Comment on amt-2020-483
Anonymous Referee #4

This work depicts the results of a harsh environment long term colocation experiment focused on performance evaluation of Ozone LCS versus a high precision and accuracy regulatory grade instrumentation. Surprisingly accurate results are obtained with a very simple open source architecture based on redundant EC sensors targeted to O3 using lab calibration. Field calibration allowed for further amelioration. Their experimental design allowed them also to

a) detect drift during two consecutive lab calibration in May and July 2018.

b) evaluate the possibility of calibration transfer correcting the response from one sensor with coefficient derived from other sensors calibration.

c) finally, reliability of the LCS sensor outperformed the regulatory grade one, which failed several times due to typical operational and technical issues.

I enjoyed the paper, it is well organized and the technical content is certainly relevant with validation encompassing several seasons.

However, it could be important to highlight that the selected sensor are actually sensitive to NO2 (also) with a sensitivity that is very similar to the one expressed towards the intended target, O3. Actually Alphasense suggests to couple this sensor with an NO2 targeted one and to subtract the NO2 estimate from the NO2+O3 estimation provided by the original sensor. From author's results, influences from low NO2 concentrations at high altitude is clearly found to be negligible.

In the paper, the description of the auxiliary electrode working principle is imprecise. Actually, auxiliary electrodes in Alphasense sensors are just architectural twins of working electrodes which are screened from the atmosphere. As a result they react only to temperature providing a way to correct from temperature dependance subtracting the response elicited by temperature itself on both electrodes. Unfortunately geometry and fabrication variance does not allow for building an exact replica of the working electrode so correction is not exactly accurate and field calibration is still necessary.

It could be interesting to the discuss the cost and the long term accuracy of the field
calibration which in the proposed version needs to use two months in different extreme seasons to obtain the presented results. Recent papers have analyzed the robustness to changing environment of field calibration, which is to say that they lose accuracy when concept drift arise. Is the relative stability of the alpine environment beneficial for field calibration robustness?

Moreover calibration transfer “absolute” results are very interesting but they could be slightly overoptimistic from the numerical point of view since, in my understanding, authors are using the entire dataset to derive the calibration coefficients and then they validate the performance of the other so calibrated LCS in the very same conditions in which the calibration have been derived. If I am correct I would like this to be discussed in the paper.