

Note that substantial changes have been made to the manuscript, including deletion and replacement of entire sections. **For reasons unknown, not all changes appear to have been captured by the TrackChanges pdf.**

The responses to reviewers generally include details of the changes made. Therefore:

Comments are shown in black

Responses are shown in blue.

Changes are **shown in blue bold**, with **clarification of changes** provided under the Response where necessary.

**Note that in addition to reviewer comments:**

- numerous minor amendments have been made to text throughout the manuscript.
- The conclusion has been updated.
- original appendix F and associated table F1 has been removed.

## **REVIEWER #1**

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

**Findings have been highlighted in the abstract** as follows: *“These case studies show that for observations of relative humidity, nitrogen dioxide and particle counts, natural pollutant variability is well observed by the aircraft, whereas SO<sub>2</sub> variability is limited by instrument precision. Good agreement is seen between observations aloft and those on the ground, particularly for PM<sub>2.5</sub>. ( $r^2 = 0.90$ ). Analysis of odd oxygen suggests titration of ozone is the dominant chemical process throughout the column for the flights analysed, although a slight enhancement of ozone aloft is seen. Finally, a preliminary evaluation of AQUUM performance for two case-studies suggests a large positive model bias for ozone aloft, coincident with a negative model bias for NO<sub>2</sub> aloft. On one case, there is evidence that an under prediction in the modelled boundary layer height contributes to the observed biases at elevated altitudes.”*

- 2) 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.

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. This has been emphasised in the abstract and main manuscript. The more detailed analyses to which the referee refers in “General Comments” is intended to follow in future work by both the authors and other database users.

**The use of acronyms has been reviewed throughout and revised as necessary.**

The authors agree with the reviewer’s comments on the applicability of the COVID discussion throughout the manuscript. **The description and subsequent analysis based on the COVID lockdown has been removed, as has mention of COVID in the title. The original section 4.4 has been replaced with a comparison of airborne vs ground-**

based observation of pollutants over greater London. The only remaining COVID reference is informing the reader that the observation period encompasses the COVID-affected period.

**Clarification of changes:** Acronyms revised throughout. Original section 4.4 has been removed. The whole of section 4 has been re-ordered as per below and has a new addition to replace the removed section.

- 4.1 The spatial scales of pollutant variability
- 4.2 Ground-based and airborne observation comparison using long term observations over London
- 4.3 Preliminary model evaluation

The manuscript title has been revised to: “Long-term airborne measurements of pollutants over the UK, to support air quality model development and evaluation”

- 3) Line 94. Please include a more specific publication for GOME than Molina and Molina, 2004. **This reference has been replaced with Liu, X., Chance, K., Sioris, C. E., Spurr, R. J. D., Kurosu, T. P., Martin, R. V. and Newchurch, M. J.: Ozone profile and tropospheric ozone retrievals from the Global Ozone Monitoring Experiment: Algorithm description and validation, *J. Geophys. Res. Atmos.*, 110(20), 1–19, doi:10.1029/2005JD006240, 2005.**
- 4) 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. **Please refer to author response to comment #2.**

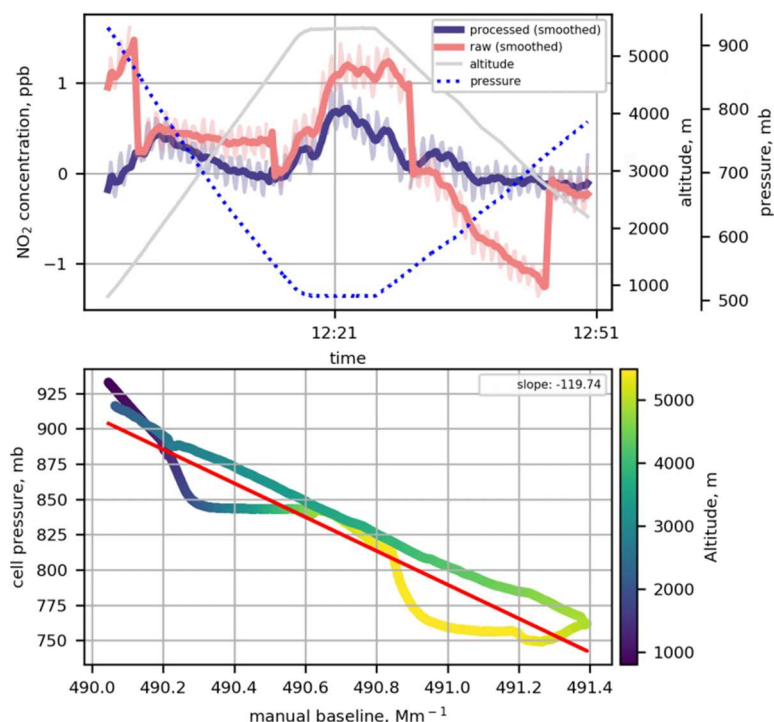
**Clarification of changes:** None.

- 5) 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. **Please refer to author response to comment #2. This section is highly applicable to the manuscript and thus remains. The authors thank the reviewer for the suggestion of a summary table which has been added.**

**Clarification of changes:** Added an instrument summary table to section 2.

- 6) 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). **Appendices have been reviewed and revised throughout.**
- 7) Line 242: 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. **The measurement ceiling has been removed and the sensitivity in line 242 has been clarified as follows: “The sensitivity of the NO<sub>2</sub>CAPS was empirically derived to be  $0.17 \pm 0.14\sigma$  ppbv (during a separate ground-based zero test, where data is also averaged over 10s intervals).”**
- 8) **Comment:** Line 234: Figure 1 and Figure 2 do not seem to be mentioned in the text. Please correct. **Mention of fig 1 and fig 2 have been added to sec. 2 and 2.1, respectively.**
- 9) 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. The authors feel figure 3 is informative to the baseline correction discussed in the main text. However, based on this feedback, **the figure has been revised and figure text has been updated accordingly**, as shown below.



*Figure 3: Top: timeseries of raw (uncorrected) and processed (corrected) NO<sub>2</sub> 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: NO<sub>2</sub> instrument 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.*

- 10) Line 259: Does it mean that the first 7 flights do not have any valid O<sub>3</sub> data? **Yes. The text has been changed to: "...following some initial issues with negative calculated ozone values during MOASA measurements (impacting the first 7 flights which do not have valid ozone data), the Dewlines were regularly replaced...."**
- 11) 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."* **This has been revised as follows: "Vertical bars show the error in response for each bead size, derived by calculating the standard error of the mean for 15 second segments of each bead response, and then taking the mean of these values."**
- 12) 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. **Appendices have been reviewed and revised throughout.**

- 13) 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. Regarding uncertainties, this has been clarified in the **text which has been changed to: "...*it is subject to potential uncertainties, such as assumptions of aerosol homogeneity and sphericity, that caution against its use as an accurate measure of the true ambient particle IOR (Frie and Bahreini, 2021)..*"**. Please refer to response to Comment #2 regarding moving essential details to the supplement.
- 14) Line 415: Is the Appendix D the table D1? **Appendices have been reviewed and revised throughout.**
- 15) Line 422 "were" instead of "are" **This has been changed.**
- 16) Line 431: Why is Figure 8 mentioned before Figure 7? Please change the numbering. **Done.**
- 17) 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. **Please refer to response to Comment #2 regarding moving essential details to the supplement. Clarification of changes: None.**
- 18) Line 517: I guess that Appendix 7 is the Table F1. See comments above related to the naming and numbering of Appendices. **Appendices have been reviewed and revised throughout.**
- 19) 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. **The text has been updated to: "...Measured values in each dataset were split into groups of equal size, with sizes corresponding to equivalent ground distances ( $d_{int}$ ) ranging from 0.42 km to 17 km, in 0.085 km (1 second) intervals (where, for this study which focuses on average variability over a campaign, an airspeed of 85 m/s is assumed to be equivalent to 0.085 km straight-line distance at ground level)..."**
- 20) 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. **We thank the reviewer for their comment. This section has been revised and shortened. The flight mean variability has been removed** as the authors feel it did not add to the manuscript. **Additional references and text have been added** to the section to substantiate the analysis and emphasize the potential use of high-horizontally resolved data to study the natural and often sub-grid variability of pollutants. **Clarification of changes: Section 4.1 has been revised and shortened. Original Figure 10 has been removed.**
- 21) 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? Please refer to the previous comment.

**Clarification of changes: None.**

22) Line 592: "... *no statistical post processing has been applied*". How will this affect the results obtained? Please clarify what you expect. **Changed text to "... *no routine statistical post-processing (SPP, which uses surface level observations to apply corrections to the surface model level only) has been applied to the data. Given this study focuses on those data above the surface level, the omission of the SSP has no impact on the evaluation.*"**

23) 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. **Agreed – this section has been reformatted as suggested, whereby flights are discussed individually and summarised at the end.**

**Clarification of changes: Original 4.3 has been revised and split into 3 sections: 4.3.1 Flight M270, 4.3.2: Flight M296 and 4.3.3 Summary.**

24) 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.

**The following has been added to the respective sections in 4.3:**

**Flight M270: " *The variability observed is poorly represented by the coarse resolution model. Variation in the AQUM model data is largely caused by changing from one grid box to the other and ozone shows a typically smooth gradient between model grid boxes. We note that in this case the stacked flight transects only cross a very small number of model cells (3 or 4) in the horizontal, which may be accountable for the low model variability seen here.*"**

**Flight M296: "...unlike flight M270, the observations and model show similar variability. This is likely due to the flight track crossing a larger number of model cells which encompass more model predictions, and may also be due to the model capturing more variability for this case.."**

**And in the summary for this section: " *Variability in modelled ozone appears to be dependent on the number grid boxes encompassed by the flight track. It is expected that ozone concentration in higher resolution models (>12km) will better match variation in the airborne observational data, as model resolution moves towards natural scale variability.*"**

25) 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.

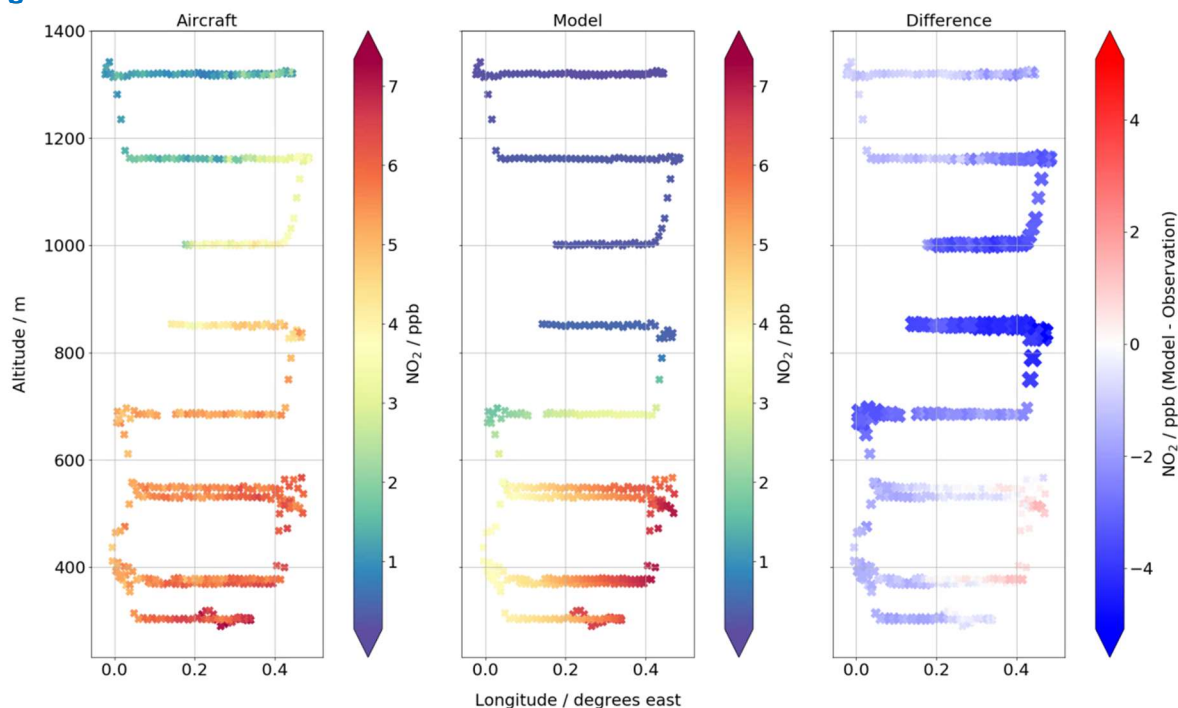
Similarly, it should be shown the effect of replacing the modelled by the observed BL height. **These more in-depth analysis is beyond the scope of the paper and is intended for future work - please refer to author response to Comment #2.**

**Clarification of changes: None.**

26) Line 613: "...shows "the" comparison" **Amended.**

27) 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. The authors feel the presentation of the data (latitude by altitude) provides good representation of spatial change with altitude. To better highlight the largest difference above 650m, **the colour-scale of the difference plot has been revised, the marker size now increases with divergence away from zero, and an additional description has been added to the figure text.**



**Figure 17:** Longitude-altitude plot of  $\text{NO}_2$  concentration for vertically stacked transects during flight M270 on 15<sup>th</sup> September 2020. The left-hand figure shows the aircraft data, the middle figure shows the model data, and the right-hand figure shows the difference between the model and aircraft, where opacity and thickness increase as the difference diverges away from zero. Data averaged over 10 second intervals.

**Clarification of changes:** Figure 17 (originally figure 14) has been revised replaced and the figure text updated.

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

**The wording has been changed to: "...This indicates a potential under-prediction in boundary layer height that may be responsible for the poor prediction of  $\text{NO}_2$  at elevated altitudes ..."**

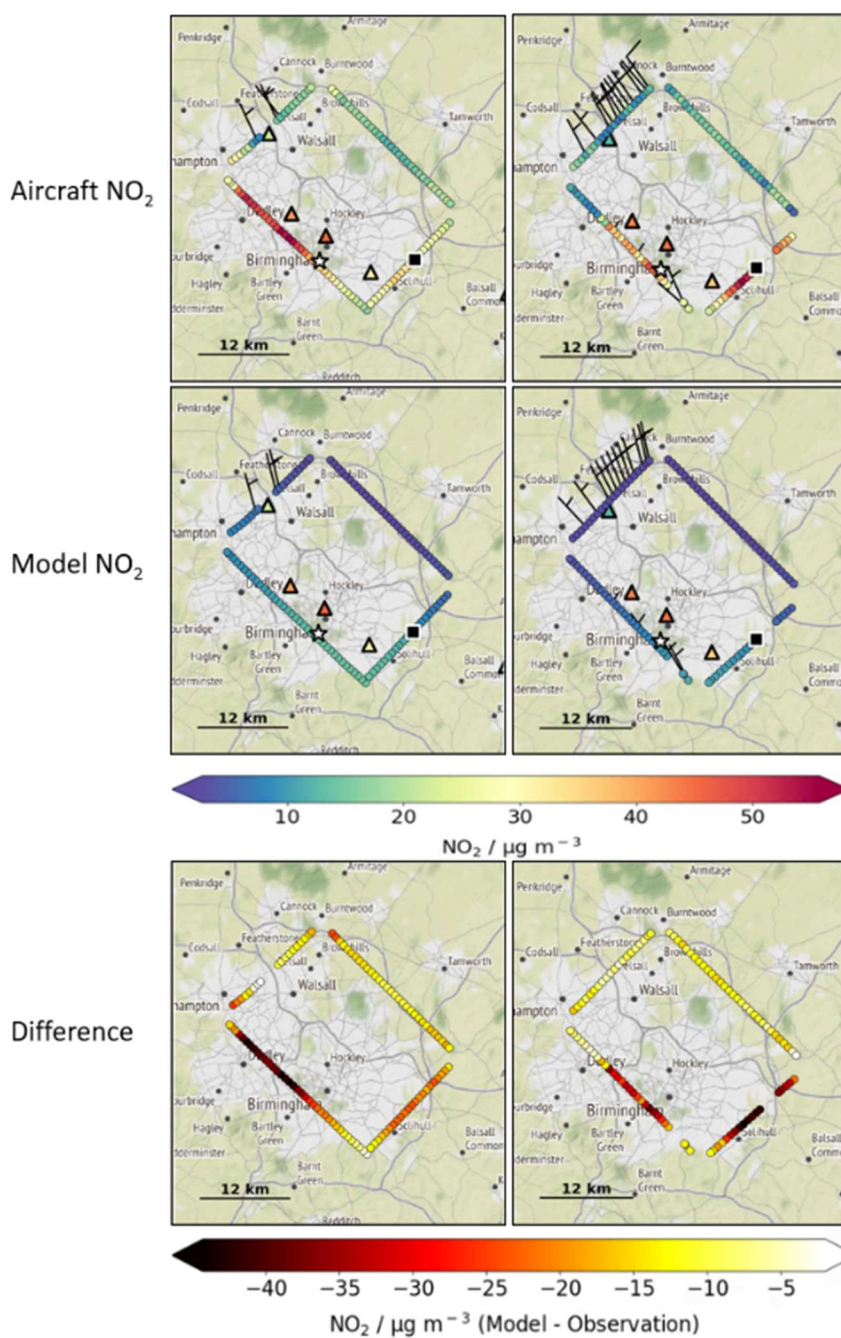
The whole of flight M296 is within the observed BL. Analysis of other flights is beyond the scope of this paper but is hoped to be addressed in future work.

- 29) 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)? The authors believe this comment refers to: "*This comparison indicates the value of evaluating model performance throughout the atmospheric column and suggests that the good agreement of  $\text{NO}_2$  seen at the surface may in this case have been somewhat fortuitous.*". This section has been revised, restructured, and reworded and – given that (in line with the reviewer’s comments) the paper does not strictly include model performance at the surface - it has been removed.

**Clarification of changes:** None.

- 30) Line 625: Please change the colour scale for the difference plot. It is impossible to see any difference by this large range. **This whole figure, including the colour scale for the difference plot has been revised** as per the below. Please note the

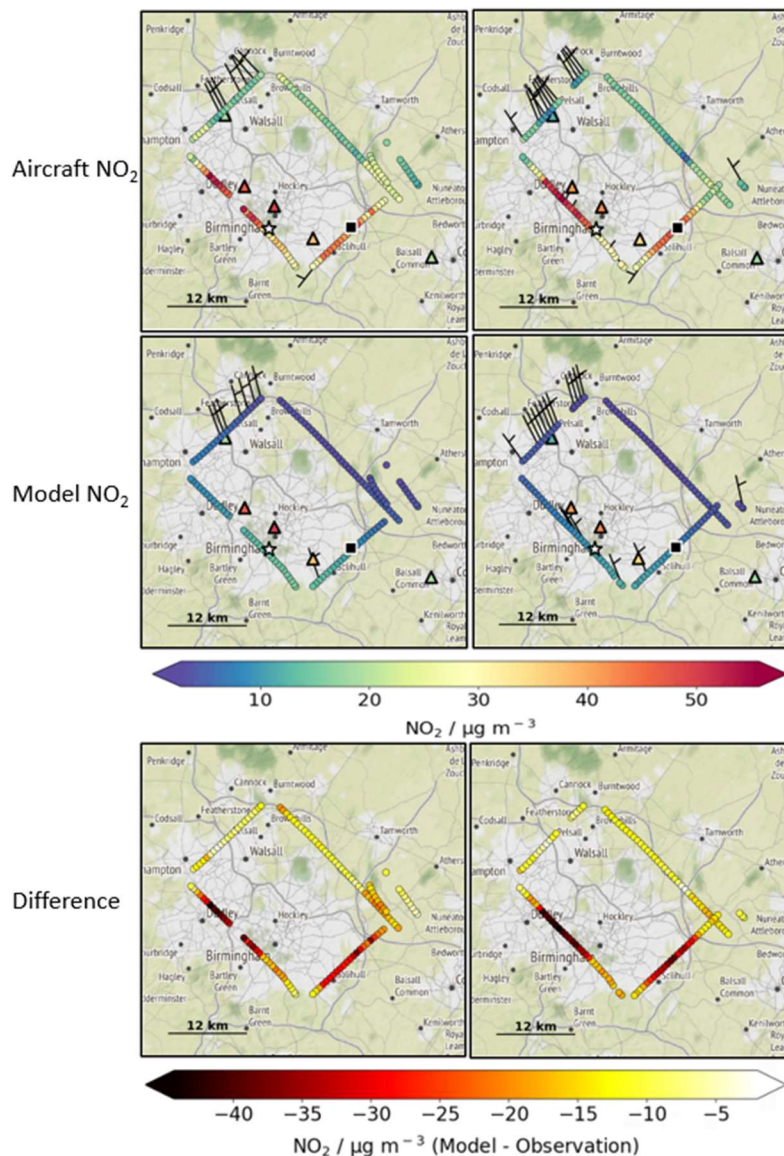
corrected ground-based concentrations which are generally in agreement with the airborne data.



**Figure 18:** Aircraft flight tracks coloured by NO<sub>2</sub> concentration ( $\mu\text{g/m}^3$ ) for the first (left, 11:23 to 11:43) and fourth (right, 12:33 to 12:52) circuit, at altitudes of 423 and 657 metres, respectively, around Birmingham during flight M296 on 1<sup>st</sup> July 2021. Top row shows the aircraft data, middle row shows the model data and bottom row shows the difference between the model and observations. Observation data is from straight and wings level transects and all data is averaged over 10 second intervals. Wind bars are only shown where the observed wind components exceed the measurement uncertainty. Data in triangles is the hourly surface level AURN NO<sub>2</sub> concentration for the circuit. Stars/squares show the location of the Birmingham supersite/airport, respectively. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

**Clarification of changes:** Figure 18 was originally figure 15. The figure text has also been updated.

31) Line 627: It would be informative to see the Circuits 2 and 3 in the supplement. Please include them. **Circuits 2 and 3 and figure text (below) have been added as appendix D, as per below:**



**Figure D1:** Aircraft flight tracks coloured by  $\text{NO}_2$  concentration ( $\mu\text{g}/\text{m}^3$ ) for the second (left, 11:43:00 to 12:10:00 and third (right, 12:10:00 to 12:33) circuit, at altitudes of 511 and 573 metres, respectively, around Birmingham during flight M296 on 1<sup>st</sup> July 2021. Top row shows the aircraft data, middle row shows the model data and bottom row shows the difference between the model and observations. Observation data is from straight and wings level transects and all data is averaged over 10 second intervals. Wind barbs are only shown where the observed wind components exceed the measurement uncertainty. Data in triangles is the hourly surface level AURN  $\text{NO}_2$  concentration for the circuit. Stars/squares show the location of the Birmingham supersite/airport, respectively. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL



- 32) 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? **The following text has been added** *“In consonance with AQUM, light north-westerly winds (0 < 5 knots) associated with the high-pressure system are observed in all circuits. These slack winds (equivalent to a maximum velocity south-eastward at 9.26 km per hour) likely pushed the plume (which is seen in the ground data to be present east of the flight track) south-eastward, accounting for the shift in the observed plume with altitude and time (approximately 1 hour between the first and final circuits). The proximity of the plume to Birmingham airport is also of note in run 4.”*
- 33) Line 636: How is the comparison between O<sub>3</sub> modelled and observed in those circuits? Refer to response to comment #32. Regarding the revised figure, **the following has been added:** *“As expected, given that NO<sub>2</sub> is photochemically split during the formation of O<sub>3</sub>, observed O<sub>3</sub> aloft (not shown) is inverse to the NO<sub>2</sub> observations, and shows a reduction of approx. 20-30 µgm<sup>3</sup> at the plume locations at all altitudes.”*
- 34) 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? **Please refer to response to comment #32.**  
**Clarification of changes: None.**
- 35) Line 646-647: Is that the case for M296? **Please refer to response to comment #32.**  
**Clarification of changes: None.**
- 36) 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” **Please refer to response to comment #2.** This figure, and associated analysis, has been removed.  
**Clarification of changes: None.**
- 37) 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. **Please refer to response to comment #2.** This figure, and associated analysis, has been removed.  
**Clarification of changes: None.**
- 38) Line 663: Please rephrase. “reform” does seem to indicate a null cycle **Please refer to response to comment #2.** This figure, and associated analysis, has been removed.
- 39) **Clarification of changes: None.**
- 40) 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. **Please refer to response to comment #2.** This figure, and associated analysis, has been removed.  
**Clarification of changes: None.**
- 41) 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?

Please refer to response to comment #2. This figure, and associated analysis, has been removed.

**Clarification of changes: None.**

- 42) 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. Please refer to response to comment #2. This figure, and associated analysis, has been removed.  
**Clarification of changes: None.**
- 43) 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. Please refer to response to comment #2. This figure, and associated analysis, has been removed.  
**Clarification of changes: None.**
- 44) 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. Please refer to response to comment #2. This figure, and associated analysis, has been removed.  
**Clarification of changes: None.**
- 45) 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. Please refer to response to comment #2. This figure, and associated analysis, has been removed.  
**Clarification of changes: None.**
- 46) Line 838: Please correct the title of the reference. **This reference is no longer included.**

## REVIEWER #2

1. 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).  
**Clarification of changes: Introduction revised.**
2. 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).”
3. 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.

**Clarification of changes: Amendments to abstract and introduction.**

4. 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<sub>2</sub>) bias at rural/urban sites and small negative biases in PM<sub>2.5</sub>. 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<sub>2.5</sub>. 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).***

1. In Sect. 1.1 (Impact of COVID-19) incorporate and discuss results from other studies in Europe related to O<sub>3</sub> and NO<sub>2</sub> 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.

**Clarification of changes: Original section 4.4 has been removed. The whole of section 4 has been re-ordered as per below and has a new addition to replace the removed section.**

- **4.1 The spatial scales of pollutant variability**
- **4.2 Ground-based and airborne observation comparison using long term observations over London**
- **4.3 Preliminary model evaluation**

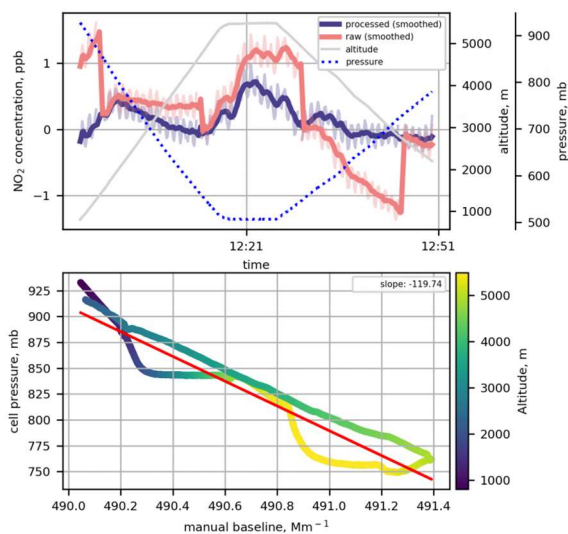
**The manuscript title has been revised to: “Long-term airborne measurements of pollutants over the UK, to support air quality model development and evaluation”**

2. In Sect. 2.1 (Instrumentation – general setup) add a table listing the instrumentation, technique, precision, and references. **Agreed, this has been added.**
3. 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 addressed in future work.**

**Clarification of changes: None.**

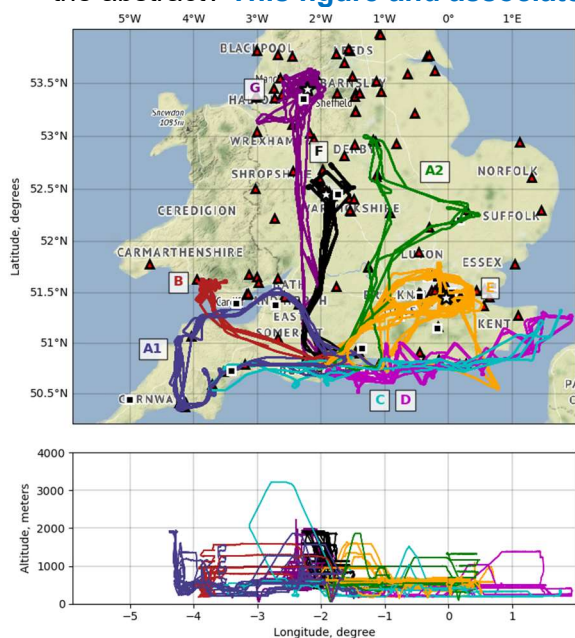
4. On Page 19 (line 676-678) you write: “We define the pre-lockdown period as 26th March 2018 to 25th March 2020 and the post-lockdown period as 26th March 2020 to 25th March 2022, where comparing like-for-like months pre- and post- lockdown minimises the impact of seasonality on the comparison.” □ would 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.](#)  
**Clarification of changes: None**
5. In general, it is recommended to give mean NO<sub>2</sub> and O<sub>3</sub> values for all flights in a table for a better overview of the airborne results. [Please refer to author response to comment #2.](#)  
**Clarification of changes: None**
6. 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.](#)  
**Clarification of changes: Section 4 has been revised and includes a summary to each of the three sub-sections.**  
Page 10, line 359: Appendix 5 not available. **Appendices have been reviewed and revised accordingly.**
7. Page 12, line 413: gcm<sup>3</sup> > g/cm<sup>3</sup>. **Changed.**
8. 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**
9. Page 14, line 517: Appendix 5 not available. **Appendices have been reviewed and revised accordingly.**
10. Page 15, line 530: Add from who. **Added “... by contacting the author.”**
11. 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.](#)
12. Page 16, line 582: Sect. 4.2 is missing. **Well spotted - corrected.**
13. 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 \(see comment 1D\).](#)  
**Clarification of changes: Sentence updated.**
14. Page 17, line 633: “The AQUM model shows little variation and comparatively low NO<sub>2</sub> 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.](#)  
**Clarification of changes: None**
15. Page 20, line 732: Add from who. **Replaced with alternative text indicating new archive.**
16. 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<sub>2.5</sub> ) in the United Kingdom, 2012 **Updated**
  - DEFRA: Clean air strategy 2019., 2019 **Updated**
  - Ecotech: Aurora3000 Integrating Nephelometer with backscatter, 2009. **Updated**
17. Page 25: Give Appendix text first after Figs. 1-16. **Done.**  
**Clarification of changes: Appendices A-E now contain text in addition to their original figures. Appendix D has been added (text and figure).**
18. Page 26, line 986: Subtitle “Case study” somehow misplaced. **Amended.**

19. Page 29, Fig. 3: Legend covers data, move it. Blue straight line better visible in red. **The figure 3 + text has been updated as follows:**



**Figure 3: Top: timeseries of raw (uncorrected) and processed (corrected) NO<sub>2</sub> 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.**

20. 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.**



21. Page 34, Fig. 11: In the figure text “0.85” is mentioned twice. **Amended.**

22. Page 43, line 1163: gcm3 > g/cm3. **Amended.**

23. Page 43, Table E1: Add NO<sub>2</sub> to the header of column 3 and O<sub>3</sub> to the header of column 4. Please refer to author response to comment #3, which advises that this section has been removed.

**Clarification of changes: None**