Response to Reviewers

Electrochemical sensors onboard a Zeppelin NT: In-flight evaluation of low-cost trace gas measurements

---

Referee comments are in **black** and authors responses are in **blue**

Reviewer #1:

Summary

"Electrochemical sensors onboard a Zeppelin NT: In-flight evaluation of low-cost trace gas measurements" evaluates the performance of a suite of sensors installed in a hatch box within the bottom of an airborne platform. The focus here is on NO and NO2 measured by Alphasense electrochemical sensors. Six units are flown together underneath the Zeppelin and intercompared. Other aspects, including results from a reference mid-infrared MIRO instrument, were described in a related paper by the authors. Many papers have been published in recent years on low-cost gas and particulate matter sensing, along with their calibration and correction for spurious environmental dependencies. In my opinion, the manuscript is well written and scientifically sound, although the scope is fairly limited. The most novel aspect is airborne deployment in Zeppelin flights, which took place in Germany.

We thank the reviewer for the positive feedback and the helpful comments. Response to each comment is provided below.

Main

My main comment is that it feels like the reference MIRO data from the flights is underutilized in terms of validating the performance of the electrochemical sensors, especially since laboratory test data is not included. Some thoughts on this point:

- While the intercomparison of the six setups in Figure 4 is interesting, is the conclusion about setup #2 performing best also supported by comparing against MIRO MGA? What do those results look like?

This comparison is shown in Figure 6. We now added more information to the caption of Figure 6 and a reference in the text to clarify that these plots show the comparison of setup #2 with the MIRO MGA. Moreover, we now added Figure S10 in the supplement that shows the linear regression of each sensor relative to the MIRO MGA. These results highlight the consistent agreement of all sensors to the MIRO MGA with the slopes of the linear fits for NOx agreeing within 10 %. We now clarify in the text why we choose setup #2 for further evaluation of the sensors as follows:

"The clocks on each of the six sensor setups were manually pre-set, therefore time synchronization was not ensured. In the first step we chose a *master setup*, here setup #2 that was operational throughout all flights with a data coverage of 99 %."

- Corrections were derived to remove dependences on T, AUX, and dRH/dt without using MIRO, and compared against manufacturer recommended corrections. This is advantageous in avoiding requiring use of a reference instrument. However, can MIRO be used independently here to evaluate how well this correction approach works?

Figure 6 (d) for setup #2 and Figure S10 for all setups now show the linear regressions of the corrected data vs. MIRO. The R^2 range from 0.75 to 0.88 which reflects the good performance of the correction method. We further clarify this in the main text as mentioned in the comment above.

- What can be said about stability of the ECS sensors and derived calibration during or between flights.

The measurements performed in this paper range from end of April to mid-June. When accounting for dependencies to T, AUX, dRH/dt the correlation of the ECS sensors to the MIRO reference system are consistent highlighting that no major instabilities were present during these measurements (Fig. 6 and S10). Wei et al. (2018) highlighted that for a 2 month period, like in our study, ECS sensors were stable with drifts < 2 ppb/month. We added a few sentences and references in the main text (Sect. 3.2.3) highlighting the importance of long term measurements to further evaluate long term instabilities.

"Besides the above direct influences, there is also the possibility of sensor drifts, i.e., a change of the sensor signal with time. Wei et al. (2018) estimated a possible drift of < 2 ppb/month whereas Mead et al. (2013) state that the sensitivity of the sensors remained unchanged over an 11-month measurement period. For our deployment duration of 1.5 months, sensor drifts are therefore expected to be within the uncertainty of the measurements which is also reflected by the good agreement of the ECS and the MIRO in Fig. 6 and S10. Furthermore, Fig. S11 shows the timeseries of all sensors during different flight days in May and June to evaluate the influence of such sensor drifts. The consistent correlation of all setups to the MIRO highlights the stability of the sensors during this study. However, we promote the need for controlled laboratory measurements in the future to evaluate long-term influences on the stability of the ECS signals including sensor drifts."

Minor

I am not completely sure what data was used to generate the figures. On L265 it talks about only showing setup #2. L160 talks about excluding #4, #6, and partly #5. Which sensor setups are included in Fig 3, 5-7?

We split up section 2.3.1 and added a new section 2.3.2 to clarify that only the temperature and relative humidity sensors of setups #4, #5 and #6 were excluded. We also added the setup number in all corresponding captions for clarification.

Similarly, Figures 4 and 6 seem to be showing aggregate data over all the flights. This could be clearer. How many individual flights and hours of data are included? Is any data excluded?

We added more information to the captions of Fig. 4-6. This includes the hours of measurements (≈ 286 hours total, ≈75 hours in-flight), the measurement height, and the number of data points. Additionally, we extended Figure 6 by adding a colorbar that shows the number of flight measurements performed per bin.

Figure 1(a) what is the height and width of the hatch box?

We added the following in section 2.1: “The hatch box dimensions are 738×538×162 (length × width × height in mm).”.

L160 “From this correlation analysis it is evident that the sensors of setups #4, #6 and partly #5 provide erroneous data.”This is not obvious to me. Figure S2 in particular is referenced in the sentence before, but is hard to read both in terms of the font size within the figure and having a fairly brief figure caption.

We agree and thank the reviewer for pointing that out. We now replaced the figure, improved the caption, and added section 2.3.2 to further describe the selection process. Fig. S4 now shows the differences between each temperature sensors to the rest and highlights that sensors #4, #5, and #6 deviate more from the others.

Figure 6; is top panel with y-axis label ‘Accuracy / ppb’ showing the +/- 2 standard deviations as shown in the bottom panel? This should be clearer.

The standard deviations $\pm 2\sigma$ reflect the precisions of the measurements. We added the following to the caption to better define the accuracy: ", i.e., the absolute discrepancies between sensor and MIRO MGA data".

“Evidently, with the manufacturer’s correction, amount fractions in the low ppb range cannot be quantified (Fig. 6) predominantly due to the high offset of -19.76 ppb.”

I don’t understand this point since an offset affects low ppb as well as high ppb measurements equally and does not determine the sensitivity of the measurement. It could be an issue if readings are filtered to be above zero.

That is a valid point. Rephrased the sentence to “Evidently, with the manufacturer’s correction, amount fractions (Fig. 6 (a)) cannot be accurately quantified predominantly due to the the high offset of -19.76 ppb.”

Acknowledgements: WRF is not directly mentioned in the main text, although perhaps it was run to generate the wind field in Figure 7. EURAD-IM is not mentioned.

We added “EURAD-IM, WRF” to the caption of Fig. 7.