First of all, we would like to thank the reviewer for reading the manuscript and for the valuable comments and suggestions for article improvements. The manuscript has been extensively revised. Due to the significant changes in the article, we have decided to modify the title of the article to better reflect its focus. New title "Measurement report: A complex street-level air quality observation campaign in the heavy traffic area utilizing the multivariate adaptive regression splines method for field calibration of low-cost sensors".

We believe that these changes will improve the readability of the article.

Listed below are our answers and the changes made to the manuscript according to the remarks and suggestions given by the reviewer. The comment of the reviewer is listed below, our responses are highlighted in blue italics.

Reviewer 2:

The topic of the manuscript is both interesting and relevant to ACP readers. However, significant revisions to the presentation of the methods and results are necessary before the validity of the findings can be assessed and their relevance fully evaluated.

We would like to thank the reviewer 2 for his valuable remarks and recommendations. The manuscript has undergone extensive revisions, including incorporation of most of the comments. We improved the description of the used methods and we moved some important results related to measurements with low-cost sensors to the article. On the contrary, most of the information related to the supplementary meteorological measurement has been moved to the Supplement. We strongly believe that the changes made have improved the readability of the article.

Abstract is somewhat confusing, more like a list of unrelated sentences and not a concise overview. Please rewrite.

The abstract has been completely rewritten. Please see the revised version of the manuscript.

Basics of MARS method (why, how, the elements of the model, meaning of splines) should be described as it cannot be considered as commonly known method.

The description of the MARS method has been modified in the methodology section in the revised version of the manuscript (section 2.3, lns 243-247 on p10), especially by adding the summary Table 3 (p11) with the design of the COR and COR2 MARS models. The detailed description of the MARS method principles has been added to the revised version of the Supplement to the section S2.3.1 (p3-4). The aspects why we used this method are mentioned in the introduction and in discussion (see lns 92-101 on p3-4 and lns 518-528 on p23) in the revised version of the manuscript.

The method description lacks sufficient detail to allow for replication of the analysis, which should be the standard level of detail. Conversely, the methods section includes numerical information that would be better suited to the results section. This misplacement makes the methods section harder to read. When transferring these details to the results section, explain the significance of the numbers (e.g., CVs, R^2) and the narrative they convey.

The methods section has been revised as recommended. The structure of the chapter has been changed for better clarity with a focus on methods used for the LCS network (methods related to meteorological measurements have been moved to the Supplement). In addition, the numerical information has been moved to the results section 3.1 in the revised manuscript.

An explanation of the contribution of the metrics used has been provided in the methodology section 2.3, lns 228-242 on p9: "Summary statistics included the coefficient of variation (CV) to express mean precision of LCS measurements during field measurements, along with mean, median, standard deviation (SD), and parameters derived from regression analyses: intercept (a), slope (b), coefficient of determination (R^2), Williamson-York regression parameters (a, b, using the maximum given RM, EM and LCS uncertainties; according to Cantrell, 2008), mean bias error (MBE) and root mean square error (RMSE)".

Statistical significance has been added to the correlation coefficients mentioned in the results in lns 306-307 on p12 (as all relationships presented were statistically significant at the $p<0.05$ level, they are not reported further).

It should be discussed whether linearity can be assumed in the data and whether the linear regression model is therefore appropriate. If linearity can be assumed I would recommend considering more robust fitting methods, as the ordinary least squares has been shown to be ill-suited for many types of atmospheric data (see e.g. Cantrell, 2008 and Mikkonen et al. 2019).

We compare the reference measurement method with the values measured by the sensor without correction (raw data) and the values corrected using the MARS method (corrected data). In accordance with the commonly used methodological approach, we perform a linear best fit of the scatter plot of the reference method versus the sensor measurement. However, we thank you for your valuable comment at this point, because we acknowledge that both methods (sensor and reference monitors) are subject to error. Therefore we have additionally calculated the Williamson-York Iterative Bivariate regression (see lns 236-238 on p9 in Sect. 2.3 in the revised version of the manuscript) and showed these results as the part of LCS summary statistics related to the initial and final field comparative measurement (see Tables S12-S15 and Tables S16-S19 in the revised version of the Supplement). The bivariate regression was calculated according to the recommended publication Cantrel (2008). The used values of variance were calculated from the maximum possible uncertainties specified for the measurement method (15 % for NO_2 and O_3 RM measurement, 25 % for PM_{10} and $PM_{2.5}$ EM measurement and 30 % for LCS measurements) since we do not have the exact (laboratory-verified) values of uncertainties, especially in case of LCSs.

Section 2.3.2 could also be improved by adding subtitles or breaking it into smaller, thematically organized segments for better readability.

We agree with this comment. Section 2 (Materials and methods) has undergone extensive changes and shortening. Newly, there are only three main subsections: 2.1 Study area and experimental design, 2.2 Technical specification of instruments used and measurement *methods, 2.3 Data processing and statistical analyses. We believe that changes made improved the readability of this section.*

Describe clearly in the main text what are the COR and COR2 methods as your results are strongly dependent on them. Show an example of a correction model in the main text and justify the form of the model. The meaning of the equations in Table S14 is not fully clear even for experienced statistician, let alone the average reader of the ACP.

Thank you for this valuable remark. The description of COR and COR2 method was modified in the text of method section 2.3 and improved by adding Table 3 with the design of the COR and COR2 MARS models. We believe that changes made will help readers to better understand the summary statistics of correction model outputs shown in Tables S2-S10 in the Supplement (i.e. Tables SS6-S14 in the original version of the Supplement), which are important in our opinion for demonstrating the performance of the individual correction models.

The manuscript leans too heavily on the massive Supplement. Reconsider which result tables and figures you are showing in the main text. If important results are shown only in 51-page supplement, it will not be found by readers.

You are correct, thank you for this important note. As already mentioned before, we moved some important results related to low-cost sensor measurements to the main text of the article, including added Figure 5 (14), which represents the quality of the sensor measurement before and after correction.

Minor remarks

There are two sections numbered as 2.3.2

We apologize for this typo in the original manuscript. This error is no longer repeated in the revised version.

Section 4.1.2: Comparing R2 values from different studies is not meaningful as the value can be calculated in numerous ways which are not necessary comparable.

Thank you for this remark. Coefficient of determination (R^2 resulting from linear regression between reference and sensor measurement) is one of the most common metrics used to express sensor measurement performance (before and after various corrections applied) in numerous studies (e.g. Vajs et al. 2021, Kumar and Sahu, 2021 or Borrego et al., 2016 and others). Therefore, we decided to leave R^2 based comparison with other studies in the discussion section. However, as part of the manuscript revision, we verified that the studies we compare with in the discussion used really R^2 from linear regression between the reference and sensor measurements.

As different calibration methods are discussed in the manuscript, it would also be interesting to hear what the authors think on use of dynamic models, like in Zaidan et al. (2020), which are able to account for evident autoregression in the data.

We thank the reviewer for recommending this article. The authors achieved interesting results using very comprehensive autoregressive models. Although we understand the direction that the paper suggests, for the purposes of our study, the choice to correct the LCS network using the MARS method was a fast yet effective method to validate the measured data collected for the purpose of microscale model validation. However, we are impressed by this method and can focus on testing it in future campaigns.

Lines 737-738: stating that transport is not the main source of PM10 and PM2.5 in European cities needs reference.

This statement has been edited with adding the reference: "These results may suggest that with the current development of cars in recent years, transport might not be the main source of aerosol pollution in European cities, unlike nitrogen oxides (see for example Scerri et al., 2023)" (lns 641-643 on p26).

Scerri, M. M., Weinbruch, S., Delmaire, G., Mercieca, N., Nolle, M., Prati, P., and Massabò, D.: Exhaust and non-exhaust contributions from road transport to PM10 at a Southern European traffic site, Environmental Pollution, 316, 120569, https://doi.org/10.1016/j.envpol.2022.120569, 2023.

References

Cantrell, C. A.: Technical Note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems, Atmos. Chem. Phys., 8, 5477–5487, https://doi.org/10.5194/acp-8-5477-2008, 2008.

Mikkonen, S., Pitkänen, M. R. A., Nieminen, T., Lipponen, A., Isokääntä, S., Arola, A., and Lehtinen, K. E. J.: Technical note: Effects of uncertainties and number of data points on line fitting – a case study on new particle formation, Atmos. Chem. Phys., 19, 12531–12543, https://doi.org/10.5194/acp-19-12531-2019, 2019.

Zaidan, M. A. et al., "Intelligent Calibration and Virtual Sensing for Integrated Low-Cost Air Quality Sensors," in IEEE Sensors Journal, vol. 20, no. 22, pp. 13638-13652, 15 Nov.15, 2020, doi: 10.1109/JSEN.2020.3010316.