Overall response: We would like to thank reviewer #3 for their useful and insightful comments. The reviewer recognises the importance of the manuscript and asks for technical clarifications. However, we would like to emphasize that this manuscript is not aiming to provide a *generic* evaluation of all portable air quality sensor types but instead aims to characterise the performance of this specific configuration with currently available sensor variants for specific environmental conditions, as a necessary precursor to a series of papers on exposure and health impacts using these PAMs. While the various methodologies we used would be of value to the scientific community, as the field of air quality monitoring evolves, different sensor types, models and configurations might be expected to lead to different results.

Detailed response

Abstract 1: "*It would appear beneficial to include brief selected quantitative summary information on the precision, accuracy…."*

Response (added): ….Overall, the air pollution sensors showed high reproducibility (R^2: 0.80 -1.00, mean R^2 = 0.93) and excellent agreement with standard instrumentation (R^2: 0.56 – 0.99, mean R^2 = 0.82) in outdoor, indoor and commuting microenvironments across seasons and different geographical settings.

Introduction 2: "this paper further aims to create a roadmap for calibration and validation of portable monitors suitable for personal exposure quantification…. *wondered if this aim could be stated in a clearer way?"*

Response (re-phrased as follows pg 2 ln 35-37): As the field of novel air pollution sensing technologies expands rapidly, this paper further aims to provide methodological guidance to researchers from diverse disciplines on how to comprehensively calibrate and validate portable monitors suitable for personal exposure quantification.

Methods 3. "*it would be helpful to clarify if the monitoring systems being tested are [or will be] commercially available to wider research communities …. also there may be a substantial number of person years invested in the design and construction of the monitoring systems…"*

Response (added as follows pg 2 ln 39-41): The PAM has been developed at the Department of Chemistry, University of Cambridge in collaboration with Atmospheric Sensors Ltd. It is now commercially available (independently from the University of Cambridge) from Atmospheric Sensors Ltd (Model AS520, http://www.atmosphericsensors.com).

Methods 4 and 5. "*It would be helpful to expand on the principle of operation of the gas sensors used… accessible to a wider audience… e.g. a diagram similar to diagrams of 3-electrode systems in earlier references"* and on the same issue of 4-electrode EC sensors "*Is it ok to directly translate the finding to the newer 4-electrode sensor, or is separate quantification of the linearity and LoD of the newer sensor also required?*"

Response: (added in paper pg4 ln1-10): These EC sensors are the A4 variant from Alphasense (ref) and operate on a four electrode system. The principle of operation of the four electrode system is identical to that of the earlier variants of three-electrode system with an additional electrode, the auxiliary (or non-sensing) electrode to compensate for the temperature dependence of the cell potential. …. The linearity and LoD of the 4-electrode sensors (when integrated in the PAM) have been tested under laboratory conditions following the same methodology as described in Mead et al. 2013, yielding very similar results.

Response to referee: Rather than include inevitably incomplete sensor details (as much is proprietary to the manufacturer), we feel it more appropriate to refer readers (and the referee) to the manufacturer's website for further details about the principles of operation of the individual sensors integrated in the PAM. The 4th electrode was implemented to correct for

systematic errors and not to change the LOD. Hence, we expect the linearity and LOD to be comparable with those of the 3-electrode sensors.

Methodology 6. ***"****A strength of your model is that you use a conceptually simple model……present results of the equations fitted… in Supplementary Information".*
Response to referee: Such calibration factors are likely to be location specific. Making them public, would risk encouraging other researchers taking them as generic.

Methodology 7. *"It would appear useful to make your field calibration data available"*
Response: Data availability: https://doi.org/10.17863/CAM.41918

Methodology 8. "*Add references for Mie"*
Response (pg5 ln 7-8, references added):
We have included a suitable reference.
Mie, G. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.* 1908, *330*, 377–445.
Bohren, C. F.; Huffman, D. R. Absorption and scattering of light by small particles; *Wiley*, 1988.

Methodology 9 &10. "*robust objective way of splitting the data to calibration-validation periods"*…. "*Currently it is not clear what calibration data has been used to adjust the sensor outputs in these figures*"
Response (added to the paper):
To evaluate the performance of the linear model, the datasets were split into training (i.e. calibration) and validation periods to first extract the calibration parameters and then apply them to the validation set and compare the measurements with those from reference instruments (referred to as "calibration - validation" method). The training sets ranged from 1 to 16 days, and the adjusted coefficient of determination ($\overline{R}^2$) remained stable for training periods longer than 3 days. Therefore, approximately a third of the dataset was selected as a training set. As relationships in these linear models should ideally not be extrapolated beyond the range of the observations (including meteorological conditions), the calibration periods covered the temperature and concentration ranges in which the sensors were deployed (Cross et al., 2017). Once the performance of the model was established in diverse environments, we used the full co-location periods to determine the agreement between PAM sensors and reference instruments.
Response to referee (not in the paper):
Calibration of gaseous sensors for the China deployment: The China deployment lasted about one year, and no evidence of a significant drift of the gaseous sensors was found. Using a unified **training** dataset from the outdoor co-locations **both in the heating and non-heating season** the coefficients of the linear model were determined. As the variation between seasons was greater than the variation within seasons and covered a wide range of environmental conditions and pollution levels, the choice of the training period did not affect the performance of the model. The coefficients determined in the training set were then applied to the **validation** periods of both seasons and were also used to convert raw measurements to ppb in the indoor deployment. In this way, we prove that the calibration parameters determined in outdoor co-locations are suitable across a range of environmental conditions and for indoor deployments too, and therefore for personal exposure as people spend as much as 90% of their time in indoor environments.

Calibration for the UK deployment: The PAMs were co-located outdoors next to reference instruments in 2015, and were then deployed for two years continuously to participants of the

COPE project. The PAMs were again co-located next to reference instruments in 2017 after the completion of the project. The change in the sensor sensitivities did not show a systematic error. Individual sensor drifts were quantified and were corrected by linearly interpolating sensitivities between the start and the end of the deployment.

Results 12. "*Dates and measurement times in Table 3"*
Response: Added to Table 3 first column

Colocation	**Start**	**End**	**Total time**
China (Dec- Jan)	28/12/2016	15/01/2017	447 hours
UK (Oct- Nov)	27/10/2017	13/11/2017	408 hours
China (June)	28/06/2017	16/07/2017	432 hours
UK (April-May)	26/03/2018	10/04/2018	342 hours

Results 13 & 14. "*Express RMSE as percentage of mean reference"* **–**
Response: We have added mean values of each pollutant during the co-location periods. We report RMSE as a percentage of the maximum (as the 95% percentile), as peak exposure events are more relevant in health studies. The reader is now also able to calculate the RMSE as percentage of the mean reference values.

Results 15. pg 11 ln 6 *"were the extreme temperatures recorded inside the monitor enclosure and within the sensors?"*
Response to referee: The temperature inside the PAM is on average 7℃ higher than ambient temperatures. The monitors were co-located in metal shelters on the roof of a research building in the Peking University campus (Figure below), and were exposed to direct sun further increasing the temperature inside the PAM.


Figure (not included in manuscript): Co-location of 60 PAMs at PKU during the heating and non-heating season for a four-week period. PAMs were placed in custom-made protective shelters in proximity to the inlets of reference instruments and were connected on mains.
Response added: In Table 3 added "internal conditions of the PAM"

Results 16. "*Why was O_3 sensor less affected by heat then NO_x and CO? - O_3 concentrations were higher than usual?"*
Response to referee: The principles of operation of the NO_2 and O_3 sensors are very similar. It is possible that the O3 sensors perform better because (a) the concentrations of O_3 were higher in the summer and (b) the sensor is further away from the internal battery of the PAM and therefore slightly cooler than the other sensors. The next generation of PAM hardware will integrate an insulation layer between battery and sensors. NO was close to the LOD (maximum= 5ppb, mean =1 ppb).

Results 17. "*Your reference to the participants on p11 could be given with further detail e.g. geographical location / time of year etc. to provide context to the comparison with the field evaluation measurements."*
Response: FIGURE A5 CAPTION Added: The co-location with the reference instruments on the roof of Peking university took place from 28/06/2017 to 16/07/2017, the field deployment was conducted in Beijing and Pinggu from 22/05/2017 to 26/06/2017.

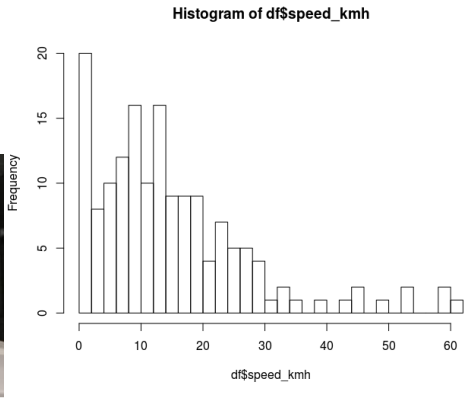
Results18. *"In Figure 4 caption it would be helpful to specify how far the 'PKU reference site' and the 'nearby government monitoring site' were from the indoor measurement location"*
Response ADD TO FIG 4 CAPTION: PKU roof is 5.3 km and the governmental reference station (Haidianwanliu) is 6 km away from the location of the indoor experiment.

Results 19. "*On p12 you assert that your results prove the suitability of the low cost sensors for quantification of indoor air pollution. It would seem appropriate to qualify the promising indication of suitability over timescales and conditions similar to the calibration period/conditions?*"
Response: Add to PAGE 12 LINE 34:
The conversion of the raw measurements to ppb used the sensitivities extracted using outdoor co-locations both during the heating and non-heating season (subsection 3.1) with the linear model (subsection 2.1).
….proving the suitability of this monitoring platform to quantify indoor air pollution levels for these species provided they have been adequately calibrated in the local environment.

Results 20. "*Do you have information on the response times of the sensors…"*
Response to referee: Response times have been determined in the Alphasense Technical Specification sheets of each sensor (see references in Section 2.1, page 3) and ranged between t_{90} < 25s for CO (from 0 to 10ppm) and t_{90} < 60s for O_3 (from 0 to 1ppm). As the response times are smaller than the time resolution for the PAM measurements we do not expect them to affect the way of correcting the sensors' cross sensitivities.

Results 21. "*When the PAM was mounted on the roof of the vehicle how were the sensor inlets orientated in relation to airflow? Did the varying speed of airflow have any effect on agreement between sensors and reference instruments?"*
Response (added to manuscript pg 15): The PAM was mounted on the roof with the OPC inlet facing forwards and the EC sensors facing to the sides (see photo below). The reference instrument inlets were located on the car roof as well. The speed mostly varied between 5 and 20 km/h (see histogram below). There was no correlation between car speed and RMSE error in the gaseous and particulate measurements. The OPC contains an airflow measurement unit which compensates for any wind or internal flow dependence.

The histogram below shows the distribution of the speed measurements covering the same time period as Figure 6 in Section 3.2.



Results 22. *"11.30- 12.30 e.g. was vehicle static Do you know why O3 measured by sensor and reference instruments diverges during this time period?"*

Thank you for bringing this to our attention. During this period the vehicle was stationary, and we have therefore removed this data.