Response to interactive comments from Referee #1

Thank you for the time you put into reviewing our manuscript and the helpful feedback. Please see our following responses and proposed changes to the original manuscript, which we believe, help to improve this paper and increase its impact. Below the comments from Referee #1 are given in black. Our responses to the comments are shown in blue. Text added or changed in the manuscript is marked in italics.

First of all I would like to congratulate the author for this very interesting and very complete work on the real use of sensor system in an epidemiologic study. The use of the uncertainty as a marker for reliability is a very interesting choice. However, I would like to emphasize that the European Directive DQO uses a specific value usually taken as the limit value set by the Directive itself. Nevertheless, the work presented in this paper has been carried out with great care resulting in a very nice publication.

We sincerely thank the reviewer for the kind words and thoughtful feedback on our work. We greatly appreciate the recognition of our efforts in applying sensor systems within an epidemiologic framework, as well as the interest in our approach to using the relative expanded uncertainty to evaluate of the sensor performance.

Regarding the valuable observation on the European Directive DQOs (Data Quality Objectives), we acknowledge the importance of aligning with the directive's specifications, particularly in reference to the limit values. However, the DQOs of EU Directive 2008/50/EC and 2024/2881 as well as the mentioned CEN/TS 17660-1 (2021) and 17660-2 (2024) set minimum temporal resolutions of 24-hour, 8-hour and hourly (depending on the pollutant) for short-term mean concentrations. In our study, due to the limited time we had for the indoor and outdoor calibrations we tested temporal resolutions of 1, 5, 10, and 15 min for NO_2 sensors and 1 and 30 min for $PM_{2.5}$ sensors, for indoor and outdoor respectively. This differ from the temporal resolution stated in the EU Directive 2008/50/EC and 2024/2881 to calculate the relative expanded uncertainty.

Moreover, in the CEN/TS 17660-1 (2021) and 17660-2 (2024) were our study takes the inspiration from, strict requirements are defined for the tests: sensors must be co-located at least two different type of stations, two seasons per type of station and 40 days per campaign. As the full project had a length of 1.5 years including design of two sensor systems, calibration, deployment and evaluation, it was not possible to co-locate the sensors with reference equivalent instruments for more than 2 weeks before each deployment. The attempt of evaluating the REU in the limit values (both the current or the future ones) could lead to a misinterpretation of the sensor classification.

We have modified the following text in the manuscript to acknowledge this limitation:

"Note that the short-term DQO EU Directives were conceived for hourly and daily averages for NO_2 and $PM_{2.5}$, respectively. For epidemiological studies, however, especially those using portable monitors, 24 h average or even 1 h average may be insufficient, as detecting short-term pollution peaks requires higher temporal resolutions. Moreover, longer co-location periods are not always possible during the exposure assessment campaigns and consequently, the use of a 1-hour average can decrease considerably the available data to train the calibration models and reduce the range of T and RH, as well as the pollution concentration range used. Therefore, in this work, we present the REUs of the NO_2 models for different averaging times, that is, 1, 5, 10, and 15 min and thus, an evaluation of the REUs at the limit value is not applicable. Similarly, co-location measurements of indoor $PM_{2.5}$ sensors in a particle chamber with high particle concentrations lasted an average of 2 to 3 hours. Therefore, the uncertainties were calculated for a 1-min average. For outdoor $PM_{2.5}$ sensors where more data points are available, a 30-min average was used so that neither REU for $PM_{2.5}$ measurements for indoor or outdoor are applicable in the region of the limit values."

You can find below some comments I had on some specific point:

- Table 2: in the comparison with outdoor AQMS, how did you define the 6km limit?

It is not a real limit, it represents the air distance of the house located the furthest from the monitoring station, that is, 6 km away. We have clarified this in the text as follows:

"Additionally, the data of the four outdoor air quality monitoring stations available in Stuttgart as well as the data of the monitoring station of the University of Stuttgart in Hauptstätter Street was also collected to qualitatively compare their NO_2 and the $PM_{2.5}$ trends with the data of the outdoor AQSSs during deployment in the houses of the patients. The air distances between the closest and the furthest monitoring station and the houses of the patients was 0.6 and 6 km, respectively (see Fig. S1)."

- Line 272-273: the REU needs to be calculated at a given value, usually the limit value also set in the Directive. You do not mention those value here, will you use a specific one?

Due to the above-mentioned reasons, we did not to evaluate the REU against a specific value (such as the current limit values). Therefore, we considered the graph displaying all measured concentrations to be a more appropriate option.

In the manuscript we have now added two tables indicating at which NO_2 concentration is reached the DQO of 25 % for indoor (Table 4) and outdoor (Table 5).

Table 4. *Concentration in ppb at which the DQO for indicative measurements (25 %) is accomplished for the outdoor calibration.*

Averaging time	***Model***	***B03-P1***	***B03-P3***	***B03-P7****	***B05-P4***	***B06-P4***	***B06-P7***	***B08-P6***
1 min	*MLR*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*
	SVR	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*
	RFR	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*
	ANN	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*	*N.A.*
5 min	*MLR*	*26*	*N.A.*	*27*	*N.A.*	*39*	*N.A.*	*N.A.*
	SVR	*23*	*27*	*N.A.*	*N.A.*	*34*	*23*	*21*
	RFR	*22*	*28*	*21*	*N.A.*	*33*	*20*	*21*
	ANN	*24*	*40*	*38*	*N.A.*	*39*	*N.A.*	*28*
10 min	*MLR*	*17*	*N.A.*	*N.A.*	*N.A.*	*29*	*N.A.*	*28*
	SVR	*17*	*23*	*19*	*N.A.*	*29*	*N.A.*	*14*
	RFR	*17*	*22*	*19*	*N.A.*	*24*	*33*	*17*
	ANN	*17*	*25*	*32*	*N.A.*	*28*	*38*	*19*
15 min	*MLR*	*18*	*N.A.*	*N.A.*	*N.A.*	*19*	*N.A.*	*24*
	SVR	*18*	*N.A.*	*26*	*35*	*19*	*21*	*11*
	RFR	*17*	*29*	*N.A.*	*30*	*20*	*N.A.*	*13*
	ANN	*18*	*44*	*N.A.*	*N.A.*	*18*	*N.A.*	*19*

N.A.: not accomplished.
**B03-P7 is an outdoor AQSS used for indoor measurements as part of an experiment to test the outdoor calibration methodology for indoor measurements.*

Table 5. *Concentration in ppb at which the DQO for indicative measurements (25 %) is accomplished for the indoor calibration.*

Averaging time	Model	B01-P4	B01-P5	B02-P1	B02-P6	B04-P2	B04-P3
1 min	MLR	8	11	N.A.	42	9	11
	SVR	3	2	N.A.	15	2	8
	RFR	1	-	40	7	-	3
	ANN	5	5	N.A.	31	5	7
5 min	MLR	7	10	N.A.	27	8	9
	SVR	1	14	N.A.	15	-	6
	RFR	1	-	60	21	-	2
	ANN	4	4	N.A.	22	4	5
10 min	MLR	6	10	N.A.	26	8	8
	SVR	1	1	N.A.	11	1	3
	RFR	1	-	39	25	-	1
	ANN	3	4	N.A.	21	4	5
15 min	MLR	7	9	N.A.	22	7	7
	SVR	1	4	N.A.	4	3	3
	RFR	1	-	N.A.	20	-	3
	ANN	3	4	N.A.	19	4	4

N.A.: not accomplished.
The cells marked with (-) do not have a value for the REU as $U_{field}(y_i)$ cannot be calculated with Eq. S6 in the Supplement due to the negative value of $u_s^2(y_i)$ (Eq. S1). This is caused due to the extremely low RSS. Near-zero RSS are an indicator of the overfitting of the RFR in the indoor calibration models.

A you can see, for some models trained with the data of the indoor calibration methodology, the calculation of the REU using the equations of the CEN/TS 17660-1 (2021) and CEN/TS 17660-2 (2024) gives for RFR calibration models negative values independently from the temporal resolution used. This behaviour has been now explained in the text:

"The lower REUs that are achieved during the calibration of AQSSs in indoor conditions may be due to the controlled conditions, as the NO_2 gas was given stepwise and kept constant for 3.5 hours, as well as the controlled changes of the T and the RH. This lack of variability in the calibration data resulted in low sum of residuals (RSS) triggered by model overfitting."

"This creates the challenge of calibrating indoor AQSSs for a wide range of NO_2 concentrations and meteorological parameters without causing model overfitting."

- Line 281: why did you mentioned both directive as the 2024 will overcome the 2008?

The work of WG42 in CEN/TS 17660-1 (2021) and CEN/TS 17660-2 (2024) is aligned with EU Directive 2008/50/EC. However, in this study, we aimed to go a step further by exploring uncertainty in the context of the upcoming regulation. At present, there are no discussions about revising the technical specifications, but this may change in the future.

We have clarified that in the manuscript:

"The new directive also specifies in Annex V new DQOs for indicative measurements (I.M.) and objective estimation (O.E.) that the Member States shall comply by 11 December 2026. Therefore, the inclusion of new DQOs is intended as forward-looking exercise."

"The CEN/TS 17660-1 (2021) and CEN/TS 17660-2 (2024) provides a classification that is consistent with the requirements of DQOs defined in the Directive 2008/50/EC."

- Figure 3: this kind of figure is not easy to read. I do understand that the main idea is to show a general trend but I'm a bit overwhelmed by so many information.

We acknowledge the difficulty on reading Figure 3 (now Figure 4). We hope the Table 4 and 5 can now help to understand and complement the REU Figures.

- Figure 3: which indicative measurement DQO did you select 2008 or 2024?

As shown in Table 1, the DQO of NO_2 for indicative measurements for short-term measurements is the same in both directives (source: Annex I Table A in EU 2008/50/EC and Appendix V, Table 2 in EU 2024/2881).

The manuscript already included the clarification "Note that both directives have the same DQO for indicative measurements (25 %)." We have now added the following clarification in the caption of Figure S3 and S4 in the supplemental material.

"The DQO for short-term indicative measurements is the same in both Directives."

- Figure 6: I would advise the author to change the color scheme, currently indoor (pink) and outdoor (dark pink) to get a more distinguishable colors.

We have changed the colours of Figure 6 and Figure 7 (now Figures 7 and 8) to white (indoor) and dark grey (outdoor).