



*Supplement of*

## **Performance validation and calibration conditions for novel dynamic baseline tracking air sensors in long-term field monitoring**

**Han Mei et al.**

*Correspondence to:* Peng Wei ([pengwei@sdnu.edu.cn](mailto:pengwei@sdnu.edu.cn)) and Zhi Ning ([zhining@ust.hk](mailto:zhining@ust.hk))

The copyright of individual parts of the supplement might differ from the article licence.

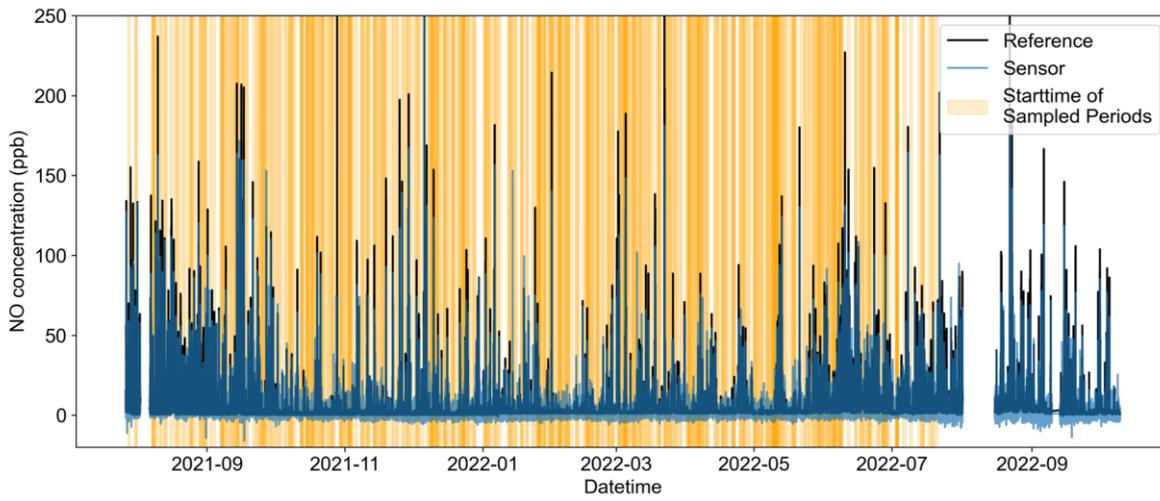


Figure S1. The start times of 500 randomly selected calibration periods for the MAS1 NO sensor.

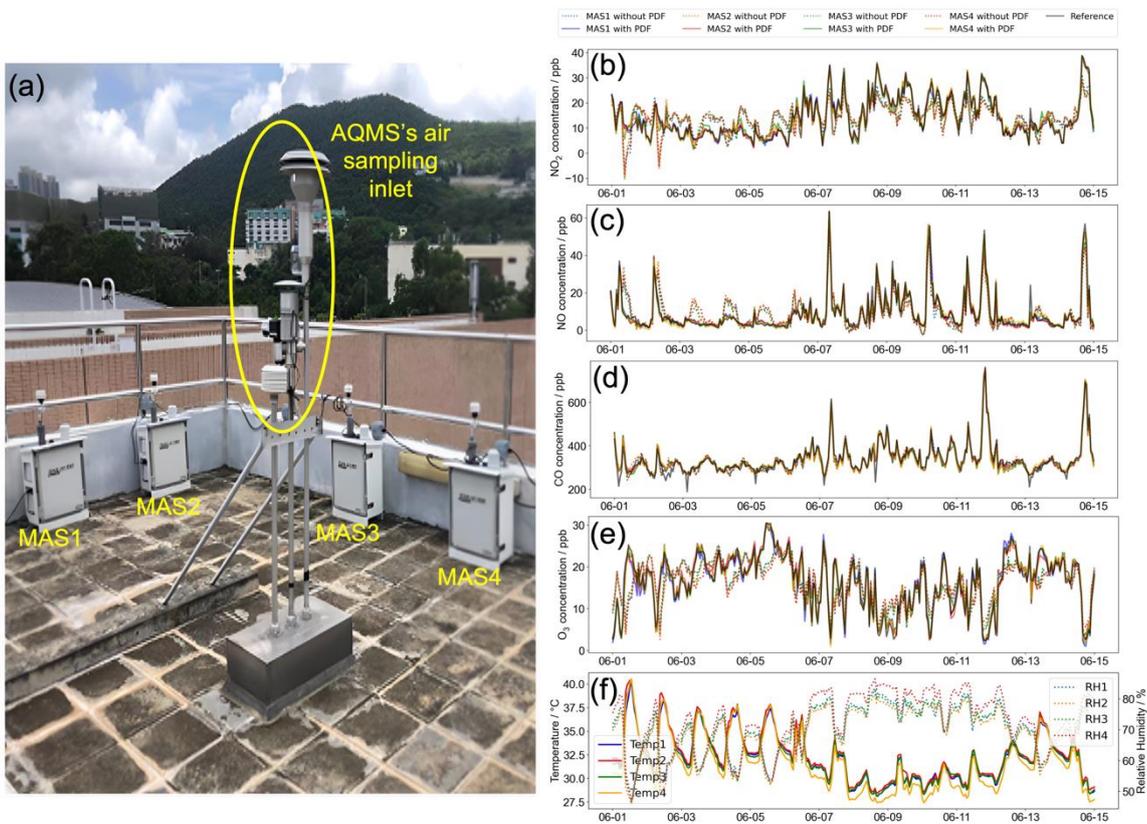


Figure S2. (a) Setup and (b-e)  $\text{NO}_2$ , NO, CO, and  $\text{O}_3$  long-term field data comparison of four MAS units with the AQMS in Hong Kong in 2019. (f) shows the temperature and RH measured inside the four MAS gas sensor modules.

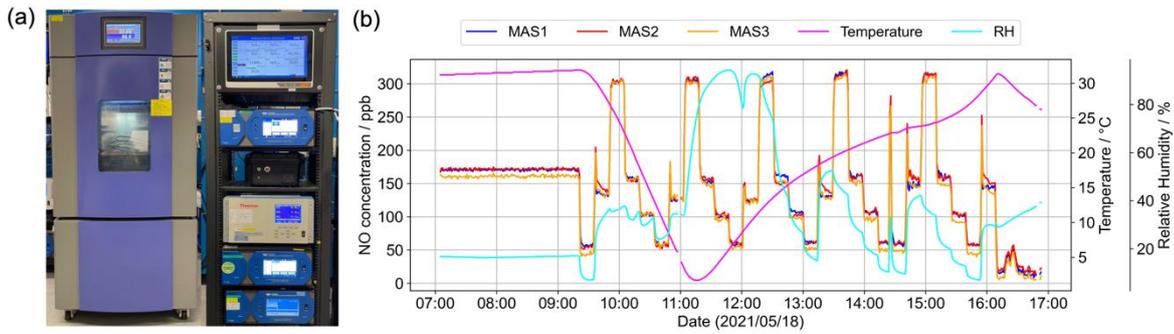


Figure S3. (a) Laboratory environmental chamber setup and (b) the response of 3 MASS' NO sensors under multiple point concentrations in laboratory temperature and humidity test.

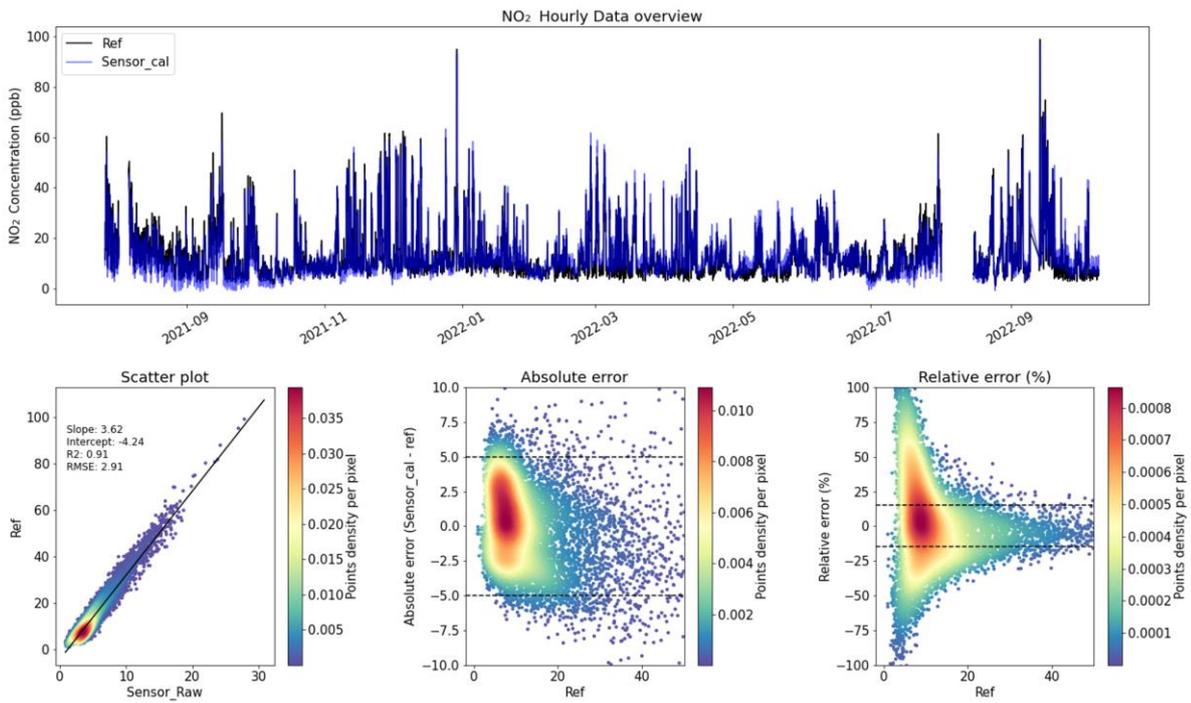


Figure S4. MAS1 NO<sub>2</sub> sensor calibrated data overview.

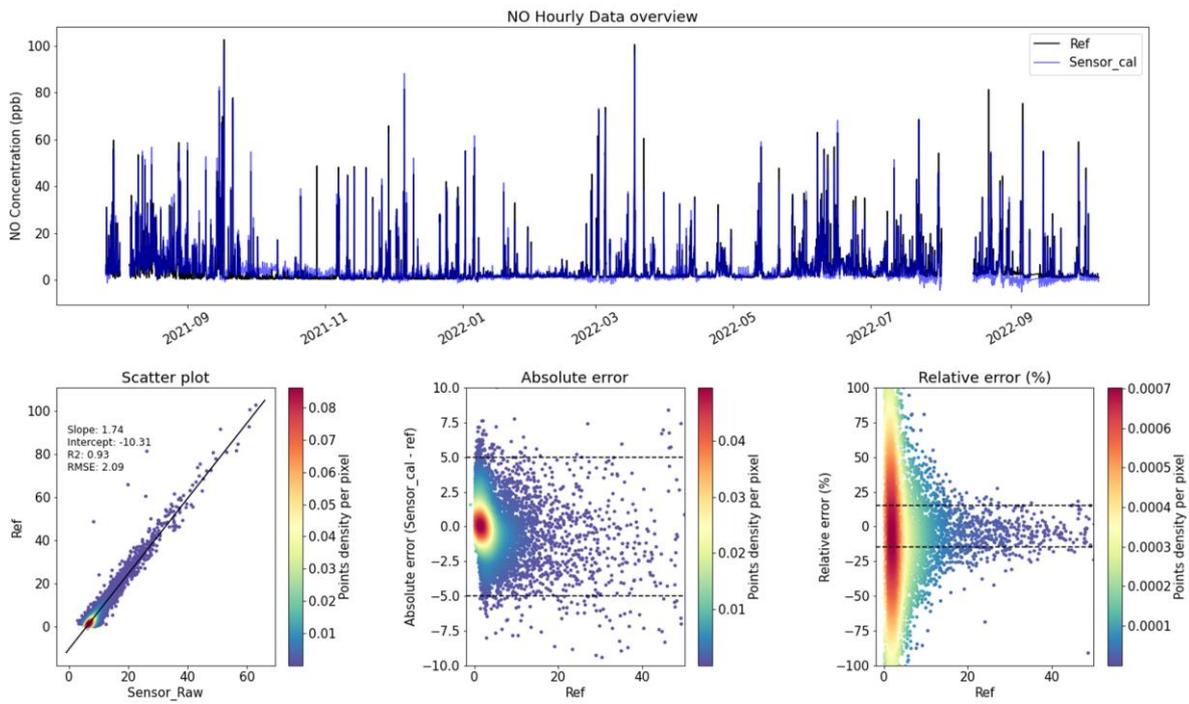


Figure S5. MAS1 NO sensor calibrated data overview.

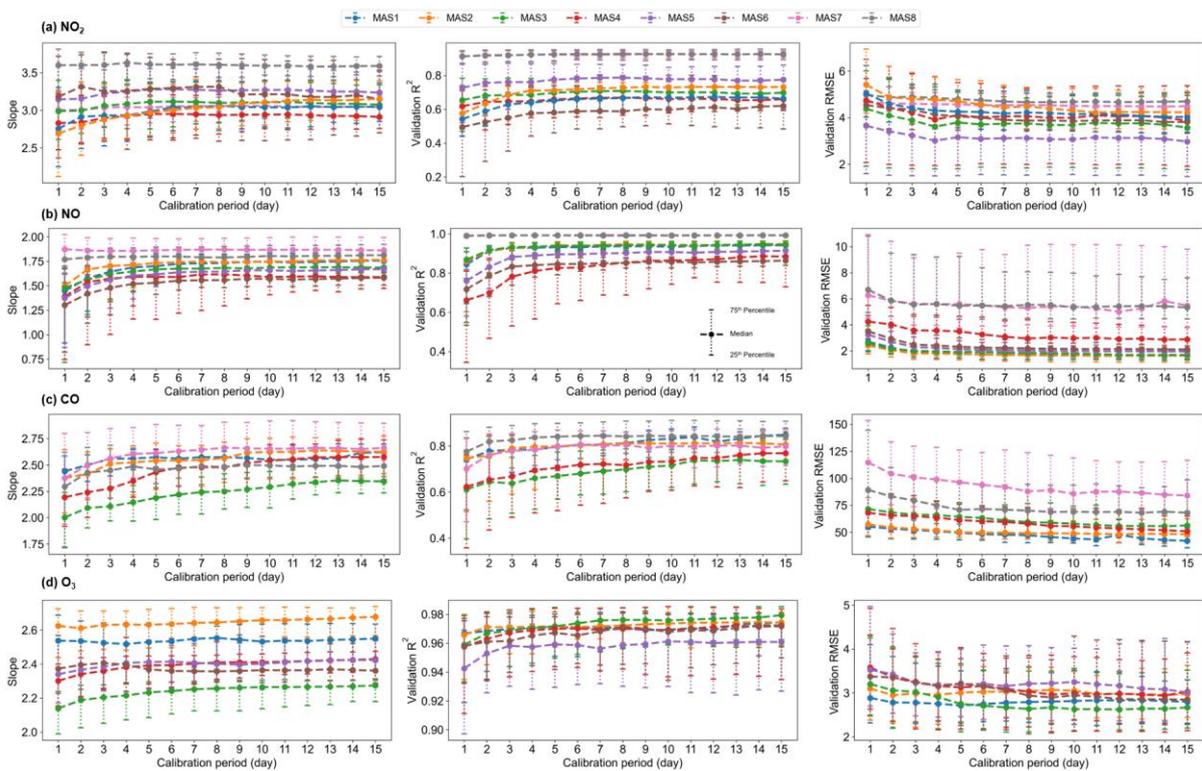


Figure S6. The potential range of the calibration slope, validation  $R^2$  and RMSE for a given calibration period for each MAS (a)  $\text{NO}_2$ , (b)  $\text{NO}$ , (c)  $\text{CO}$ , and (d)  $\text{O}_3$  sensors. Different colored lines represent the results of different MAS units. The vertical error bar is the 25%–75% distribution of the results under different calibration periods.

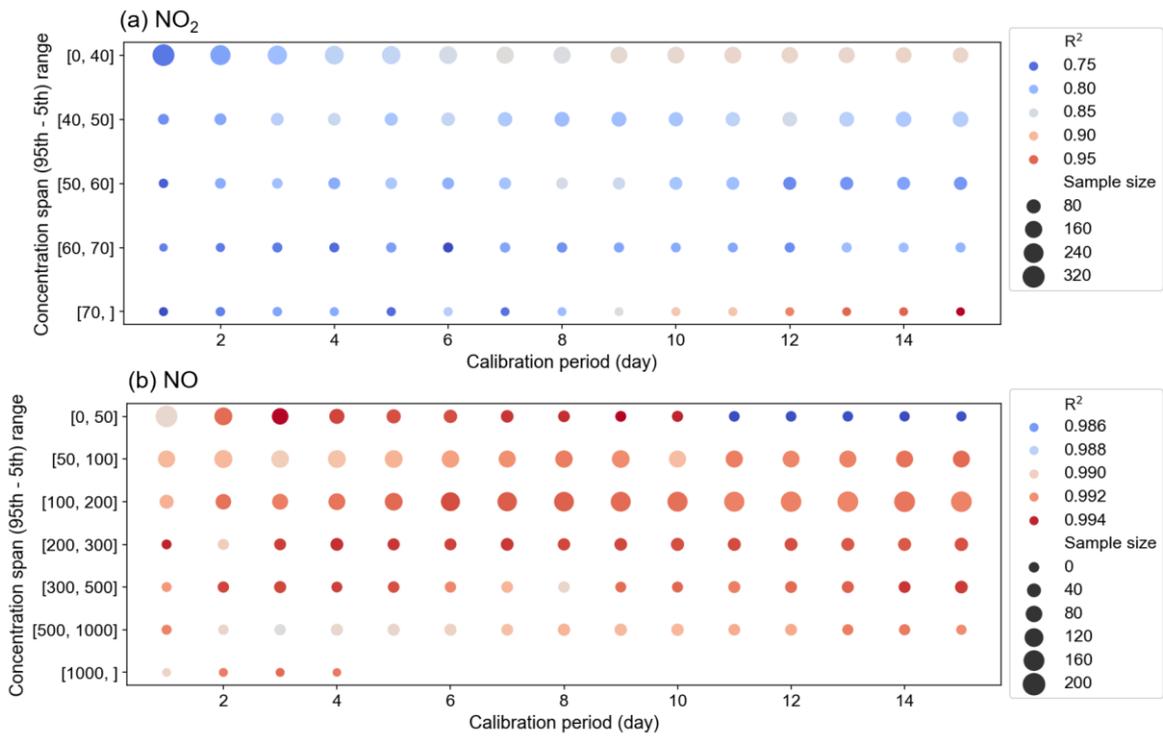


Figure S7. (a) NO<sub>2</sub> and (b) NO bubble plot of median R<sup>2</sup> of MAS units 7 and 8 (located in the high-concentration region (Shanghai)) and two factors: calibration period and concentration span. The size of the bubbles represents the number of samples. The color of these bubbles represents the median R<sup>2</sup> values in corresponding categories. Red represents higher R<sup>2</sup> values, while blue represents lower R<sup>2</sup> values.

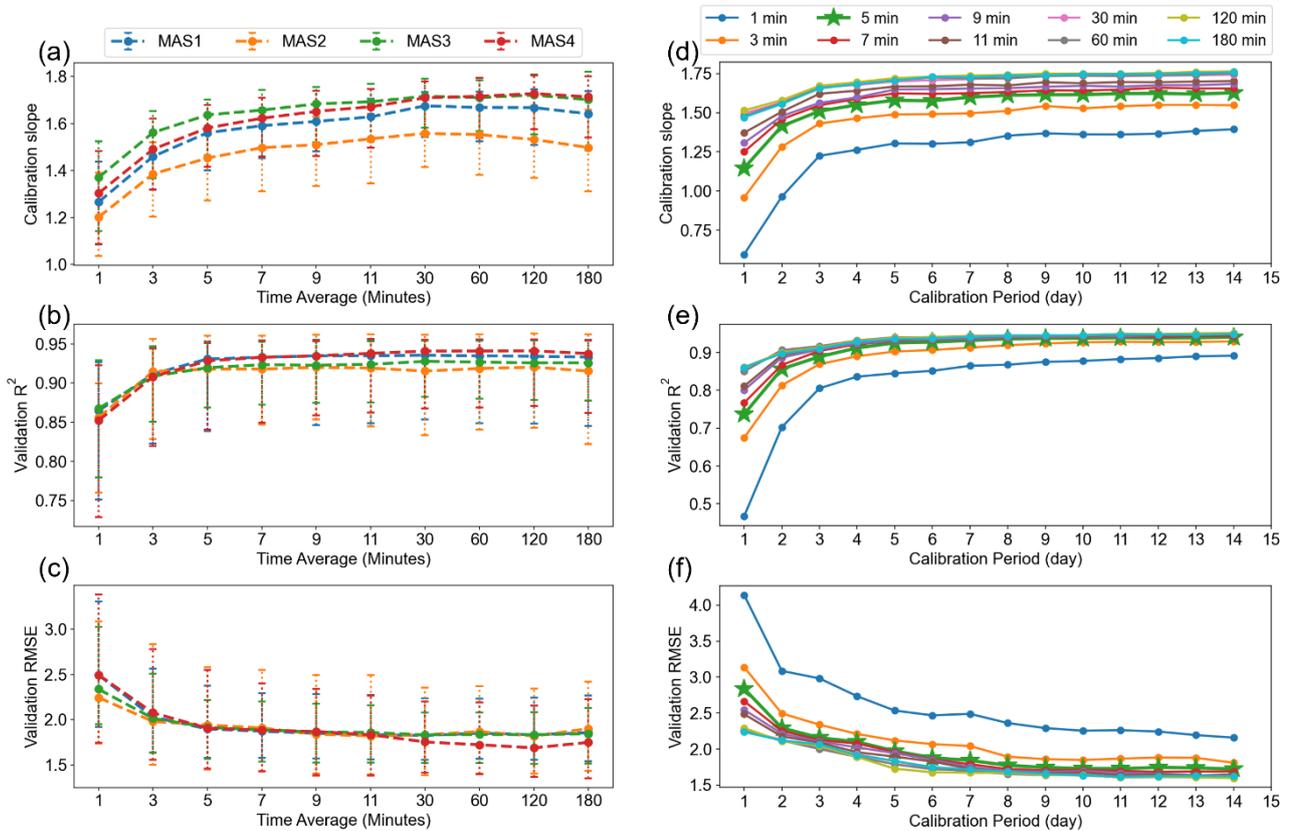


Figure S8. (a)-(c) The potential range of calibration slope, the  $R^2$ , and the RMSE of the validation set for MASs 1-4 NO sensors, under various time averaging with a calibration period of 1 day. Different colored lines represent the results of different MAS units. The vertical error bar is the 25%–75% distribution of the results under different categories. (e)-(f) The calibration slope median, the  $R^2$  median, and the RMSE median of the validation set for MAS1 NO sensors across all calibration periods, with different colors denoting time averages ranging from one minute to three hours.

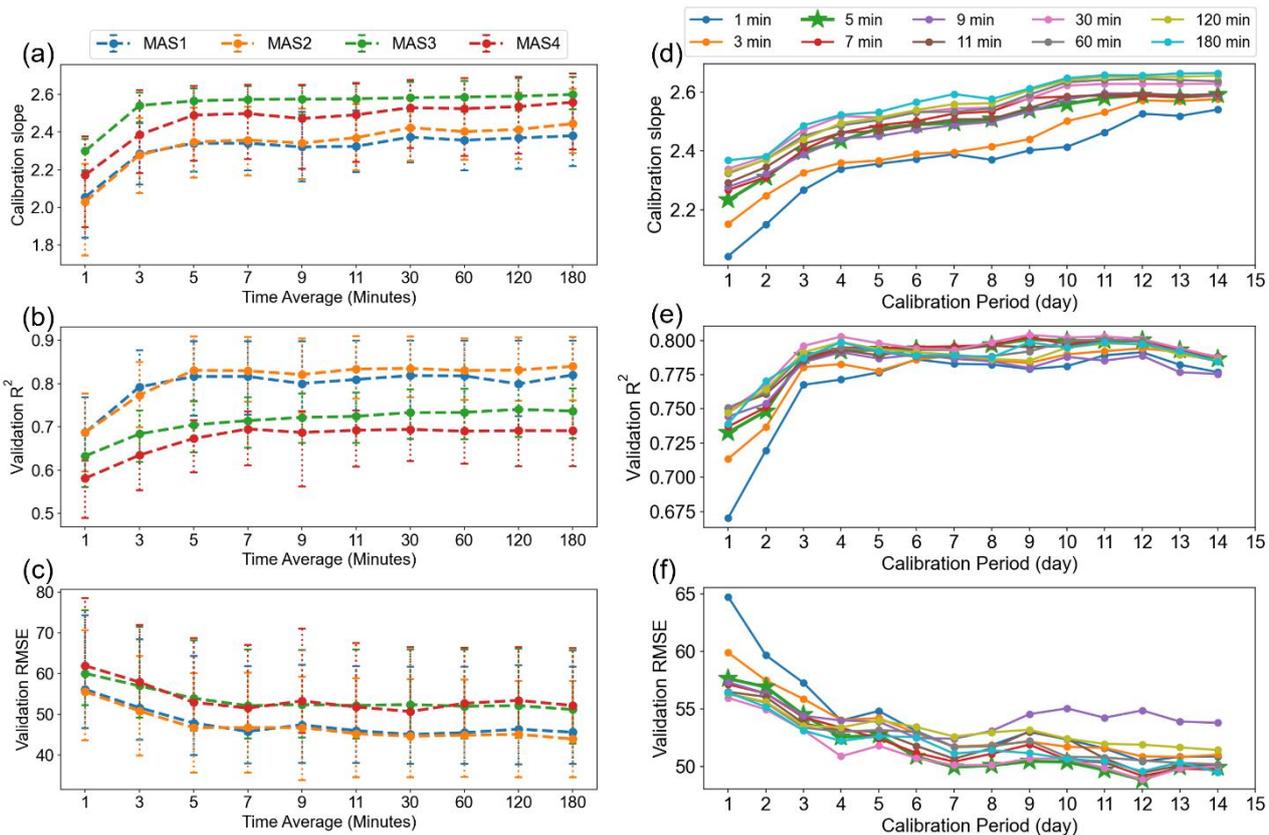


Figure S9. (a)-(c) The potential range of calibration slope, the  $R^2$ , and the RMSE of the validation set for MASs 1-4 CO sensors, under various time averaging with a calibration period of 1 day. Different colored lines represent the results of different MAS units. The vertical error bar is the 25%–75% distribution of the results under different categories. (e)-(f) The calibration slope median, the  $R^2$  median, and the RMSE median of the validation set for MAS1 CO sensors across all calibration periods, with different colors denoting time averages ranging from one minute to three hours.

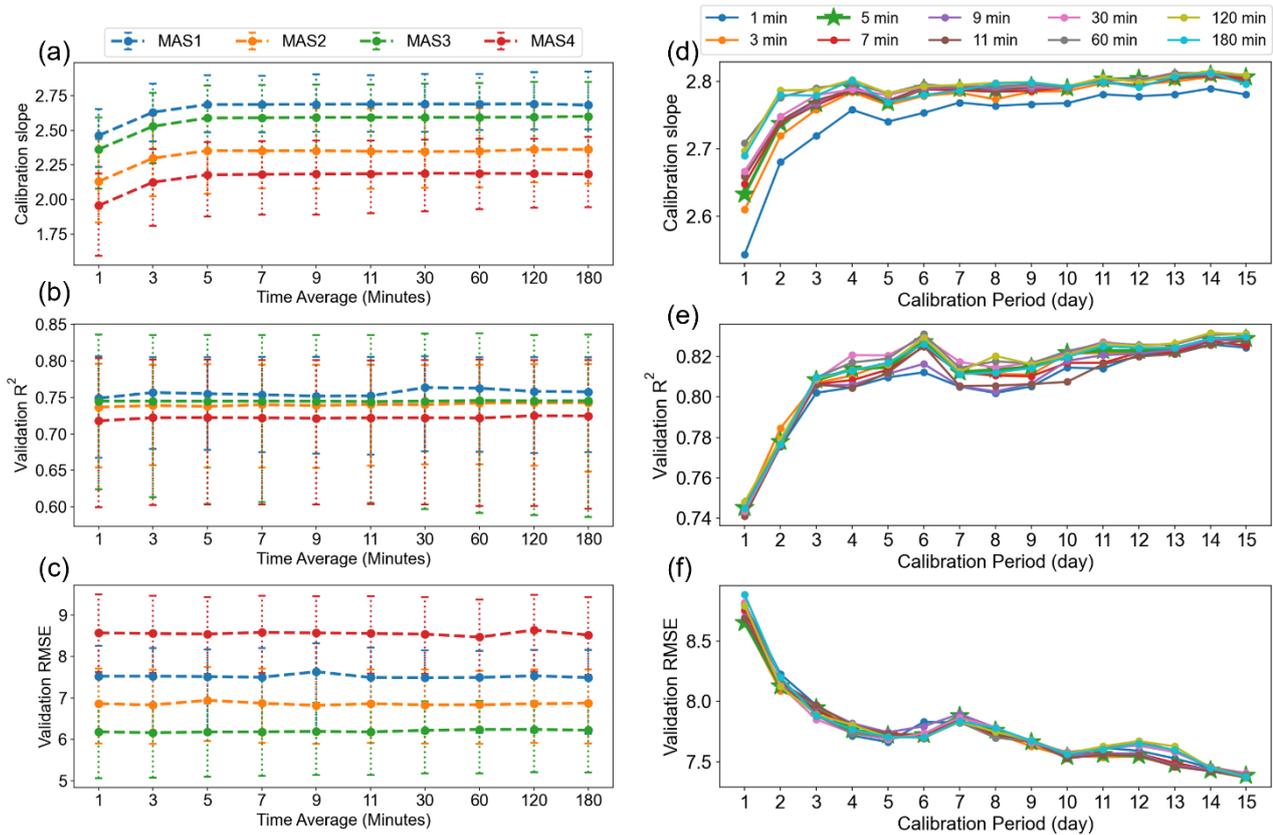


Figure S10. (a)-(c) The potential range of calibration slope, the  $R^2$ , and the RMSE of the validation set for MASs 1-4  $O_3$  sensors, under various time averaging with a calibration period of 1 day. Different colored lines represent the results of different MAS units. The vertical error bar is the 25%–75% distribution of the results under different categories. (e)-(f) The calibration slope median, the  $R^2$  median, and the RMSE median of the validation set for MAS1  $O_3$  sensors across all calibration periods, with different colors denoting time averages ranging from one minute to three hours.

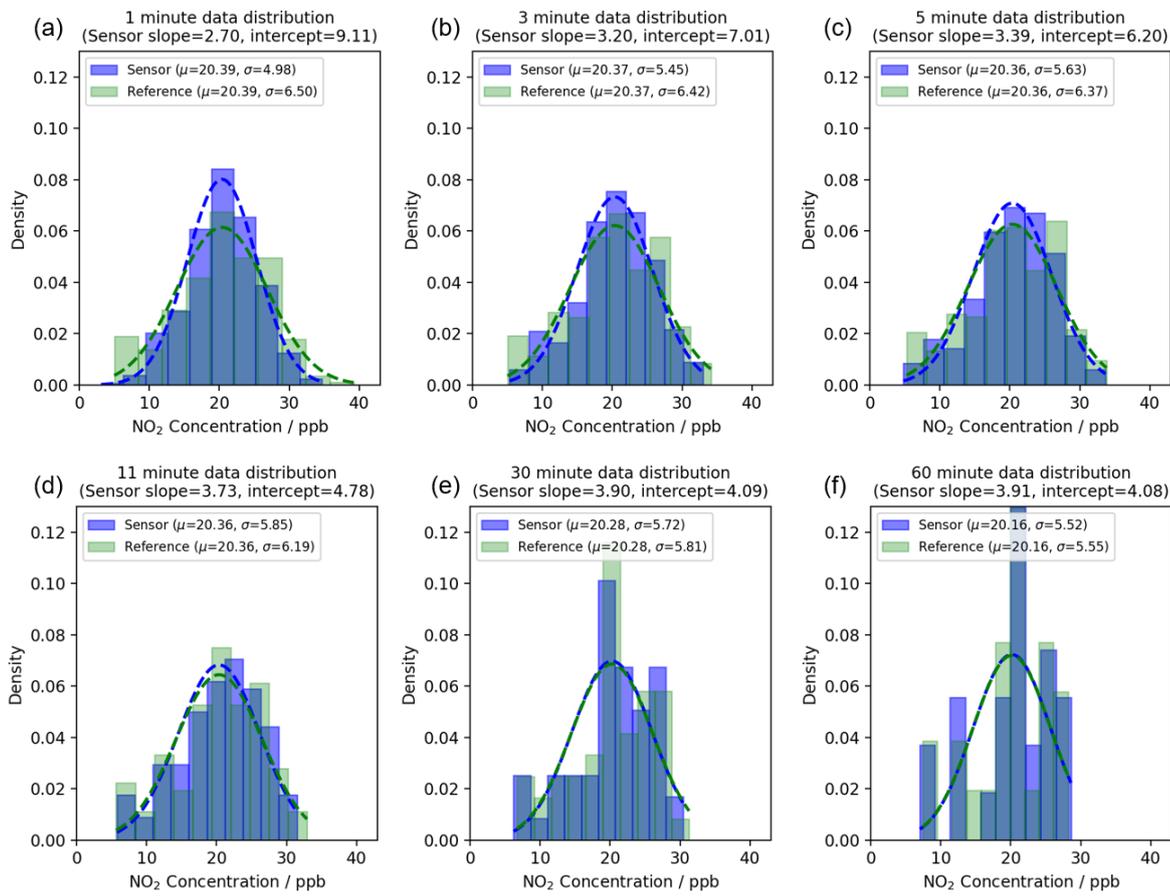


Figure S11. Comparison of the distribution between reference data and the MAS1 NO<sub>2</sub> sensor data after calibration using varying time averaging. The purple and green dashed lines represent the normal distribution of the sensor and reference fitting, respectively.  $\mu$  and  $\sigma$  represent the mean and standard deviation of the normal distribution, respectively.

### S1. Potential causes of sensor calibration coefficient variation from mathematical perspective

To delve deeper into the enhanced calibration performance with increased time averaging, the principles of linear regression are pivotal (Marill, 2004). In calibrating sensor data via linear regression (Reference =  $a \times \text{Sensor\_raw} + b + \varepsilon$ ), it is posited that the dependent variable  $Y$  (Reference) consists of the linear portion ( $a \times \text{Sensor\_raw} + b$ ) and a residual component ( $\varepsilon$ ) adhering to a normal distribution (Tripepi et al., 2008). The residuals in this model should conform to the white noise criteria (Kulperger, 1998; Rahmatullah Imon, 2009), signifying their independence, identical distribution, and lack of autocorrelation. These residuals should also be uncorrelated with both the independent variable  $X$  (Sensor\_raw) and dependent variable  $Y$ , maintaining a zero mean and constant variance, indicative of 'homoscedasticity' or its absence, 'heteroscedasticity'.

Analyzing the residuals is essential for grasping sensor data nuances and validating the model's calibration coefficients (Law and Jackson, 2017). Figure S12 displays the residual plot in the  $X$  and  $Y$  axes, where residuals display a random, homoscedastic distribution over  $X$  values but turn

heteroscedastic with a strong linear tie to  $Y$  in minute-level data. This pattern suggests overlooked influential factors in the calibration model (Tripepi et al., 2008), which significantly interact with  $Y$ , affecting the residuals systematically.

Consequently, the predictive capability of the calibration model may be compromised since it fails to encapsulate these crucial variables' effects on  $Y$ . However, averaging  $X$  and  $Y$  over time tends to normalize the residuals' homoscedasticity in the hourly data along the  $Y$  axis, possibly due to mitigating heteroscedasticity-inducing elements during time averaging (Long and Ervin, 2000). This leads to a more homoscedastic arrangement in  $Y$  and a mitigated impact on the  $X$ - $Y$  regression relationship. While further analysis is needed to pinpoint the factors affecting residual behavior, it's clear that extended time averaging of sensor data facilitates calibration coefficients nearing the ideal solution, highlighting the importance of appropriate time averaging in achieving optimal calibration.

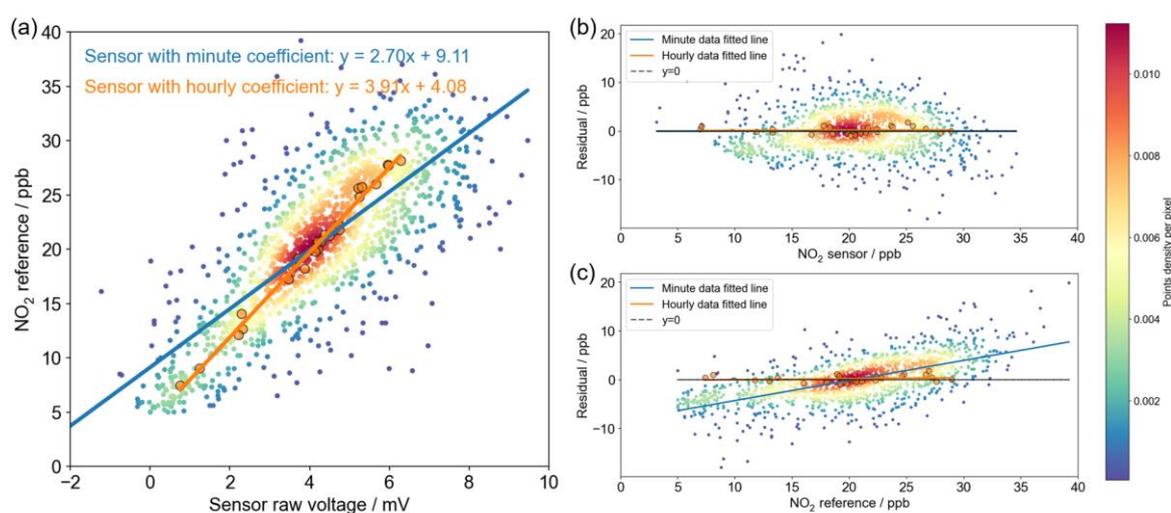


Figure S12. (a) Scatter plots of the reference and sensor\_raw data of MAS1's NO<sub>2</sub> sensor at hourly and minute-level time averaging, respectively. (b) Distribution of sensor residuals (Reference - Sensor) in the sensor direction. (c) Distribution of sensor residuals in the reference direction. The red and blue lines in (b) and (c) are the fitted trends of the residuals.

## References

Kulperger, R.: A regression residual process, *ASYMPTOTIC METHODS IN PROBABILITY AND STATISTICS: A VOLUME IN HONOUR OF MIKLOS CSORGO*, 741–757, <https://doi.org/10.1016/B978-044450083-0/50049-6>, 1998.

Law, M. and Jackson, D.: Residual plots for linear regression models with censored outcome data: A refined method for visualizing residual uncertainty, *COMMUNICATIONS IN STATISTICS-SIMULATION AND COMPUTATION*, 46, 3159–3171, <https://doi.org/10.1080/03610918.2015.1076470>, 2017.

Long, J. and Ervin, L.: Using heteroscedasticity consistent standard errors in the linear regression model, *AMERICAN STATISTICIAN*, 54, 217–224, <https://doi.org/10.2307/2685594>, 2000.

Marill, K.: Advanced statistics: Linear regression, Part I: Simple linear regression, *ACADEMIC EMERGENCY MEDICINE*, 11, 87–93, [https://doi.org/10.1197/S1069-6563\(03\)00600-6](https://doi.org/10.1197/S1069-6563(03)00600-6), 2004.

Rahmatullah Imon, A.: Deletion residuals in the detection of heterogeneity of variances in linear regression, *JOURNAL OF APPLIED STATISTICS*, 36, 347–358, <https://doi.org/10.1080/02664760802466237>, 2009.

Tripepi, G., Jager, K. J., Dekker, F. W., and Zoccali, C.: Linear and logistic regression analysis, *Kidney International*, 73, 806–810, <https://doi.org/10.1038/sj.ki.5002787>, 2008.