

Specific Points

1. Abstract

- 1.1. Line 3: remove “(empirical)” in favour of just “statistical” or “empirical”
- 1.2. Line 10: clarify the bias metric used similar to how it was done for MAE
- 1.3. Line 12: “using our OEM retrieved allowed” - replace retrieved with retrieval?
- 1.4. Line 13: include the accepted ranges in parenthesis here if possible

2. Introduction

- 2.1. Line 21: add proper citation for map.purpleair.com and move to references
- 2.2. Line 22 and 24: It is incorrect to say the PurpleAir monitor makes particulate “measurements” - they estimate the concentrations based on measurements of light scattering and an assumed particle composition
- 2.3. Line 25: the term “Low-cost sensors” is used but not defined for the reader
- 2.4. Line 35: Barkjohn et al. expanded on this work in this publication:
<https://www.mdpi.com/1424-8220/22/24/9669>
- 2.5. Line 44: Suggest including <https://amt.copernicus.org/articles/15/3315/2022/> as a reference as it compares many of the cited models with sites across Canada/USA
- 2.6. Line 45: remove “(also called empirical)” in favour of just “statistical” or “empirical”
- 2.7. Suggest adding paragraph(s) describing the OEM method and hygroscopic growth

3. Methodology

- 3.1. What time period is the data from? I believe 2021 based on the comment in this section on data removed for a period in August 2021, but it should be clearly stated what date range was used.
 - 3.1.1. Suggest adding a study site paragraph to start this section.
- 3.2. Section 2.1 belongs mostly in the introduction as it is a review of what was done in another study. Lines 75 and 76 should be expanded on here instead with specifics on what was done in this study
- 3.3. Line 74: what total is “or about 30%” in reference to?
- 3.4. Line 79: citation for PurpleAir is missing
- 3.5. Line 80: having two sensors is for precision, not accuracy
- 3.6. Line 84: clarification is needed whether the A/B comparison was done before or after averaging to daily averages
- 3.7. Line 86: “typically eliminated about 3% of measurements” - unclear what this 3% applies to, each day?
- 3.8. Line 90: “due to internal heating and insolation effects”
- 3.9. Line 93: In my experience, the bias in PA RH has a diurnal cycle due to insolation effects on the temperature within the unit. Given this, adding 21% to PA RH may be okay for this dataset but it may not be transferable between sites/time/averaging periods. This should be emphasised in the discussion/conclusion, and I would suggest including justification for the 21% adjustment in the results (ex. a scatter of RH from both monitors before and after)
- 3.10. Line 94: “were about 2C high” is vague
- 3.11. Line 101: link should be cited properly and moved to the reference list
- 3.12. Lines 108 - 121: this paragraph belongs in the introduction

- 3.13. Line 108: suggest improving the paragraph transition here. This paragraph should start with the *measurement/monitor* differences, not the *price* differences.
- 3.14. Line 111: the plantower sensors independently produce particle counts and concentration estimates using two separate proprietary algorithms (ie. concentrations are not derived from the particle counts). See:
<https://amt.copernicus.org/articles/13/6343/2020/>
- 3.15. Line 117: I don't believe "swelling" is the correct term. The water accretes on the surface of the particle; swelling implies the absorption of water by the particle
- 3.16. Line 119: "detect higher concentrations of larger-diameter particles" is incorrect - due to the hygroscopic growth of the particles the sensors detect higher scattering and estimate a higher concentration. The assumed particle density does not change.
- 3.17. Lines 123 - 135: belongs in the introduction
- 3.18. Line 156: what bias metric was used?

4. Results

- 4.1. Line 163: Remove "One model parameter is the particle diameter."
- 4.2. Line 171: define "reasonable ambient range"
- 4.3. Figure 1:
 - 4.3.1. suggest increasing font size as it is difficult to read even with zooming
 - 4.3.2. Make the x/y axes have the same limits so seasons can be compared visually. As it is now, the Raw fall values look biased higher than that for the winter panel, however the winter axes go out to 60 ug/m3 whereas the fall axes go out to 40 ug/m3
 - 4.3.3. "OEM" and "MLR" need to be defined in the figure caption
- 4.4. Line 175: what form was the MLR? And what were your coefficients? This needs to be discussed in the methods and results
- 4.5. Lines 177-187: use these sections to describe the results displayed in the figures, not to describe the presentation of the figure. For example, "The raw PurpleAir observations tended to be biased higher than the Ministry PM2.5, which worsens as concentrations increase."
- 4.6. Line 188: "The physical calibration has a tendency to over-correct at high relative humidity" - is this true? Figure 2 has high and low RH on both sides of the 1:1 line for all seasons. I don't think this is sufficient justification to disregard higher humidity values.
- 4.7. Figure 2:
 - 4.7.1. Increase font size as it is difficult to read without zooming
 - 4.7.2. The figure caption should state that these are daily observations
 - 4.7.3. The colour scale used makes it difficult to see the mid-range values, this makes the extremes visually stand out. I would suggest binning the RH values into low (<55%) moderate (55- 65) and high (> 65%) and using 3 easily differentiable colours. This would also make it more clear what "higher values of relative humidity" (Line 190) entails.
- 4.8. Line 195: the PurpleAir spec sheet lists an accuracy tolerance of 3%
(<https://www2.purpleair.com/products/purpleair-pa-ii>)
- 4.9. Figure 3:
 - 4.9.1. Replace "PM2.4" with "PM2.5" for both (a) and (b)

- 4.9.2. The error bars are not clearly visible on (b) - this should be mentioned in the results section
- 4.9.3. (b) has "Month" as an x axis label not "Season"
- 4.10. Additional recommendations:
 - 4.10.1. A daily mean time series of the observation data would benefit this paper, especially for justifying the bias-adjustment of 21% for the PurpleAir RH.
 - 4.10.2. Instead of Figure 2, try RH on the x axis and daily mean bias on the y. That could more clearly show if high RH has an effect on the PurpleAir bias.

5. Discussion

- 5.1. Lines 199-201: this is an important finding that could be highlighted in the abstract
- 5.2. Lines 209 - 211: something like this would be great for the results section
- 5.3. Line 217: particle composition/age also varies seasonally as well and is an important factor. The optical properties and hygroscopicity of particulates from a residential wood stove that have not been airborne for long will differ from those from wildfire smoke that travelled from western Canada to eastern Canada. Another factor could be the concentration ranges within each season - concentrations of PM2.5 tend to be lower in Spring/Fall due to less periods of stagnation
- 5.4. Line 218: "biggest" is vague and it is unclear to me what supports this claim
- 5.5. Line 220: "are more strongly affected" is vague
- 5.6. Line 228: it is not clear to me how the apparent seasonal differences in overcorrection at high humidities indicates that the physical model does not perform well at high humidities.
- 5.7. Line 231: Table 2 and the summary of it belongs in the results, not discussion
- 5.8. Lines 232-233: "The statistically-calibrated data consistently had no bias" - this seems suspect to me (potentially a result of not splitting training/testing data), and is vague. Where was this presented in the results?
- 5.9. Table 2: What about the raw R squared? Did the physical/statistical models improve/worsen the correlation?
- 5.10. Line 236: "greatest" is subjective and was not statistically evaluated - recommend removing this sentence
- 5.11. Line 237: "it is known" should be replaced with "we noticed" and the sentence should be clarified that it is for this site/region - PM2.5 can have dramatic spatial variation
- 5.12. Additional limitations
 - 5.12.1. It is a concern that only one year for a single PurpleAir/Ministry pair was used- are these results transferable to other areas and concentration ranges?
 - 5.12.2. The data were not split into training and testing datasets, likely overfitting the model and producing overoptimistic performance measurements.
 - 5.12.3. The concentration range is fairly moderate, it would be interesting to see the performance during wildfire smoke events where daily mean PM2.5 can exceed 100 ug/m3 (more than twice of what was observed at this site/period)

6. Conclusions

- 6.1. Line 258: include the reasonable ranges in parenthesis here if possible
- 6.2. Line 260: use the updated url for aqmap (<https://aqmap.ca/aqmap>), and move to references with a proper citation