Dr Peer Nowack

University Lecturer
Atmospheric Chemistry and Data Science
Climatic Research Unit
School of Environmental Sciences
University of East Anglia
Norwich, NR4 7TJ
United Kingdom
p.nowack@uea.ac.uk

Imperial College Research Fellow
Grantham Institute
Department of Physics
Data Science Institute
Imperial College London
London, SW7 2AZ
United Kingdom
p.nowack@imperial.ac.uk

24th June 2021

Atmospheric Measurement Techniques Manuscript amt-2020-473

We thank Dr Malings for the overall positive feedback and helpful comments, which have helped to further improve our manuscript. We address each of his comments (*italic font*) point-by-point below.

Reply to Reviewer #1:

This paper presents an application of machine learning techniques to the calibration of data from low-cost sensors. It particularly focuses on (1) the effects of different subsets and combinations of inputs, including transformed inputs, on the resulting calibration performance, (2) the comparison of regularized Ridge regression to un-regularized regression, and of Gaussian Process regression to the more common Random Forest approach, and (3) the transferability of performance between locations, especially in cases where there is a greater range of concentrations at one location as compared to another. Overall, the paper presents and explains the issues well, will a good description of the motivation and methods. The discussion of the machine learning approaches in particular shows a good understanding of these techniques. The results presented are mainly in line with previous work in this area and help to highlight and provide greater context for some of these issues, in particular the question of transferability of calibrations between locations. The paper is fairly well written, and I believe it is suitable for publication, provided some steps are taken to clarify certain aspects and statements (as outlined below).

Specific Comments

Lines 1-8: This background information can probably be condensed to 1 or 2 sentences within the abstract.

We agree. We have shortened this text to:

"Low-cost air pollution sensors often fail to attain sufficient performance compared with state-of-the-art measurement stations, and typically require expensive laboratory-based calibration procedures. A repeatedly proposed strategy to overcome these limitations is calibration through co-location with public measurement stations."

Lines 18-19: For the sentence "In particular, none of the methods is able to extrapolate to pollution levels well outside those encountered at training stage.", I believe it should say ". . .none of the non-parametric methods. . ." or ". . .none of the non-linear methods. . .", since you later state that the linear Ridge regression is able to extrapolate. Alternatively, if you mean that the methods are able to extrapolate but may not do so well, I suggest phrasing that as ". . .none of the methods is able to extrapolate well to pollution levels. . .".

We can see how this statement could be misunderstood. Indeed, Ridge regression can extrapolate - to a degree - outside the training range, but is ultimately also limited in this sense. In other words, Ridge can also not extrapolate "well outside" the training range. Gaussian Process regression is also better at extrapolating than are Random Forests (cf. Figure 6). We agree that our current wording is difficult to follow for someone who has not read the paper yet. To clarify this, we have rephrased to:

"We also highlight several key limitations of the machine learning methods, which will be crucial to consider in any co-location calibration. In particular, all methods are fundamentally limited in how well they can reproduce pollution levels that lie outside those encountered at training stage. We find, however, that the linear Ridge regression outperforms the non-linear methods in such extrapolation settings. GPR can allow for a small degree of extrapolation, whereas RFR can only predict values within the training range. This algorithm-dependent ability to extrapolate is one of the key limiting factors when the calibrated sensors are deployed away from the co-location site itself."

Line 76: I would recommend removing the "1 – residual sum of squares/total sum of squares" part of this sentence, as this is more of a calculation formula than a definition of the term. Instead, I would suggest including this as a numbered equation in your paper, e.g., in the results section.

We have removed this text in response and have added the equation to section 3.1 (page 13 of the revised manuscript).

Figure 2: I believe this is the first time "AirPublic" is mentioned in the context of the sensor nodes. I suggest that this be explicitly stated as the maker of the sensor nodes in the body of the paper where the sensor nodes are described.

In section 2.1, we now write:

"Depending on the measurement location, we deployed one set or several sets of air pollution sensors, and we refer to each set (provided by London-based AirPublic Ltd) as multi-sensor 'node'."

Line 301: Same comment as for line 76.

Done.

Line 316: "logarithmic plus exponential" can be ambiguous, i.e., did you use both as separate inputs, or add them together? I would instead phrase this as "both logarithmic and exponential".

Thank you for pointing out. We have changed the text accordingly.

Lines 386-389: I would also suggest mentioning the importance of measuring potential interferents, like ozone and NO, since this seems to be indicated by your results as well and is a separate issue from the temperature and humidity effects.

We agree and have changed to paragraph to:

"In summary, using all sensor signals in combination is a robust and skilful set-up for our NO_2 sensor calibration and is therefore a prudent choice, at least if one of the machine learning methods is used to control for the curse of dimensionality. In particular, the B43F sensor is important to consider in the calibration, but further calibration skill is gained by also considering

environmental factors, the presence of interference from ozone and NO, and additional NO_2 devices.”

Lines 449-451: It is not clear to me why dividing the data based on time in this way would guarantee the largest variability in pollutant concentrations.

The chosen sub-periods contain the highest PM10 concentrations across the total measurement periods for either location. For Figure 7a, the Min-Max possible range is well reflected and for location CR9 (Figure 7b), we sample both the maximum (at around 650 hours) and close to minimum values (at around 980 hours). However, it is true that slightly different choices could have been made (i.e. where to center the 400-hour period). We therefore agree that it is better to change the text to:

“To mitigate issues related to extrapolation (Figure 6), we selected the last 400 hours of the time series for location CarPark (Figure 7a), and hours 600 to 1000 of the time series for location CR9 (Figure 7b). This way we still emulate a possible minimal scenario of 400 consecutive hours of co-location while also including near maximum and minimum pollution values within our training data (given the available measurement data). We note that alternative sampling approaches, such as random sampling with shuffling of the data, could lead to artificial effects at validation and testing stage, because of autocorrelation effects that could inflate apparent calibration skill.”

Where we now also highlight disadvantages of e.g. data shuffling and random sampling.

Lines 510-512: You should specify whether this statement (in particular the concentration range given) refers to NO2, PM10, or both.

Done.

Another suggestion I forgot to include: I would suggest considering changing the title of the paper. Currently, it is very generic. You may want to focus on some specific result and allude to that in the title, for example, relating to how well these methods can be applied beyond their calibration site, or how the some techniques (linear/Ridge regression) are better suited for extrapolation beyond the training data range.

For the revised manuscript, we have changed the title to:

“Machine learning calibration of low-cost NO_2 and PM10 sensors: non-linear algorithms and their impact on site transferability”

We have also taken over all technical corrections suggested by Dr Malings.