https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

Retrieval of Bulk Hygroscopicity From PurpleAir PM2.5 Sensor Measurements

Jillian Psotka[1,*], Emily Tracey[1,*], and Robert Sica[1,*]

[1]Department of Physics and Astronomy, The University of Western Ontario, London, Canada
[*]These authors contributed equally to this work.

Correspondence: Robert Sica (sica@uwo.ca)

Abstract. PurpleAir sensors offer a unique opportunity for a large-scale and densely populated array of sensors to study surface air quality. While the PurpleAir sensors are inexpensive and abundant, they require calibration against a validated coincident measurement to ensure the quality of the measurement. Traditionally, this calibration is performed using statistical (empirical) methods. We propose a method to extend the aerosol properties determined by PurpleAir sensors to include estimates of the hygroscopic growth of aerosols using a novel calibration method based on the optimal estimation method (OEM). The hygroscopic growth can be estimated during calibration due to the calibration's sensitivity to relative humidity, which influences the measured size distribution of the aerosols. Our OEM-based retrieval for calibration includes an estimation of the bulk hygroscopicity. By employing the physically-based calibration with the London's Ministry of the Environment, Conservation and Parks site calibrated measurements, the average daily Mean Absolute Error (MAE) of the PurpleAir PM2.5 measurements decreased from 5.58 $\mu g/m^3$ to 1.68 $\mu g/m^3$, and the average daily bias from 4.75 $\mu g/m^3$ to -0.23 $\mu g/m^3$. This improvement is comparable to the improvement seen using conventional statistical methodologies. In addition to calibration, using our OEM retrieved allowed us to estimate seasonal bulk hygroscopicity values ranging from 0.33 to 0.40. These values are consistent with the accepted ranges of bulk hygroscopicity values determined in previous studies using sophisticated air quality measurement instruments.

1 Introduction

Home air quality monitors are becoming increasingly popular both for the general public and in the atmospheric science community due to their low cost and ease of use. One notable inexpensive air quality sensor is the PurpleAir PA-II sensor, which can be utilized both indoors and outdoors and is priced at approximately $300 USD. These sensors connect to WiFi and provide users with real-time air quality and ambient condition readings at two-minute resolution. Each PurpleAir sensor can be integrated with the PurpleAir Network, where data from over 30,000 sensors across the globe are publicly available (https://map.purpleair.com/).

PurpleAir sensors employ light scattering photometry to make particulate measurements. Every two minutes, an air sample is drawn through the instrument and the optical laser beam interacts with the particulates in the sample. The scattered light is

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

measured by a photocell detector plate, which converts the detected photons to a measurement of the number of particles in the sample and their sizes.

Low-cost sensors have been successfully applied not only to home-use but for scientific use. Bi et al. (2020) incorporated PurpleAir measurements into large-scale $PM_{2.5}$ modelling. They paired 54 PurpleAir sensors to nearby U.S. Environmental Protection Agency Air Quality System (AQS) stations across California. They calibrated the PurpleAir sensors using a generalized additive model that included linear corrections for relative humidity, temperature, and sensor operating time. The calibrated PurpleAir data was then combined with the AQS data to create high-resolution daily $PM_{2.5}$ estimates. In their model, the contribution of data from each individual PurpleAir sensor was down-weighted depending on the residual errors. They found that the model that incorporated PurpleAir measurements was more effective at modelling $PM_{2.5}$ predictions than a strictly AQS-based model. This study is promising in terms of wider applications of PurpleAir data use and could have been improved further with more accurate PurpleAir calibration.

Barkjohn et al. (2021) developed statistical calibrations using 50 PurpleAir sensors at 39 unique sites in the United States. They tested 15 linear and multi-linear models of varying complexity, with a mixture of additive and multiplicative interaction terms. They used only parameters that were provided or could be explicitly calculated from PurpleAir measurements so as to make their calibration applicable to any PurpleAir site, regardless of proximity to reference instruments. Their study found that PurpleAir sensors' over estimations of $PM_{2.5}$ readings could be adequately corrected by a multiple linear model of the form

$$\mathrm{PM}_{2.5,\mathrm{corr}} = a\mathrm{PM}_{2.5,\mathrm{meas}} + b\mathrm{H} + c, \tag{1}$$

where a, b, and c are constant coefficients and H is the relative humidity as measured by the sensor. Increasing the complexity of the model did not have significant advantages for the calibration. They were successful in creating a single nationwide calibration that could be applied to their PurpleAir sensors and reduce errors.

Other studies (Ardon-Dryer et al. (2020), Magi et al. (2019), Malings et al. (2020), Tryner et al. (2020)) have also applied statistical (also called empirical) methods to calibrate the PurpleAir sensors to standard research-grade instruments. In this study, we used an inverse modeling technique called the Optimal Estimation Method (OEM) to calibrate the PurpleAir measurements using a Thermo Scientific SHARP (Synchronized Hybrid Ambient Real-time) Model 5030 particulate monitor and a physical model of hygroscopic growth factor given by Malings et al. (2020). The hygroscopicity of bulk aerosol was retrieved during the calibration, providing a possible advantage of this technique over previous calibration methods.

2 Methodology

2.1 The Physical Model

Malings et al. (2020) calibrated nine PurpleAir sensors using two different correction methods, one physics-based and one statistical. The physical model was based on the hygroscopic growth of different aerosols and the composition of the air pollution in Pittsburgh, Pennsylvania (USA). The hygroscopic growth factor of $PM_{2.5}$ which quantifies the hygroscopic growth

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

of the aerosols is calculated as

$$\mathrm{f}(T,H) = 1 + \kappa_{\mathrm{bulk}} \frac{w(T,H)}{1 - w(T,H)}\,, \tag{2}$$

where T and H are temperature and relative humidity, κ_{bulk} is the hygroscopicity of bulk aerosol, and w is the water activity. The hygroscopicity of bulk aerosol was calculated as the sum of the fractional component, x_i, of each main aerosol multiplied by its hygroscopicity, κ_i:

$$\kappa_{\mathrm{bulk}} = \sum_{i=1}^{n} x_i \kappa_i\,. \tag{3}$$

Malings et al. (2020) used four main aerosols, carbonaceous mass, sulfate, nitrate, and ammonium. The fractional composition of these aerosols varies by location, but these are consistently the most abundant $PM_{2.5}$ aerosols in the United States (Bell et al., 2007).

The water activity was calculated as a function of temperature and relative humidity as

$$w(T,H) = H \exp\left(\frac{4\sigma_w M_w}{\rho_w R T D_p}\right)^{-1}, \tag{4}$$

where σ_w, M_w, and ρ_w are the surface tension, molecular weight, and density of water. R is the ideal gas constant, and D_p is the average particle diameter. A linear correction was also applied to account for the unknown factory calibration of PurpleAir sensors. The total correction was as follows:

$$\mathrm{PM}_{2.5,\mathrm{corr}} = a\left(\frac{\mathrm{PM}_{2.5,\mathrm{meas}}}{\mathrm{f}(T,H)}\right) + b, \tag{5}$$

where a and b are constant coefficients, $\mathrm{PM}_{2.5,\mathrm{corr}}$ and $\mathrm{PM}_{2.5,\mathrm{meas}}$ are the corrected and measured $PM_{2.5}$, respectively, and $\mathrm{f}(T,H)$ is defined in Eq. 2. They tested this model using a statistical correction which was a multiple linear correction with terms for relative humidity, air temperature, and dew point temperature. They found that the two correction approaches yielded comparable improvements on $PM_{2.5}$ readings. Large uncertainties were still present in hourly-averaged readings (mean absolute errors 3-4 μg/m^3 or about 30%) but yearly-averaged readings were more accurate (errors less than 1 μg/m^3). The Malings et al. (2020) work established that their physics model plus a constant term is reasonably complete, and is the basis for the forward model we will use with our OEM method.

2.2 Treatment of the Measurements

2.2.1 PurpleAir Sensor

Measurements from all public PurpleAir sensors are available for download from the PurpleAir network map. Each sensor has two particle counters, Channel A and Channel B, that report independently for purposes of accuracy. The data are provided as unfiltered, two-minute readings for each channel. Daily averages of the $PM_{2.5}$, relative humidity, and temperature measurements were made for our London site on The University of Western Ontario campus (43.01°latitude, -81.27°longitude, 258 m above sea level).

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

$PM_{2.5}$ readings that had discrepancies between Channels A and B of more than 4 $\mu g/m^3$ or more than 25% of their average were removed. Any readings that only had a value from one channel were also removed. This procedure typically eliminated about 3% of the measurements. It should be noted that there were some days, and in one case over three consecutive weeks (August 2021), where all measurements were discarded to due to clearly erroneous readings. In some cases the cause of these errors remain unknown, and in others there were insects or debris in the sensor that had to be cleaned out.

Relative humidity and temperature averages were compared to measurements from the Environment and Climate Change Canada weather station at the London International Airport, about 12 km from the PurpleAir sensor. Due to internal heating, PurpleAir temperature readings can be up to 5.3°C higher and humidity readings up to 24.3% drier than ambient conditions (Barkjohn et al., 2021; Malings et al., 2020). The relative humidity readings from the PurpleAir were consistently about 21% lower than the Airport station values, so a simple correction of adding 21% to each PurpleAir relative humidity measurements was made. The PurpleAir temperature readings were about 2°C high, which was not a large enough discrepancy to require correction, since the effect of temperature in the forward model is primarily through the relative humidity.

To perform the calibration, the measurements were split into four seasons: Spring (March - May), Summer (June - August), Fall (September - November), and Winter (December - February). This binning was chosen to investigate seasonal variance in PM composition on the bulk hygroscopicity.

2.2.2 Ontario MECP Validation Measurements

Measurements from all Ontario Ministry of the Environment, Conservation and Parks (MECP) air quality sites are available online (http://www.airqualityontario.com/history/summary.php). They provide hourly readings of ozone, $PM_{2.5}$, and nitrogen dioxide. For our reference data, we used readings from the London Ambient Air Monitoring Site (42.97°latitude, -81.20°longitude, 244 m above sea level), which is about 6 km away from the PurpleAir sensor. The MECP $PM_{2.5}$ readings are given as integers values. To test that this rounding was not impacting our calibration, after we believed we had a robust calibration, we applied various rounding schemes (floor, half round up, and ceiling) to the PurpleAir measurements and redid the calibrations. None of these rounding choices had a significant impact on the calibration results.

2.2.3 Comparison of the Measurement Instruments

There are two main factors that account for the price difference between the PurpleAir sensor and its research-grade counterpart at the MECP site, the SHARP 5030. The first is the ability of the SHARP 5030 to directly measure the mass of the particles through beta attenuation. Since the PurpleAir instrument does not have this technology, it must make assumptions and estimations in order to convert the photometric particle count (measured in particles per decilitre) to mass density. These corrections are done onboard the unit using a proprietary algorithm, which introduces additional uncertainty into the data product. The same PurpleAir sensor could read too high or too low in different locations depending on the particle density in that area and how it compares to the assumptions made by the manufacturer. Hence, a correction factor must be applied to the measurements to account for the unknown difference in average particle density.

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

Additionally, PurpleAir sensors do not have a mechanism to correct for ambient humidity. Humidity can greatly affect the accuracy of PM measurements because particulates experience swelling in the presence of water (hygroscopic growth). Hygroscopic growth introduces uncertainties in low-cost sensor measurements because they artificially detect higher concentrations of large-diameter particles. The extent of hygroscopic growth depends not only on the ambient humidity, but also on the aerosol composition. Each type of particulate has its own hygroscopicity, which is a measure of its ability to absorb water. Therefore, a correction for hygroscopic growth must consider $PM_{2.5}$ composition, which varies spatially and seasonally.

2.3 The Optimal Estimation Method

OEM is an inverse method that allows the retrieval of the atmospheric state using a set of measurements and a forward model of the physical system. The forward model, $\mathbf{F}$, is represented as

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}), \tag{6}$$

where $\mathbf{y}$ is the measurement vector, $\mathbf{x}$ is the state vector, the vector which contains the retrieved quantities, and $\mathbf{b}$ are additional parameters required by the forward model.

OEM is based on Bayes' theorem which describes the calculation of conditional probabilities. Bayes' theorem allows the most likely state to be determined consistent with the *a priori* knowledge, the performed measurement, and their associated uncertainties. The cost function, $\mathbf{C}$:

$$\mathbf{C} = [\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})]^T \mathbf{S}_y^{-1} [\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})] + [\hat{\mathbf{x}} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\hat{\mathbf{x}} - \mathbf{x}_a], \tag{7}$$

is then minimized to find the optimum value of the retrieval parameters. Here $\mathbf{S}_y$ is the measurement error covariance matrix, $\hat{\mathbf{x}}$ is the normalized state vector, and $\mathbf{x}_a$ and $\mathbf{S}_a$ are the *a priori* estimate of the state vector and its error covariance matrix. In a successful OEM retrieval, one should be able to slightly vary the *a priori* estimates without having an effect on the retrieved state. This indicates that the *a priori* estimates are guiding the solution, not dictating it.

2.3.1 Implementing OEM

We implemented OEM using Qpack, a free Matlab function developed for atmospheric instrument simulation and retrieval work (Eriksson et al., 2004) using a forward model similar to that in Eq. 5,

$$\mathrm{PM}_{2.5,\mathrm{corr}} = \frac{\mathrm{PM}_{2.5,\mathrm{meas}}}{\mathrm{f}(T, H)} + c. \tag{8}$$

Our forward model does not include the constant a appearing in the form of the model described by Malings et al. (2020). The constant factor is not required in OEM as we are directly retrieving the hygroscopicity of the bulk aerosol. Descriptions of all variables, along with their affiliation to the state vector $\mathbf{x}$ or the input parameters $\mathbf{b}$, are given in Table 1. We took the measurement vector $\mathbf{y}$ to be the reference $PM_{2.5}$ readings from the MECP site, since the goal was to have the corrected data match these reference readings. We used values from Malings et al. (2020) for our *a priori* state vector.

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

Table 1. Parameters used in OEM code. The state vector $\mathbf{x}$ was retrieved by the code while the parameter vector $\mathbf{b}$ was inputted.

Variable	Description	Vector
c	constant linear term	**x**
κ_{bulk}	bulk hygroscopicity	**x**
D_p	average particle diameter	**x**
$PM_{2.5,\mathrm{meas}}$	$PM_{2.5}$ as measured by PurpleAir	**b**
H	relative humidity as measured by PurpleAir	**b**
T	temperature as measured by PurpleAir	**b**
σ_w	surface tension of water	**b**
M_w	molecular weight of water	**b**
ρ_w	density of water	**b**
R	ideal gas constant	**b**

2.3.2 Sensitivity to Model Parameters

One of the many advantages of OEM is the ability to investigate the sensitivity of the retrieval to each model parameter. For each model parameter, the error covariance matrix, $\mathbf{E}$, is given by

$$\mathbf{E} = \mathbf{G} \cdot \mathbf{J}_b \cdot \mathbf{S} \cdot {\mathbf{J}_b}^{\mathbf{T}} \cdot \mathbf{G}^{\mathbf{T}}, \tag{9}$$

where $\mathbf{G}$ is the gain matrix,

$$\mathbf{G} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}}, \tag{10}$$

$\mathbf{J}_b$ is the Jacobian for the parameter,

$$\mathbf{J}_b = \frac{\partial \mathbf{F}}{\partial b}, \tag{11}$$

and $\mathbf{S}$ is the uncertainty covariance matrix (Rodgers, 2000). In Section 3 we will investigate the sensitivity of the retrieval of the hygroscopic growth factor on the model parameters.

2.4 Statistical Metrics used to Assess the Calibration

The accuracy of the calibration was assessed using Mean Absolute Error (MAE) and bias. MAE is used to assess how well, on average, a data set agreed with the reference data. Lower values of MAE and bias indicate better agreement between our data and the reference data. We also used the adjusted coefficient of determination, adjusted-R^2, to assess how well our calibrated data correlated with the reference data.

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

3 Results

3.1 Physical Model Analysis

3.1.1 Sensitivity to Average Particle Diameter

One model parameter is the particle diameter. Using the parameter sensitivity analysis method described in Section 2.3.2, we tested the sensitivity of the model to uncertainty in average particle diameter. We found that due to the low sensitivity of the forward model to particle diameter, the choice of D_p had a negligible effect on the total retrieval error. Hence, large uncertainties in average particle diameter do not impact the retrieval, particle diameter can not be retrieved and the choice of particle diameter has negligible effect on the retrieval.

3.1.2 Sensitivity to Temperature

We investigated the behaviour of the physical model in response to environmental controls to examine the impact of using PurpleAir sensors to report ambient conditions. We considered the hygroscopic growth factor's (Eq. 2) dependence on temperature and relative humidity. We found that temperature throughout a reasonable ambient range did not impact the calibration significantly and thus, we use T as measured by the Purple Air sensor.

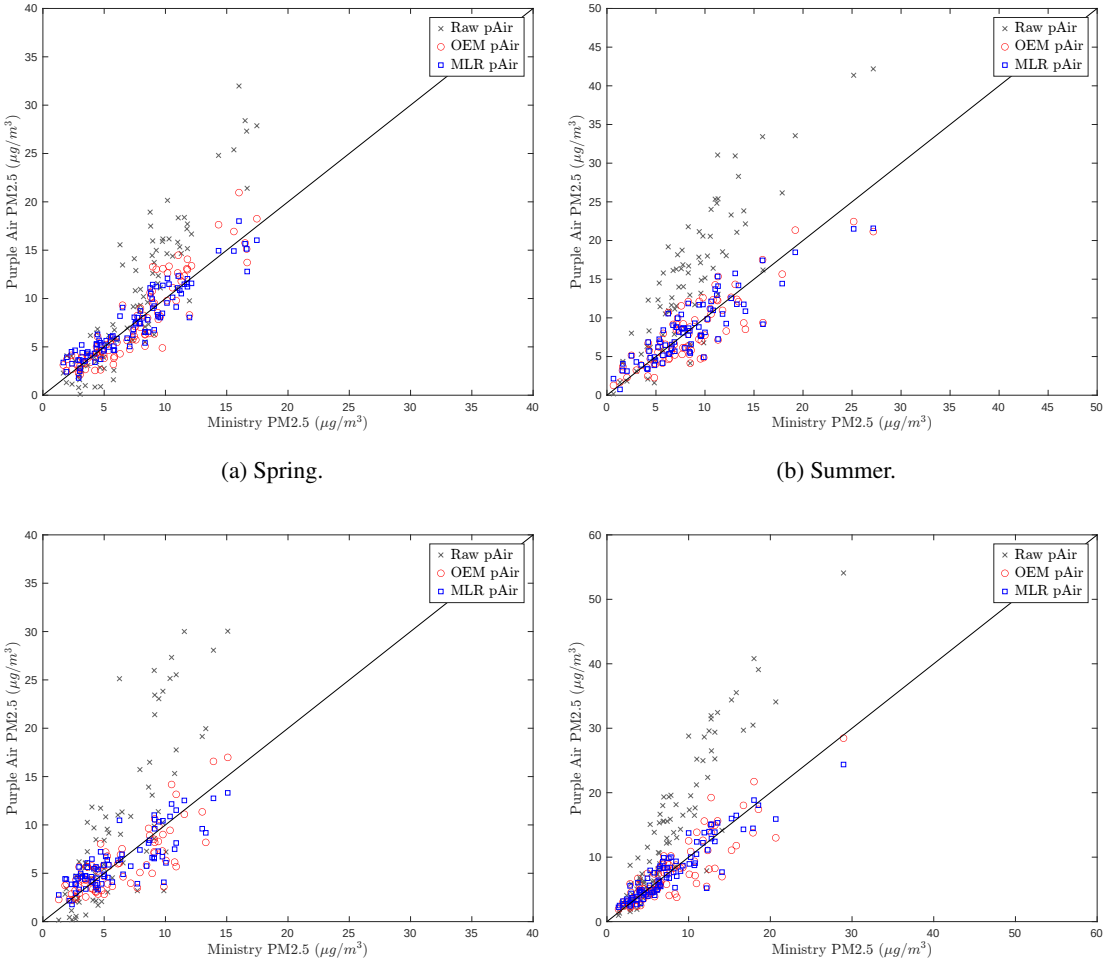
3.2 Physical Calibration Results

After using OEM to retrieve the necessary parameters, the physical model (Eq. 8) was applied to calibrate PurpleAir measurements from each season. The PurpleAir measurements were also calibrated statistically using multiple linear regression (MLR) from for comparison. The physically-calibrated, statistically-calibrated, and raw $PM_{2.5}$ measurements for each season are shown in Fig. 1. The horizontal axes are $PM_{2.5}$ readings by the MECP reference sensor, and the vertical axes are averaged $PM_{2.5}$ readings from the PurpleAir sensor. The location of each point in the plots signifies one day of $PM_{2.5}$ measurements by the MECP sensor and the PurpleAir sensor measurement for the same day averaged from two-minute readings. If the calibration had perfectly corrected the PurpleAir data, all calibrated points would be along the 1:1 line. The physical calibration succeeded in bringing the raw measurements closer to the reference measurements.

3.2.1 Impact of Relative Humidity on the Physical Calibration

To investigate the impact of relative humidity on the calibration, the physically-calibrated data from each season are plotted using a colour scale showing relative humidity (Fig. 2). As in Fig. 1, the horizontal axes are daily $PM_{2.5}$ readings by the MECP reference sensor, and the vertical axes are daily-averaged $PM_{2.5}$ readings from the PurpleAir sensor. The location of each point in the plots signifies one day of $PM_{2.5}$ measurements by the MECP sensor and the PurpleAir sensor measurement for the same day averaged from two-minute readings, with colour showing the ambient relative humidity during that day. The physical calibration has a tendency to over correct at high relative humidity; many of the high-relative humidity points have

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024



(a) Spring.

(b) Summer.

(c) Fall.

(d) Winter.

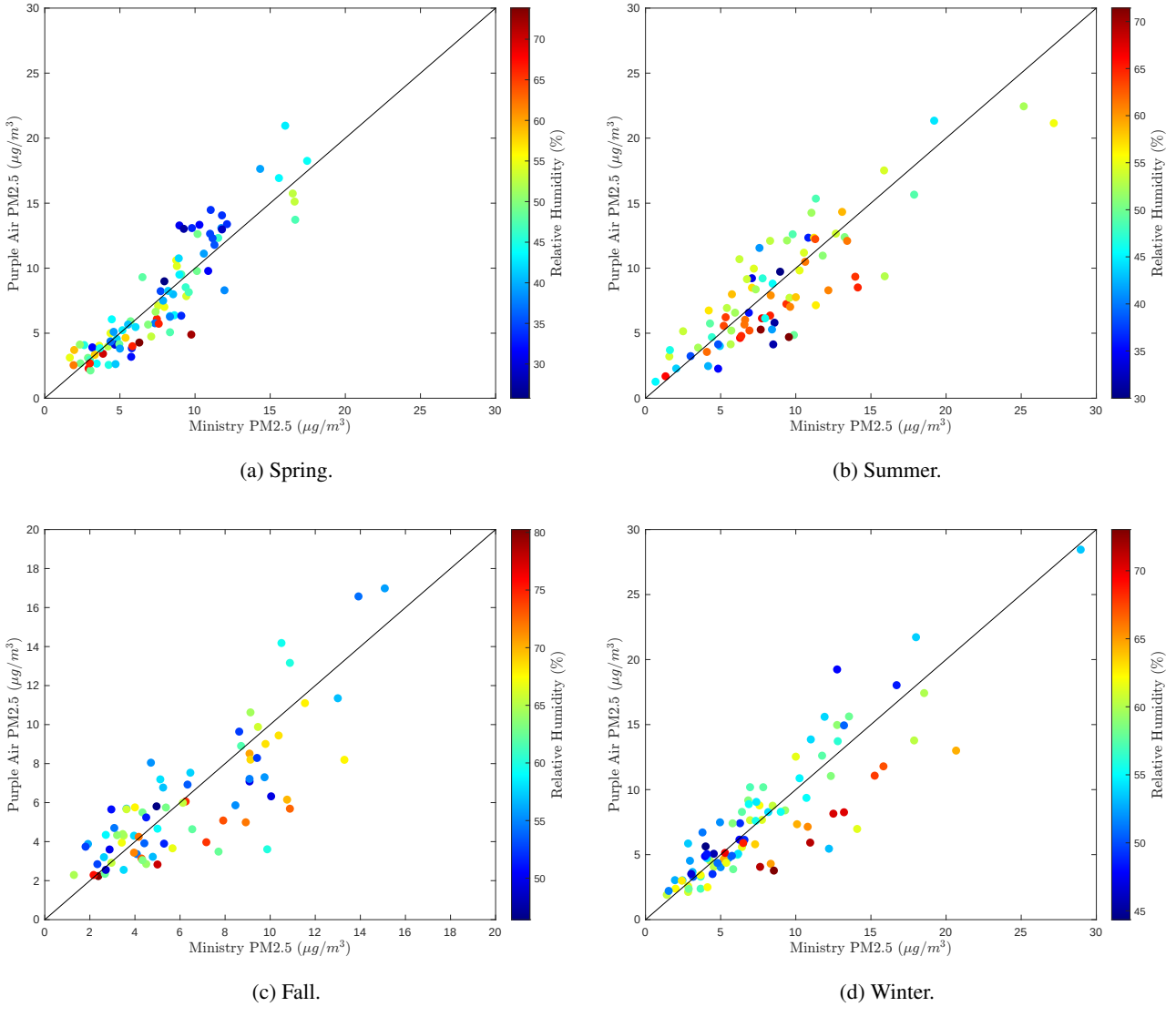
Figure 1. Daily-averaged $PM_{2.5}$ as measured by MECP vs PurpleAir. Physically-corrected data (red), raw data (grey), empirically corrected data (blue) and 1:1 line (black).

been corrected to a value that is too low and now lay far below the 1:1 line. Thus, corrections at higher values of relative humidity are insufficient.

3.3 Bulk Hygroscopicity Results

The bulk hygroscopicity of particulates was one of the parameters of the physical model retrieved through OEM. The seasonal values of bulk hygroscopicity and the associated hygroscopic growth factor retrieved are shown in Fig. 3a and 3b respectively. The error bars on these values were calculated assuming that the relative humidity was known within 2.5% uncertainty as this

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024



(a) Spring.

(b) Summer.

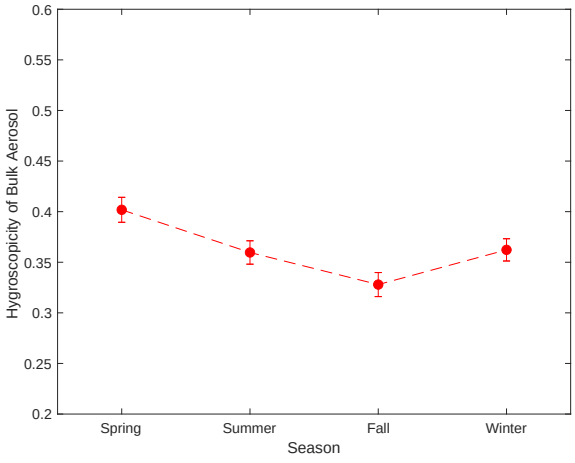
(c) Fall.

(d) Winter.

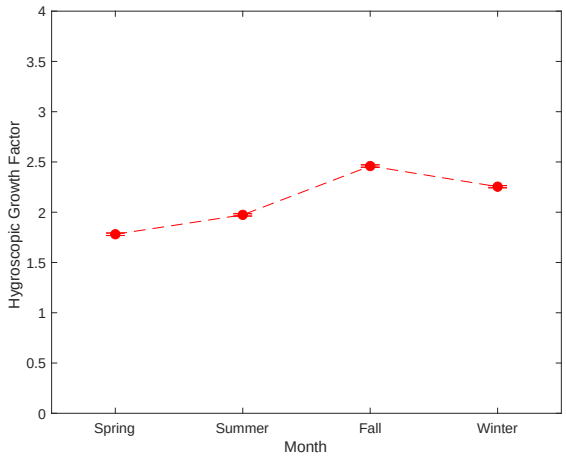
Figure 2. Physically-corrected data, colour-scaled by relative humidity. The 1:1 line is in black. There is a tendency to over correct at high relative humidity, especially during summer and winter.

195 is a reasonable uncertainty for a relative humidity sensor to achieve using measurements from a co-located, calibrated weather station. For our pilot study the relative humidity values were not of sufficiently high quality, as previously discussed.

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024



(a) Bulk hygroscopicity retrieved from PM2.4 data.



(b) Hygroscopic growth factor retrieved from PM2.4 data.

Figure 3. Bulk hygroscopicity and the hygroscopic growth factor retrieved through OEM for each season. Uncertainties are based on the assumption that relative humidity is known within 2.5% uncertainty.

4 Discussion

4.1 Updates to the Physical Model

We found that the forward model was not sensitive to the particle diameter, D_p. This result is significant because it indicates that additional instrumentation that measures D_p is not needed in order to apply the physical model. Using a general estimate of average particle diameter, in our case a value of $D_p = 200\,\text{nm}$, can be recommended for future applications of this model. This makes the physical model more accessible to PurpleAir sites that do not have the additional instrumentation necessary to estimate this value.

We found that the model was sensitive to relative humidity, so we recommend correcting the PurpleAir relative humidity measurements before use. A constant correction was sufficient for our site. We found that the model was not explicitly sensitive to temperature, so the raw PurpleAir temperature measurements may be used without correction. It should be noted that the physical model is still implicitly sensitive to temperature since it is sensitive to relative humidity, which depends on temperature.

4.2 Investigating Retrieved Bulk Hygroscopicity

The retrieved values of hygroscopicity of bulk aerosol are between 0.33 and 0.4 as shown in Fig. 3, which are consistent with values in the literature (Cerully et al., 2015; Petters and Kreidenweis, 2007). The hygroscopicity of bulk aerosol is the smallest in the Fall and largest in the Spring. This is consistent with the result from (Akpootu and Gana, 2013) that describes an inverse relationship between hygroscopicity and relative humidity. Our results show highest relative humidity in the Fall and the lowest

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

in the Spring, with the hygroscopic growth factor varying proportionally as expected. The results shows that our method may have the potential to estimate bulk hygroscopicity.

4.3 Effectiveness of Physical Calibration

As seen in Fig. 1, the raw data were more linearly-related to the reference data in the winter and summer months. This difference is likely due to seasonal changes in aerosol particle size distributions because changes in temperature impact the atmospheric reactions that are secondary sources of $PM_{2.5}$. The biggest impact on measurements due to changes in size distributions comes from more or less particles falling below the 300 nm detection limit of the sensor. During transitional seasons when weather greatly varies from the beginning of the season to the end, the PurpleAir readings are more strongly affected. For instance, the winter and summer PM2.5 data sets generally had lower daily standard deviations. This indicates that $PM_{2.5}$ levels were more steady throughout the course of each day, which could lead to better correlation with the reference instrument.

The physical calibration succeeded in reducing the MAE and bias of the raw data. Averaged over all seasons, the MAE was reduced by 69% and the magnitude of the bias was reduced by 95%. Overall, the spring data was the best of the physically-calibrated results for all of our metrics (MAE, bias, and R^2).

The physical calibration tended to over correct the raw data, as indicated by negative biases. The over-corrected data points tended to be those associated with high relative humidity (Fig. 2). This effect was especially prominent in the summer and winter data sets, indicating that the physical model does not perform well at high humidities and is perhaps not valid over a certain humidity threshold. These departures are due to the exponential behaviour of the hygroscopic growth factor. This effect is not observed in the statistically-calibrated data due to the linear nature of its relative humidity dependence.

The MAE from both models is compared with the raw data in Table 2. Although both models performed similarly, the statistically-calibrated data had slightly less MAE and higher values of R^2. The statistically-calibrated data consistently had no bias. It can be concluded that the statistical model had better overall calibration performance, while the physical model allows a new data product to be retrieved.

Table 2. MAE of all raw, physically-calibrated, and statistically-calibrated measurements along with the R^2 value of the physical and statistical calibrations.

Season	Raw MAE $\mu g/m^3$	Physical MAE $\mu g/m^3$	Physical R^2	Statistical MAE $\mu g/m^3$	Statistical R^2
Spring	3.48	1.38	0.78	1.05	0.87
Summer	6.95	1.98	0.72	1.74	0.78
Fall	5.00	1.59	0.61	1.50	0.66
Winter	6.90	1.78	0.75	1.34	0.75

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

4.4 Limitations of Our Approach

The greatest limitation of our method was that the reference instrument was not co-located with our PurpleAir sensor. From observations of the spatial spread of $PM_{2.5}$ from the PurpleAir website, it is known that regions in close proximity follow the same trends in $PM_{2.5}$ and generally have very similar readings. It is due to this that we were comfortable carrying out this calibration with our reference site approximately 6 km away. We also attempted to work around this limitation through daily averaging, which should allow two sites in the same city to reach similar values of $PM_{2.5}$, but it is still impossible to know exactly the effect that this limitation could have had on our work.

Another limitation is regarding the calibration of daily-averaged data sets. The main purpose of this calibration is to correct for effects of relative humidity. These effects cannot be fully represented when taking daily averages since relative humidity varies significantly throughout the span of 24 hours and averaging greatly smooths these variations. Therefore it is undetermined if the daily-averaged data sets fully encapsulate the model of hygroscopic growth in the presence of humidity.

Finally, this method requires high quality measurements of relative humidity, beyond what the Purple Air sensor is capable of. Best conditions for the application of this technique would include a co-located, calibrated weather station and close proximity to the calibration source.

5 Conclusions

We applied a physical model based on the hygroscopic growth of particulates to calibrate $PM_{2.5}$ measurements from a PurpleAir sensor. We calibrated daily-averaged data for one year split into four seasons. The physical calibration reduced average daily MAE and bias from 5.58 $\mu g/m^3$ to 1.68 $\mu g/m^3$ and from 4.75 $\mu g/m^3$ to -0.23 $\mu g/m^3$, respectively. The physical model tended to over correct data points with relative humidity above approximately 70%. This relative humidity bias was not seen in the statistical calibration, which reduced bias to 0 $\mu g/m^3$ and average MAE to 1.46 $\mu g/m^3$.

The physical calibration did not perform quite as well as statistical calibration, but it did provide insight into the physical model of hygroscopic growth of particulates. We found that the average particle diameter does not need to be measured and can simply be estimated for future implementations of this model. This makes the physical model applicable to more sites that do not have access to these measurements. Additionally, we were able to use OEM to retrieve reasonable values of bulk hygroscopicity. Furthermore, our method is extremely fast computationally, making it ideal to apply to "real time" situations such as air quality maps like the hourly $PM_{2.5}$ UNBC/ECCC map (https://cyclone.unbc.ca/aqmap/).

The main limitation of this study was our inability to access co-located reference measurements. We encourage researchers with dedicated air quality observatories with more sophisticated, co-located equipment to test our method and compare our bulk hygroscopicity estimates with other techniques. Furthermore, it would be of interest to investigate the inability of the physical model to represent $PM_{2.5}$ at high levels of relative humidity. Finally, it should be noted that the OEM method could in principle retrieve the individual values of the hygroscopicity as well as the bulk values as in the current retrieval.

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

Data availability. The data set used is available on the Zenodo database at https://doi.org/10.5281/zenodo.14146969.

Author contributions. JP was responsible for the initial assembly and processing of the Purple Air and Ministry measurements and writing an initial report of the work. ET performed much of the reprocessing of the data as well as contributions to the manuscript. RJS supervised both students, helped in the implementation and coding of the OEM method, and wrote an initial draft of the manuscript.

Competing interests. The authors declare that they have no conflict of interests.

Acknowledgements. This project has been funded in part by the National Science and Engineering Research Council of Canada through a Discovery Grant (Robert J. Sica) and by a Canadian Space Agency FAST grant (Kimberly Strong, PI).

https://doi.org/10.5194/egusphere-2024-3618
Preprint. Discussion started: 9 December 2024

References

Akpootu, D. and Gana, N. N.: The Effect of Relative Humidity on the Hygroscopic Growth Factor and Bulk Hygroscopicity of water Soluble Aerosols, The International Journal Of Engineering And Science, 2, 48–57, 2013.

Ardon-Dryer, K., Dryer, Y., Williams, J. N., and Moghimi, N.: Measurements of PM2.5 with PurpleAir under atmospheric conditions, Atmospheric Measurement Techniques, 13, 5441–5458, 2020.

Barkjohn, K., Gantt, B., and Clements, A.: Development and application of a United States-wide correction for PM2.5 data collected with the PurpleAir sensor, Atmospheric Measurement Technology, 14, 4617–4637, 2021.

Bell, M., Dominici, F., Ebisu, K., Zeger, S., and Samet, J.: Spatial and Temporal Variation in PM2.5 Chemical Composition in the United States for Health Effects Studies, Environmental Health Prospectives, 115, 989–995, 2007.

Bi, J., Wildani, A., Chang, H. H., and Liu, Y.: Incorporating Low-Cost Sensor Measurements into High-Resolution PM2.5 Modeling at a Large Spatial Scale, Environmental Science and Technology, 54, 2152–2162, 2020.

Cerully, K. M., Bougiatioti, A., Jr., J. R. H., Guo, H., Xu, L., Ng, N. L., Weber, R., and Nenes, A.: On the link between hygroscopicity, volatility, and oxidation state of ambient and water-soluble aerosols in the southeastern United States, Atmospheric Chemistry and Physics, 15, 8679–8694, 2015.

Eriksson, P., Jiménez, C., and Buehler, S. A.: Qpack, a general tool for instrument simulation and retrieval work, Journal of Quantitative Spectroscopy and Radiative Transfer, 91, 47–64, 2004.

Magi, B. I., Cupini, C., Francis, J., Green, M., and Hauser, C.: Evaluation of PM2.5 measured in an urban setting using a low-cost optical particle counter and a Federal Equivalent Method Beta Attenuation Monitor, Aerosol Science and Technology, 54, 147–159, 2019.

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., and Subramanian, R.: Fine particle mass monitoring with low-cost sensors: Corrections and longterm performance evaluation, Aerosol Science and Technology, 24, 160–174, 2020.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmospheric Chemistry and Physics, 7, 1961–1971, 2007.

Rodgers, C.: Inverse Methods for Atmospheric Soundings, World Scientific Publishing Co. Pte. Ltd., Singapore, 2000.

Tryner, J., L'Orange, C., Mehaffy, J., Miller-Lionberg, D., Hofstetter, J. C., Wilson, A., and Volckens, J.: Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers Author links open overlay panel, Atmospheric Environment, 220, 1352–2310, 2020.