



EGUsphere, referee comment RC1
<https://doi.org/10.5194/egusphere-2022-200-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-200

Anonymous Referee #1

Referee comment on "Identifying optimal co-location calibration periods for low-cost sensors" by Misti Levy Zamora et al., EGU sphere,
<https://doi.org/10.5194/egusphere-2022-200-RC1>, 2022

General Comments

Overall the paper is very well written and presents its conclusions clearly. I recommend it for publication following some minor additions and corrections noted below.

My biggest concern is that, using a fixed total amount of sensor data, as the calibration period is increased, the evaluation period is decreased. Comparing results across calibration periods of different lengths could potentially be misleading. Ideally, calibration periods of the same length would be used in all cases; however, this is practically difficult with limited data. A comment to this effect should be added in the paper as a caveat for the presented results.

While the use of linear regression approaches to calibration is a reasonable way to approach the analysis, it is by no means the only approach to low-cost sensor calibration. In particular, methods for accounting for the non-linear impacts of various predictors, including quadratic regressions and various machine learning approaches, may be more appropriate. While it is not necessary to exhaustively investigate these here, some mention of these alternative approaches should be made, for example as a topic of future work. Similarly, while using simple "coverage" as a metric to test the appropriateness of the calibration period to the evaluation period is a reasonable first approach, more sophisticated comparisons of the statistical distributions of predictors across these periods could also be applied in future analysis and might also be mentioned here.

I would strongly suggest that the datasets used for this analysis be made publicly available if this has not already been done, and the data repository be linked in the paper. This will facilitate other researchers investigating the dataset to determine appropriate calibration strategies for their particular needs.

Specific Comments

Line 16: "mm" should be micrometers.

Line 18: "randomly" should be "randomly selected".

Line 80: "was" should be "were".

Line 90: What was the increment of the calibration durations? E.g., "ranging from 1 to 180 consecutive days in X day increments". This can be inferred from the presented results, but it is best to explicitly state it as well.

Line 115: Please elaborate on what is meant by "time", e.g., hour of the day, day of the week, age of the sensor, etc. Based on later comments I assume it is the age of the sensor, but this should be specified.

Line 202: "2.5" should be subscripted.

Figure 4: For completeness, plots similar to these should be created for all sensors and all predictors and included in the supplemental information.

Line 269: Remove "compound".

Table 3 Caption: The bottom of the caption may be cut off. Also, the "required conditions" should be specified here.

Line 302: "was" should be "were".

Line 311: Remove "and".

Line 317-319: Regarding the statement "...the co-location duration was not as predictive of data accuracy..." this might not be entirely supported by your results as you present

them, since you do not explicitly perform a meta-analysis of using either duration or coverage as a predictor of performance metrics. You might consider doing such an analysis, or slightly rephrasing this statement.

Line 334: The "<link>" is missing here.