Long-term airborne measurements of pollutants over the UK, including during the COVID-19 pandemic, to support air quality model development and evaluation by Mynard et al.

General comments:

This manuscript describes long-term measurements in the UK of a variety of atmospheric trace species carried out with a Cessna-421 aircraft from June 2019 to April 2022 (including the COVID-19 impacted period). The measurements mainly covered altitudes between the ground and the top of the boundary layer. Specific flight patterns, repeated sorties, were flown sampling both rural and urban regional background pollution in addition to polluted city plumes. The objective was to sample an extended data set to be used for evaluations of the regional air quality model AQUM of the UK Met Office. Here the gained airborne data set is described and first intercomparisons with the model are presented.

The paper addresses scientific questions relevant to the scope of AMT (technical description of instrumentation, novel measurement, modelling).

The manuscript is well-written and generally logical structured. Manuscript is written by native speakers, language fluent and precise, no improvements needed. The number of figures and tables in comparison to the text is appropriate. In general, the figures are of good quality. Proper credit is not always given to related work. The conclusions and statements are rather sparse, based on first results from two case studies. Details on suggestions for improvements on these topics are given in the specific comments.

For the reasons mentioned above and below the paper is appropriate for publication in AMT after a major revision described below.

Specific comments:

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. 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. 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.

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.

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 Sect. 2.1 (Instrumentation – general setup) add a table listing the instrumentation, technique, precision, and references.

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?

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 (pre-lockdown, main lockdown, post-lockdown?

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.

After Sect. 4 and before Sect. 5, a section on discussion of the results is missing.

Minor comments and technical corrections:

Page 10, line 359:
Appendix 5 not available.

Page 12, line 413:
gcm3 → g/cm3

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.

Page 14, line 517:
Appendix 5 not available.

Page 15, line 530:
Add from who.

Page 15, line 547:
Why data only used until July 2021 (44 flights) and not all available data until April 2022 (63 flights)?

Page 16, line 582:
Sect. 4.2 is missing.

Page 17, line 607:
"Savage et al. (2013) also reported biases during a ground-site AQUM comparison." → discuss in the manuscript

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? Discuss the reasons for the differences.

Page 20, line 732:
Add from who.

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

- DEFRA: Clean air strategy 2019., 2019

- Ecotech: Aurora3000 Integrating Nephelometer with backscatter, 2009.

Page 25:
Give Appendix text first after Figs. 1-16.

Page 26, line 986:
Subtitle "Case study" somehow misplaced.

Page 29, Fig. 3:
Legend covers data, move it.
Blue straight line better visible in red.

Page 32, Fig. 8:
Here only 45 flights shown, why not all 63 flights shown as described in the abstract?

Page 34, Fig. 11:
In the figure text "0.85" is mentioned twice.

Page 43, line 1163:
gcm3 → g/cm3

Page 43, Table E1:
Add NO2 to the header of column 3 and O3 to the header of column 4.