and associated text has been updated to include all flights.



23. Page 34, Fig. 11: In the figure text “0.85” is mentioned twice. Amended.
24. Page 43, line 1163: gcm3 > g/cm3. Amended.
25. Page 43, Table E1: Add NO2 to the header of column 3 and O3 to the header of column 4. Please refer to author response to comment #3, which advises that this section has been removed.