We would like to thank the reviewer for the constructive comments. We have tried to address these comments in the attached response document, in the manuscript and in the code. Reviewer comments are reproduced in black, our responses are in blue.

General Comments:

This manuscript details the calibration process of low-cost metal oxide NO2 and O3 sensors. The authors evaluated the performances of several univariate, multivariate, linear, and non-linear calibration models. For these models the authors also analyzed the impact of individual predictors on model performance. The authors reccommended using multiple covariates in multiple regression models and to analyze the importance of the features used. Additionally, that machine learning models can greatly improve accuracy but have a harder time on data outside the calibration dataset.

Specific Comments:

At times the novelty of the approach seems to be overstated. Line 67 and 298 talk about the use of internal temperature as a calibration factor. Off the shelf sensors such as the Clarity Node S (measures NO2 with an electrochemical cell) use RH and internal temperature to adjust their NO2 readings). Additionally, in Line 221 you state that there is no statisical difference between using internal or external temperature. On line 298 you reference Figure 4 to explain why internal temperature was chosen but you do not show the same analysis for external temperature or for NO2.

Thanks for drawing attention to this aspect. As referenced in the study by Miech, Jason A., et al., (2021), the Clarity Node S adopts an electrochemical cell to measure NO2 levels. Notably, the NO2 measurement process is particularly intriguing, as the Alphasense NO2-A43F electrochemical cell comes equipped with a crucial ozone filter at its front end.
To further enhance precision, the device accounts for internal relative humidity and internal temperature while adjusting the NO2 measurements. Despite the comprehensive approach to NO2 measurement, the study does not explicitly specify the specific sensor used to measure environmental variables.

In our study, the correlation matrix and Bland-Altman plots revealed a robust positive correlation between the external and internal temperatures, with a relatively constant delta of 8°C (due to the significant heat radiation from the MOSs, causing the internal temperatures to be significantly higher than ambient conditions). The strong multicollinearity suggests that using both temperatures together would not be useful. Therefore, it would be appropriate to interchangeably use either of the temperatures.

As pointed out by the reviewer, upon examining Table 2 we observed a small yet minimal difference in using the internal temperature. However, we proceeded with the analysis using the internal temperature based on insights from the literature, such as the study by Schmitz, Seán, et al. (2021). This research emphasizes the importance of use internal temperature and relative humidity as essential factors because they more accurately represent normal MOS operating conditions.

In conclusion, we recognized the multicollinearity issue and opted to use only one of the temperatures in our analysis. Additionally, the use of external temperature as depicted in the figure below (fig. 1), does not lead to any definitive conclusions beyond what has already been discussed. Consequently, the highlighted paragraphs, as pointed out by the reviewer, have been revised to enhance clarity and improve its contribution to the overall analysis.

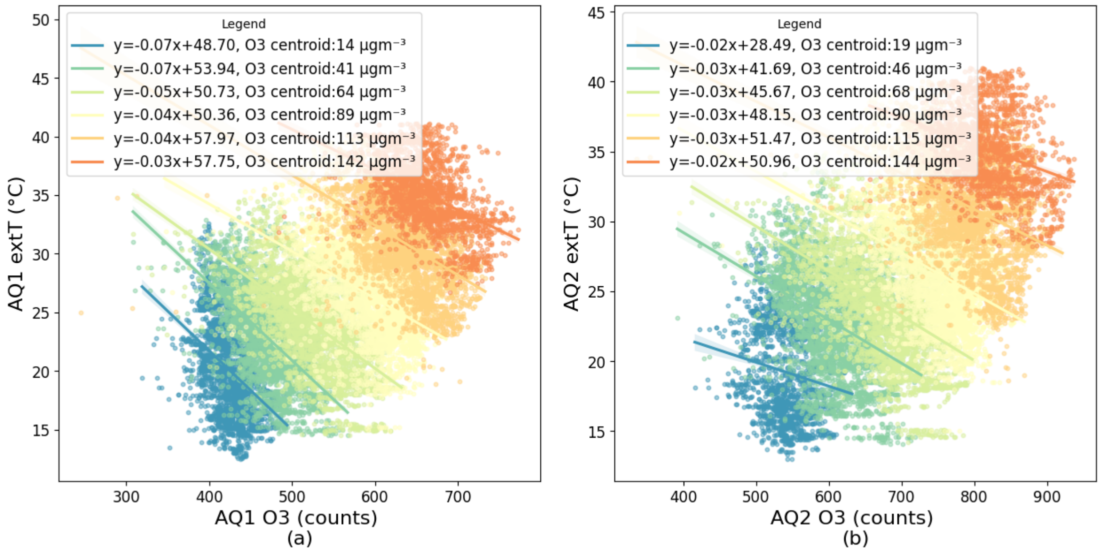

Fig. 1: Relationship between O3 raw data and extT in K-means cluster for AQ1 (a) and AQ2(b) stations. Regression lines were fitted to each cluster.

References

- Miech, J. A., Stanton, L., Gao, M., Micalizzi, P., Uebelherr, J., Herckes, P., & Fraser, M. P. (2021). Calibration of low-cost no2 sensors through environmental factor correction. Toxics, 9(11), 281. https://doi.org/10.3390/toxics9110281

- Schmitz, S., Towers, S., Villena, G., Caseiro, A., Wegener, R., Klemp, D., Langer, I., Meier, F., and von Schneidemesser, E. (2021). Unravelling a black box: an open-source methodology for the field calibration of small air quality sensors. Atmospheric Measurement Techniques, 14(11), 7221-7241. https://doi.org/10.5194/amt-14-7221-2021

The last paragraphs of introduction are then restructure on LL67-73 as:

These goals have been pursued by using ten among parametric, non-parametric univariate and multiple algorithms. Additionally, the investigation focused on delving deeper into the influence of internal temperature on LC sensors. To ensure comprehensive analysis, the covariate set for the multiple models was expanded to incorporate other essential factors such as humidity and gaseous interference compounds.

The paragraph of Discussion are then restructure on LL298-230 as:

Moreover, taking into account the observed multicollinearity issue between temperatures and the slightly higher mean accuracy, as well as the lower mean RMSE observed when using the internal one, the study drew upon insights from existing literature to identify the most suitable set of covariates (e.g., Schmitz, Seán et al., 2021; Miech, Jason A. et al., 2021). As a result, the inclusion of internal temperature as a significant factor was given priority, as it offers a more accurate representation of the operating conditions of the MOSs within the system. This approach was also adopted to tackle potential challenges in the board's analog-to-digital converter circuit.

Section 2.3: Please include more information on the sensor pre-deployment calibration with the HORIBA instruments. It is unclear whether this calibration was conducted indoors or outdoors or the spatial relationship between the AQ stations and the HORIBA instruments. If indoors please explain the lab environment where testing occurred.

As mentioned by the reviewer, it is unclear whether the calibration was conducted indoors or outdoors or the spatial relationship between the AQ stations and the HORIBA instruments. To address this, setup details are explained below.
The AQs are installed outdoors at the same height, securely mounted on a dedicated rack as shown in Figure 2a. Meanwhile, the HORIBA instruments are positioned indoors within the laboratory setting. To ensure an accurate representation of outdoor air conditions, two sampling probes, each approximately two meters in length and equipped with rain covers, are employed. These probes collect the outside air and channel it directly to the reference instruments (as depicted in Figure 2b). This setup enables us to sample and compare the air quality data collected by the AQs with the measurements obtained by the HORIBA instruments. For completeness we included Figure 2 below.

Fig. 2: AQs on the left (a); Horiba setup on the right (b).

<u>The paragraph of Section 2.3 are then restructure on LL298-230 as:</u>

As detailed in Table S1 of the Supplementary material, Pre–deployment calibration of AQ1 and AQ2 stations against HORIBA analyzers was performed at CNR-IBE headquarters in Florence, Italy (43°47’52” N, 11°11’ E, Figure 1). The AQ stations were mounted on a dedicated outdoor rack, while the HORIBA instruments were placed indoors in a laboratory setting. For outdoor air pollution sampling, approximately two–meter–long sampling probes were employed to collect outside air and channel it directly to each of the reference instruments.

Line 196: Do you have any explanation for why more data was withdrawn from AQ2 compared to AQ1?

We appreciate the reviewer for bringing attention to this difference between the AQs. We identified that AQ1 had a 2% withdrawal due to the MOS NO2 sensor, while AQ2 had a 12% withdrawal, primarily from RH (7%) than from MOS NO2. It's worth mentioning that the RH sensor in AQ2 reached saturation (data > 99%) more frequently than in AQ1, impacting the overall data.

Figure 8a: Should this legend read "AQ1 O3 MLR" rather than NO2?

We thank the reviewer for pointing out this typo, we have fixed it in the manuscript.

While Figure S10 summarizes the NO2 and O3 concentrations across the validation period and Table 6 for the field validation it would also be useful to include a table detailing the historical environmental conditions of both the field validation and calibration period, such as RH and temperature. This could help support the points made in line 340, as when the environmental conditions differ between pre-deployment calibration and the deployment/validation period the MRF model may suffer.

Table S1 now includes environmental parameters, namely temperature and relative humidity. Additionally, for each process, we have specified the reference intervals, Horiba for the pre-deployment calibration and ARPAT for the field validation. We also corrected an error in the date interval of pre-deployment.

Line 331: Please re-word this sentence as the point is unclear

Thanks for the suggestion.

Line 331 it has been corrected as in LL331-333:

Ultimately, this resulted in robust LC performances outside the training conditions and the ability for easy adjustments to cope with changes in sensor performance over time.

Line 339: You mention global impacts of this analysis but provide no other information of how this work extends to beyond Italy.

Thanks for the suggestion. In the ongoing activity, the AIRQino LC stations are planned to be deployed outside Italy and also in extreme environmental conditions (Carotenuto et al., 2020). This will allow NO2 and O3 sensors to be tested under different meteorological and air pollution conditions. The conclusions now include some details of ongoing projects and perspective.

Reference

- Carotenuto, Federico et al., 2020 IOP Conf. Ser.: Earth Environ. Sci. 489 012022. https://doi.org/10.1088/1755-1315/489/1/012022

The conclusions now include some details of ongoing project and perspective, as follows (LL349-end):

A limitation of the present work is that the LC stations have been calibrated during a period not particularly long (70 days) and a typically summer one, thus when pollution levels are generally meaningful for O3, but they are not for NO2 concentrations. Indeed, conducting a pre–deployment calibration during a winter period, when NO2 concentrations are typically higher, would be a valuable addition to the study. This step would provide a more comprehensive understanding of the AQs validation performance under varying pollution conditions and help address the limitation of the current calibration period biased towards summer data. Moreover, conducting a similar validation outside of Italy, in regions with differing pollution and meteorological conditions would be of great interest. For this purpose, in the ongoing activity, the AIRQino LC stations are planned to be deployed outside Italy, such as in Nice and Aix–en–Provence (France), Barcelona (Spain), Budapest (Hungary), Tirana (Albania), and Niamey (Niger).

Furthermore, in the future, a new sensor for monitoring NO could hopefully be integrated into the LC stations and validated. As such, the combined monitoring of NO, NO2 and O3 concentrations and their daily and seasonal variability would allow a comprehensive pattern of the oxidant capacity of the atmosphere, particularly effective in southern Mediterranean countries such as Italy (Pancholi et al., 2018). In addition, once the AQ VOC sensor is validated, it will enable the monitoring of all O3 precursors (VOC and NOx). This comprehensive monitoring, combined with the application of SHapley Additive exPlanations (SHAP) method, will lead to a full characterization of photochemical pollution in various areas of interest, including urban, sub-urban, or rural regions. Moreover, portability of LC sensors makes them ideal devices for filling knowledge gaps in regions that are difficult to access such as the open sea. Mounted on buoys or ships, for example, LC sensors could collect the high O3 levels that typically occur over these areas in summer due to high solar activity and rather low mixing height combined with a lack of O3–consuming NO emissions.

References

- Pancholi, P., Kumar, A., Bikundia, D. S., & Chourasiya, S. (2018). An observation of seasonal and diurnal behavior of O3–NOx relationships and local/regional oxidant (OX= O3+ NO2) levels at a semi-arid urban site of western India. Sustainable Environment Research, 28(2), 79-89. https://doi.org/10.1016/j.serj.2017.11.001

- Li, W., Wang, Y., Liu, X., Soleimanian, E., Griggs, T., Flynn, J., & Walter, P. (2023). Understanding offshore high-ozone events during TRACER-AQ 2021 in Houston: Insights from WRF-CAMx photochemical modeling. EGUsphere, 2023, 1-21. https://doi.org/10.5194/egusphere-2023-1117.