Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-285-AC1, 2019

Atmospheric Measurement Techniques Discussions

Interactive comment on "An improved low power measurement of ambient NO_2 and O_3 combining electrochemical sensor clusters and machine learning" *by* Kate R. Smith et al.

Kate R. Smith et al.

ks826@york.ac.uk

Received and published: 4 February 2019

Specific Comments

C1 : How accurate is 'good enough'? I think a bit more context regarding this question would be helpful to readers.

C1. Author's Response Thank you to the reviewer for the advice to be clearer when describing the requirements for sensor performance before they are considered able to perform as instruments in the field. The answer to this question is very application dependant and we cannot therefore provide a definitive statement on how good is

good enough. It is for this reason that we included the comparison for NO2 measured by two identical reference grade instruments in order to provide some reference for our comparisons. In order to try and expand on this point, we have added some text to the manuscript (page 3, lines 2 - 8) describing the standards set for reference grade instrument performance as set by the EU Directive 2008/50/EC, Annex 1(EU, 2008). Conforming to these standards is an obvious target for low cost sensor measurement performance, but that is not to say that reduced accuracy observations do not hold value providing the uncertainties are quantified.

C1. Changes to manuscript Page 3, lines 2 - 8 For reference monitors in the UK, NOX, CO and O3 instruments must produce reproducible measurements for three months that are within 5% of the average for a certain concentration in the field, and results that are linear over a set range (EU, 2008). For NOX this is 0 – 2000 ppb and O3: 0-500 ppb and CO: 0 – 50 ppm to ensure that both rural and urban concentration ranges are taken into account. Although the target performance of low-cost sensors is highly application dependent, these standards do provide a benchmark for comparison and highlight the need not only for high accuracy measurements but also reproducibility over long (months) timescales. In order for low cost sensors to be used in atmospheric monitoring or research applications the uncertainty and reproducibility must be quantified across a range of likely environmental conditions.

C2. Overall, using SLR and ML techniques seems to be the largest source of improvement. Is sensor clustering even necessary?

C2. Author's response Ultimately the clustering and statistical calibration methods are performing different functions in improving sensor performance. In terms of measurement accuracy, the SLR and ML calibration algorithms provide significant improvement over simple linear regression, due to their ability to correct for the multiple cross-sensitivities on sensor signals. In contrast, the function of the clustering approach is not to improve measurement accuracy, but rather sensor reproducibility. As shown in our previous paper (Smith et al., 2017) many sensors show variability in both signal and sensitivity over timescales of days or longer. This variability is very difficult to remove through time averaging, however the lack of correlation of this noise / drift between identical sensors means it can be addressed by instead averaging over multiple sensors. The conclusion of Smith (2017) was that clustering greatly reduces medium-term random noise in the average sensor signal, thus improving confidence in sensor signals and in theory prolonging the time requirements between calibrations. The time series in Fig 1. illustrates how sensor signals drift apart over time. The plot on the left shows all six sensor signals immediately after calibration to a reference monitor, showing a tight clustering around the median value (red). The plot on the right however shows the drift in individual sensors after a period of 16 days. The use of an average sensor reduces some of this signal variability enabling a more robust calibration to be applied, using algorithms such as SLR or Gaussian Process etc. The following text was added to the manuscript to emphasise that clustering and ML are used to target different issues. Page 4, line 26 -27.

C2. Changes to manuscript The clustering approach was used to improve sensor reproducibility as previously discussed in (Smith et al., 2017), whereas the SLR and ML techniques were applied to improve sensor accuracy by correcting for cross sensitivities.

C3. Does sensor accuracy vary over the observed concentration ranges? C3. Author's response This is an excellent question by the reviewer, and we have performed additional analysis below to investigate this. For each calibration method used in the paper, the data was 25 % of the observed reference concentration range bins. The Root Mean Squared Error (RMSE) and the Normalised Root Mean Squared Error (NRMSE) were calculated for each concentration bin and the results for NO2 and OX are summarised in the tables below. The NRMSE was calculated by dividing the RMSE between the reference observations and the sensor values by the mean reference concentration for the respective bin. Table 1 and 2 nicely displayed how the different analytical techniques improved the sensor performance at different concentrations. Therefore, we

decided to include Table 1 (Fig. 3 this document) and a description of the results in the manuscript summarising the NO2 RMSE and NRMSEs. The OX summary was very similar so wasn't included, but the authors are happy to include it if the editor wishes. Table 2 (Fig 4. of this document) summaries the results for the OX analysis.

C3. Changes to manuscript- added in a Table (Fig 3) to summarise these results on page 16 of the manuscript.

Text changes: Page 9, line 18 - 26 The RMSE and NRMSE was calculated after the application of SLR and ML for different reference concentration ranges to indicate where the greatest improvement of the sensor data occurred (see, Table 2). The RMSE and NRMSE (calculated by dividing the RMSE by the mean of the concentration bin) were determined between the reference NO2 observations and the sensor values for four equally spaced reference concentration bins. The ML techniques produced the greatest improvements in the concentration estimates for the lower concentrations of the target measurand where the effect of cross interreferences is more significant. The BRT and GP in particular displayed large improvements for the lower NO2 reference observations. At the higher concentrations of NO2, the ML algorithms displayed less improvement, where the conditions were outside those of the training data variable space. This was very noticeable for the BRT algorithm due to its inability to extrapolate.

Page 10 line 8 - 11. The NRMSE was calculated for 4 equally sized reference OX concentration bins for each analytical method used, in a similar manner to Table 2 for NO2. The NRMSE improved for SLR and the ML algorithms across all concentration ranges, with BLR and BRT optimal for reducing the error estimate the most. The error was the highest at the higher OX concentrations for BRT, which was expected due to BRTs inability to extrapolate.

C4. Technical Comments p1 l30. 'site'->situated C4. Authors response The wording has been changed on page 1, line 32.

C5 . p2 l8, l19, l21... Check reference parentheses throughout p4 l25. C5. Author's response Removed the extra brackets between multiple references and inserted a semi-colon to differentiate two citations for the same reference.

C6. Also Hagan et al. AMT 2018 C6. Author's response Added the reference into the manuscript on page 5, line 3 as it was relevant to the manuscript.

References EU: Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, Eur. Union, 1–62, 2008.

Smith, K., Edwards, P. M., Evans, M. J. J., Lee, J. D., Shaw, M. D., Squires, F., Wilde, S. and Lewis, A. C.: Clustering approaches that improve the reproducibility of low-cost air pollution sensors, Faraday Discuss., 00(0), 1–17, doi:10.1039/C7FD00020K, 2017.

Please also note the supplement to this comment:
https://www.atmos-meas-tech-discuss.net/amt-2018-285/amt-2018-285-AC1-supplement.pdf

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-285, 2018.

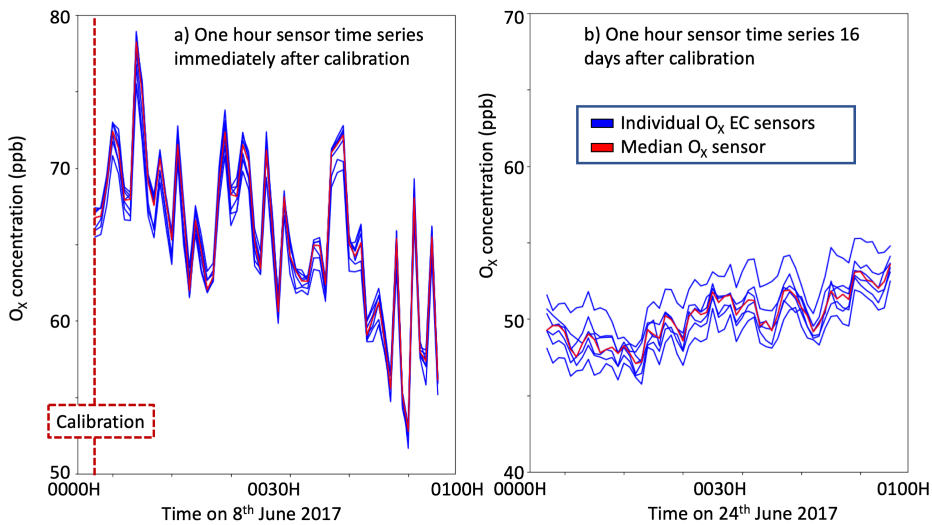

Fig. 1. Six individual OX EC (blue) with the median OX EC (red), a) immediately after SLR calibration with the reference observations and b) 16 days after the calibration.

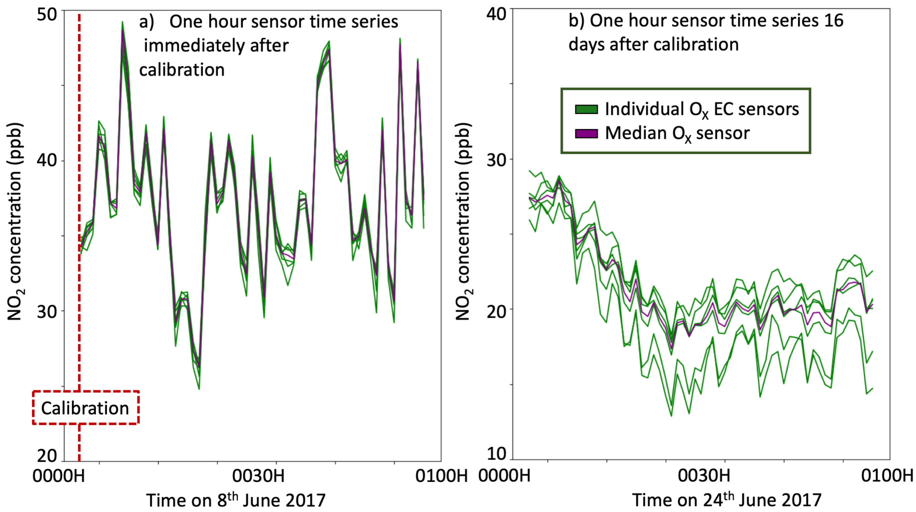

Fig. 2. Six individual NO2 EC (green) with the median NO2 EC (purple), a) immediately after SLR calibration with the reference observations and b) 16 days after the calibration.

	NRMSE of Reference vs. NO_2 concentration estimate (RMSE / ppb)				
Concentration range as a % of the max. conc. of reference NO_2	Median	SLR	BLR	BRT	GP
0 - 25 %	1.04 (20.7)	0.59 (11.7)	0.32 (6.3)	0.28 (5.6)	0.29 (5.8)
25 - 50 %	0.69 (47.5)	0.19 (13.3)	0.12 (8.2)	0.22 (15.2)	0.11 (7.9)
50 - 75 %	0.72 (94.9)	0.23 (30.8)	0.26 (34.6)	0.55 (72.5)	0.26 (33.5)
75 - 100 %	0.85 (153.1)	0.10 (17.4)	0.10 (18.8)	0.67 (120.0)	0.10 (18.2)

Fig. 3. Table 1 The NRMSE and RMSE between the NO2 reference and sensor data sets at different concentrations ranges. For each calibration method used in the paper, the data was binned into 25% of the observe

	NRMSE of Reference vs. O_X concentration estimate (RMSE / ppb)				
Concentration range as a % of the max. conc. of reference O_X	Median	SLR	BLR	BRT	GP
0 - 25 %	0.21 (11.0)	0.16 (8.4)	0.10 (5.4)	0.12 (6.0)	0.18 (9.2)
25 - 50 %	0.30 (26.4)	0.12 (10.2)	0.11 (9.4)	0.11 (9.7)	0.14 (12.4)
50 - 75 %	0.36 (50.4)	0.12 (16.3)	0.12 (16.1)	0.10 (14.0)	0.16 (22.4)
75 - 100 %	0.52 (116.1)	0.20 (44.7)	0.26 (58.0)	0.49 (110.9)	0.27 (60.6)

Fig. 4. Table 2. The NRMSE (and RMSE) between the OX reference and sensor data sets at different concentration ranges.