



Retrieval of bulk hygroscopicity from PurpleAir PM_{2.5} sensor measurements

Jillian Psotka^{1,★}, Emily Tracey^{1,★}, and Robert J. Sica^{1,★}

¹Department of Physics and Astronomy, The University of Western Ontario, London, Canada

★These authors contributed equally to this work.

Correspondence: Robert J. Sica (sica@uwo.ca)

Received: 19 November 2024 – Discussion started: 9 December 2024

Revised: 11 March 2025 – Accepted: 21 April 2025 – Published: 15 July 2025

Abstract. PurpleAir sensors offer a unique opportunity for a large-scale and densely populated array of sensors to study surface air quality. While PurpleAir sensors are inexpensive and abundant, they must be corrected to better agree with validated coincident measurements from more sophisticated instrumentation. Traditionally, this correction is performed using statistical methods. We propose a method to both correct PurpleAir PM_{2.5} measurements and allow for an estimate of the hygroscopic growth of aerosols, using a novel correction approach based on the optimal estimation method (OEM). The hygroscopic growth of aerosols can be retrieved using the sensitivity of the correction to water activity, which influences the measured size distribution of the aerosols. By employing the physically based correction using calibrated measurements from the nearby Ontario Ministry of the Environment, Conservation and Parks air quality site, the average daily mean absolute error (MAE) of the PurpleAir PM_{2.5} measurements is decreased from 5.58 to 1.68 $\mu\text{g m}^{-3}$, and the average daily bias decreases from 4.75 to $-0.23 \mu\text{g m}^{-3}$. This improvement in the correction is comparable to that seen using conventional statistical methodologies. Our OEM retrieval also 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 for atmospheric particulate matter (0.1 to 0.9) determined in previous studies using calibrated air quality measurement instruments, which suggests that our method allows a new aerosol product to be determined from a large sensor network.

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 USD 300. These sensors connect to WiFi and provide users with real-time air quality and ambient condition readings at a 2 min resolution. Each PurpleAir sensor can be integrated with the PurpleAir (2016) Network, where data from over 30 000 sensors across the globe are publicly available (<https://map.purpleair.com>, last access; 27 February 2025).

PurpleAir sensors employ light-scattering photometry to estimate the concentration of particulates based on an assumed particle composition. Every 2 min, an air sample is drawn through the instrument and the optical laser beam interacts with the particulates in the sample. The scattered light is 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. The Plantower PMS5003 sensor used in the PurpleAir sensors is described in more detail in Ardon-Dryer et al. (2020) and Ouimette et al. (2022).

There are two main differences in sensor function that account for the low cost of the PurpleAir sensors relative to research-grade air quality sensors. The first is that the PurpleAir instrument does not have the technology to measure the mass of the particles directly through beta attenuation; therefore it must make assumptions and estimations in order to calculate the photometric particle count (measured in particles per decilitre) and the mass density. These corrections

are done on board 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.

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 an increase in diameter in the presence of water (hygroscopic growth). Hygroscopic growth introduces uncertainties in low-cost sensor measurements because they detect higher scattering, leading to estimations of higher concentrations. 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.

PurpleAir sensors have been successfully applied not only for home use but also for scientific purposes. Bi et al. (2020) incorporated PurpleAir measurements into large-scale PM_{2.5} modelling. They compared 54 PurpleAir sensors to nearby US Environmental Protection Agency Air Quality System (AQS) stations across California. They corrected the PurpleAir sensors using a generalized additive model that included linear corrections for relative humidity, temperature, and sensor operating time. The corrected PurpleAir data were 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. The study by Bi et al. (2020) is promising in terms of wider applications of PurpleAir data use and could have been improved further with a more accurate PurpleAir correction.

Barkjohn et al. (2021) developed statistical corrections using 50 PurpleAir sensors at 39 unique sites in the US. 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 correction applicable to any PurpleAir site, regardless of proximity to reference instruments. Their study found that PurpleAir sensors' overestimations of PM_{2.5} readings could be adequately corrected by a multiple linear model of the form

$$\text{PM}_{2.5,\text{corr}} = a\text{PM}_{2.5,\text{meas}} + bH + c, \quad (1)$$

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.

They were successful in creating a single nationwide correction that could be applied to PurpleAir sensors and reduce errors. This work was expanded on in Barkjohn et al. (2022), where a correction was developed in cases of extreme smoke concentrations ($> 300 \mu\text{g m}^{-3}$).

Other studies (Ardon-Dryer et al., 2020; Magi et al., 2019; Malings et al., 2019; Tryner et al., 2020; Nilson et al., 2022) have also applied statistical methods to correct the PurpleAir sensors to standard research-grade instruments. In this study, we used an inverse modelling technique called the optimal estimation method (OEM) to correct the PurpleAir measurements using a Thermo Scientific Synchronized Hybrid Ambient Real-time (SHARP) model 5030 particulate monitor and a physical model of a hygroscopic growth factor given by Malings et al. (2019). The OEM is explained in detail by Rodgers (2000), and we give a brief description in Sect. 2.3.

The hygroscopicity of bulk aerosol was also retrieved during the correction process, providing a possible advantage of this technique over previous correction methods. Hygroscopicity is a fundamental parameter describing the ability of aerosol particles to absorb water (Kreidenweis and Asa-Awuku, 2013; Tang et al., 2016). It is important to measure hygroscopicity because it impacts the ability of aerosols to act as cloud condensation nuclei (CCN); thus it affects the formation and properties of clouds and their indirect radiative forcing (Farmer et al., 2015; Petters and Kreidenweis, 2007; Reutter et al., 2009; Su et al., 2010; McFiggans et al., 2006).

2 Methodology

2.1 The physical model

The forward model we used with our OEM is based on the physics-based correction model used by Malings et al. (2019). This study used two different correction methods, one physics-based and one statistical, to ensure the PurpleAir data would better match regulatory-grade data for nine PurpleAir sensors. The physical model was based on the hygroscopic growth of different aerosols and the composition of the air pollution in Pittsburgh, Pennsylvania (US). The hygroscopic growth factor of PM_{2.5}, which quantifies the hygroscopic growth of aerosols, is given by

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

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 :

$$\kappa_{\text{bulk}} = \sum_{i=1}^n x_i \kappa_i. \quad (3)$$

Malings et al. (2019) 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 US (Bell et al., 2007). The water activity was calculated as a function of temperature and relative humidity as

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

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:

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

where a and b are constant coefficients; $\text{PM}_{2.5,\text{corr}}$ and $\text{PM}_{2.5,\text{meas}}$ are the corrected and measured PM_{2.5}, respectively; and $f(T, H)$ is defined in Eq. (2). This physics-based correction was compared to 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 $\mu\text{g m}^{-3}$), but yearly-averaged readings were more accurate (errors less than 1 $\mu\text{g m}^{-3}$). Malings et al. (2019) established that their physically based model with a constant term is reasonably complete and serves as the basis for the forward model we use with our OEM.

2.2 Treatment of the measurements

This study uses measurements from January to December 2021 taken by two air quality sensors described below. Daily averages were taken to produce one data point per day. To perform the correction, 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 and its effect on bulk hygroscopicity.

2.2.1 PurpleAir sensor

Measurements from all public PurpleAir sensors are available for download from the PurpleAir (2016) Network map. (<https://map.purpleair.com>, last access: 27 February 2025). Each sensor has two particle counters, Channel A and Channel B, which report independently for the purpose of precision. The data are provided as unfiltered, 2 min 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).

If the raw, 2 min PM_{2.5} readings had discrepancies between Channels A and B of more than 4 $\mu\text{g m}^{-3}$ or more than 25 % of their average, they were removed before performing the daily averages. Any readings that only had a value from one channel were also removed. This procedure eliminated about 3 % of the measurements. In addition to this quality control procedure, there were some days, and in one case over 3 consecutive weeks (August–September 2021), where all measurements were discarded due to clearly erroneous readings. In some cases the cause of these errors remains unknown, but it is likely due to insects or associated debris in the sensor that had to be removed.

Relative humidity and temperature averages were compared to measurements from the Environment and Climate Change Canada weather station at London International Airport, about 12 km from the PurpleAir sensor. Due to internal heating and insolation effects, 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., 2019). The relative humidity readings from PurpleAir were consistently about 21 % lower than the airport station values, so a simple correction of adding 21 % to each PurpleAir relative humidity measurement was made. The relative humidity measurements before and after the correction was applied are shown in Appendix A. The PurpleAir temperature readings were, on average, about 2 °C higher than the airport station values, 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 (see Sect. 3.1.3).

2.2.2 Ontario MECP validation measurements

Measurements from all Ontario Ministry of the Environment, Conservation and Parks (MECP) air quality sites are available on the MECP (2021) website (<http://www.airqualityontario.com/history/summary.php>, last access: 27 February 2025). 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 sensor used at this location is SHARP 5030, and its PM_{2.5} readings are given as integer values. To test that this rounding did not impact our correction, we applied various rounding schemes (floor, half round up, and ceiling) to the PurpleAir measurements and redid the corrections. None of these rounding choices had a significant impact on the correction results.

2.3 The optimal estimation method

OEM is an inverse method that allows for 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

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

where $\mathbf{y}$ is the measurement vector; $\mathbf{x}$ is the state vector, the vector which contains the retrieved quantities; $\mathbf{b}$ is an additional parameter required by the forward model; and $\boldsymbol{\epsilon}$ is the measurement noise vector.

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 estimates of the state vector and its error covariance matrix. In a successful OEM retrieval, one should be able to vary the a priori estimates slightly without having an effect on the retrieved state. This indicates that the a priori estimates guide the solution rather than dictate it.

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):

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

Our forward model does not include the constant a appearing in the form of the model described by Malings et al. (2019). The constant factor is not required in OEM as we directly retrieve the hygroscopicity of the bulk aerosol. Descriptions of all variables, along with their model parameter type, are given in Table 1.

We took the measurement vector $\mathbf{y}$ to be the reference $\text{PM}_{2.5}$ readings from the MECP site, since the goal was to have the corrected data match these reference readings. The a priori state vector consists of a constant linear term and bulk hygroscopicity. The a priori bulk hygroscopicity was informed by Cerully et al. (2015). The a priori constant linear term was estimated from simple straight-line fits to the measurements. The components of the a priori state vector and their errors are presented in Table 2.

Table 1. Parameters used in the 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	$\mathbf{x}$
κ_{bulk}	bulk hygroscopicity	$\mathbf{x}$
D_p	average particle diameter	$\mathbf{b}$
$\text{PM}_{2.5,\text{meas}}$	$\text{PM}_{2.5}$ as measured by PurpleAir	$\mathbf{b}$
H	relative humidity as measured by PurpleAir	$\mathbf{b}$
T	temperature as measured by PurpleAir	$\mathbf{b}$
σ_w	surface tension of water	$\mathbf{b}$
M_w	molecular weight of water	$\mathbf{b}$
ρ_w	density of water	$\mathbf{b}$
R	ideal gas constant	$\mathbf{b}$

Table 2. The components of the a priori state vector and their errors used in the OEM code. The errors were used to create the a priori covariance matrix.

Parameter	Value
$c \text{ (}\mu\text{g m}^{-3}\text{)}$	1.9 ± 0.66
κ_{bulk}	0.25 ± 0.025

2.4 Statistical metrics used to assess the correction

The accuracy of the correction was assessed using mean absolute error (MAE) and bias. MAE is used to assess how well, on average, a data set agrees with the reference data. It is calculated as

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |r_i - \hat{r}_i| \tag{9}$$

for n measurements of PurpleAir readings (r) and reference readings ($\hat{r}$). Similarly, the bias of each data set is used to assess the systematic differences between data sets and is calculated as

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (r_i - \hat{r}_i). \tag{10}$$

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 corrected data correlated with the reference data.

3 Evaluation of the OEM model

3.1 Sensitivity analysis

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

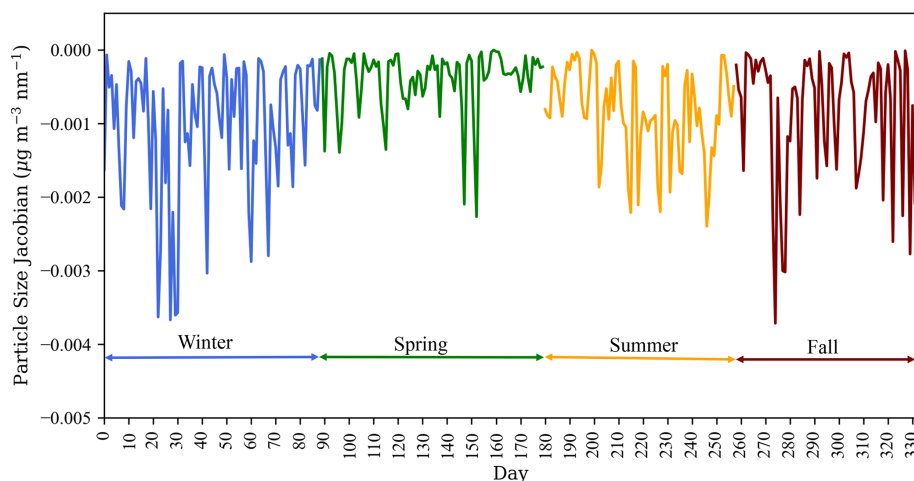


Figure 1. A time series of the sensitivity of the forward model to the particulate diameter parameter. The y axis depicts the change in the corrected $\text{PM}_{2.5}$ value for every 1 nm change in the assumed particle diameter.

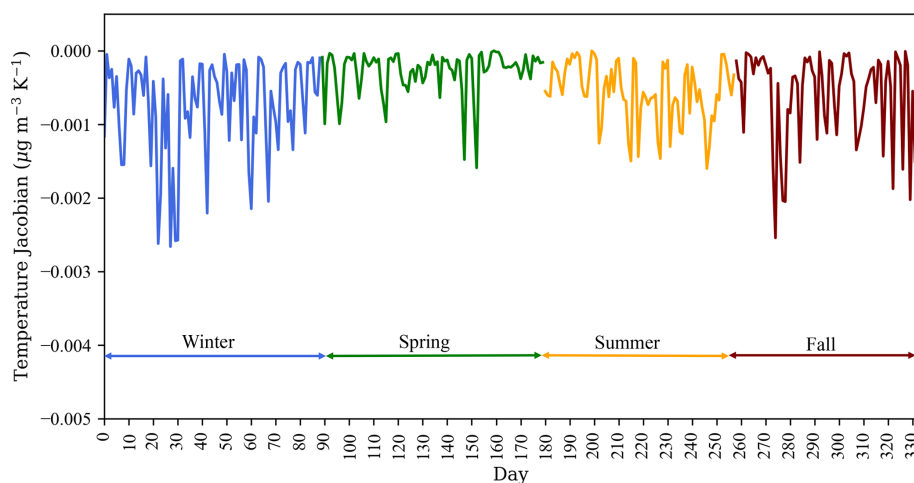


Figure 2. A time series of the sensitivity of the forward model to the temperature parameter. The y axis depicts the change in the corrected $\text{PM}_{2.5}$ value for every 1 K change in the temperature.

given by

$$\mathbf{E} = \mathbf{G} \cdot \mathbf{J}_b \cdot \mathbf{S} \cdot \mathbf{J}_b^T \cdot \mathbf{G}^T, \quad (11)$$

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

$$\mathbf{G} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}}; \quad (12)$$

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

$$\mathbf{J}_b = \frac{\partial \mathbf{F}}{\partial \mathbf{b}}; \quad (13)$$

and $\mathbf{S}$ is the uncertainty covariance matrix (Rodgers, 2000). We use these equations to investigate the impact of particle diameter, temperature, and relative humidity on the retrieval.

3.1.1 Sensitivity to average particle diameter

We tested the sensitivity of the model to the average particle diameter by investigating the Jacobian given in Eq. (13), where the parameter, $\mathbf{b}$, is the particle diameter. This Jacobian shows that the change in the corrected $\text{PM}_{2.5}$ for a change in particle size of 1 nm is 3 orders of magnitude smaller than the $\text{PM}_{2.5}$ measurements and thus insignificant (Fig. 1). 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, even large uncertainties in assumed average particle diameter do not impact the retrieval.

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 aver-

age particle diameter, in our case a value of $D_p = 200$ nm, can be recommended for future applications of this model.

3.1.2 Sensitivity to temperature

We investigated the behaviour of the physical model in response to temperature to check if the PurpleAir sensor temperature measurements are of sufficient quality for the intended use. The Jacobian is shown in Fig. 2. We found that the forward model is not sensitive to temperature, and changes in temperature throughout a range of typical London annual temperatures (-20 to 35 °C) would not impact the retrieval significantly. Therefore, the PurpleAir temperature measurements may be used without correction. It should be noted that the physical model is still sensitive to temperature through the relative humidity dependence, which is discussed in the next section.

3.1.3 Sensitivity to relative humidity

We also investigated the sensitivity of the forward model to relative humidity. The Jacobian for relative humidity is shown in Fig. 3a. The Jacobian is large enough to indicate that the model is sensitive to relative humidity. The Jacobian varies from day to day due to its correlation with $PM_{2.5}$. Figure 3b shows this relation.

Since the model is sensitive to relative humidity, we recommend correcting the PurpleAir relative humidity measurements before use. For this data set, a constant correction of 21 % was sufficient to bring the PurpleAir measurements into better agreement with measurements from the nearby airport meteorological station, but this correction may not be applicable to other sites or time periods. The relative humidity measurements before and after the correction compared to the airport measurements are given in Appendix A.

3.2 OEM correction results

After using OEM to retrieve the hygroscopic growth factor and the constant term, we applied the physical model (Eq. 8) to correct the PurpleAir measurements from each season. The physically corrected and raw $PM_{2.5}$ measurements are shown in Fig. 4. For comparison, the PurpleAir measurements were also corrected using multiple linear regression (MLR), as was done, for example, by Barkjohn et al. (2021). The MLR result is also shown in Fig. 4. Details of the MLR correction are given in Appendix B. The physical correction, shown in red, succeeded in bringing the raw measurements closer to the reference measurements and performs similarly to the statistical correction shown in blue. The cost function of the OEM retrieval is around 5.5, signifying a good fit without overfitting, which occurs when the cost function is less than 1.

The MAE and R^2 values from both models are given in Table 3. The physical correction 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 were the best of the physically corrected results for all of our metrics (MAE, bias, and R^2). Although both models performed similarly, the statistically corrected data had smaller MAEs and slightly higher values of R^2 . It can be concluded that the statistical correction had better overall performance, while the physical correction allows a new physical data product to be retrieved with slightly poorer $PM_{2.5}$ correction.

To investigate the impact of relative humidity on the correction, the biases of the physically corrected data from each season are plotted as a function of relative humidity (Fig. 5). This was done to illustrate how the correction behaves at higher humidities. We can see that the physical correction tends to overcorrect the data when the relative humidity is above about 65 %, as indicated by a negative bias. This effect is not observed consistently with the statistically corrected data. Thus, corrections using OEM at higher values of relative humidity may be insufficient. Mathieu-Campbell et al. (2024) suggest that a clustering approach is more effective at correcting PurpleAir measurements in high-humidity conditions, which allows the non-linearity associated with hygroscopic growth to be captured. The average daily bias is less than $1 \mu\text{g m}^{-3}$ for both correction models.

3.3 Bulk hygroscopicity results

The bulk hygroscopicity of particulates is 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. 6a and b, respectively. The error in bulk hygroscopicity is the sum of the observation error from OEM and the error due to relative humidity calculated using Eq. (11), assuming that the relative humidity is known within 2.5 % uncertainty. This 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.

Investigating retrieved bulk hygroscopicity

The retrieved values of hygroscopicity of bulk aerosol are between 0.33 and 0.4, as shown in Fig. 6, 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 the largest in the spring. This is consistent with the result from Akpootu and Gana (2013), which describes an inverse relationship between hygroscopicity and relative humidity. Our results show the highest relative humidity in the fall and the lowest in the spring, with the hygroscopic growth factor varying proportionally as expected. The error due to relative humidity is on average 4 % of the total error, which means that it is small in comparison to the observation error. Therefore, even though it was found

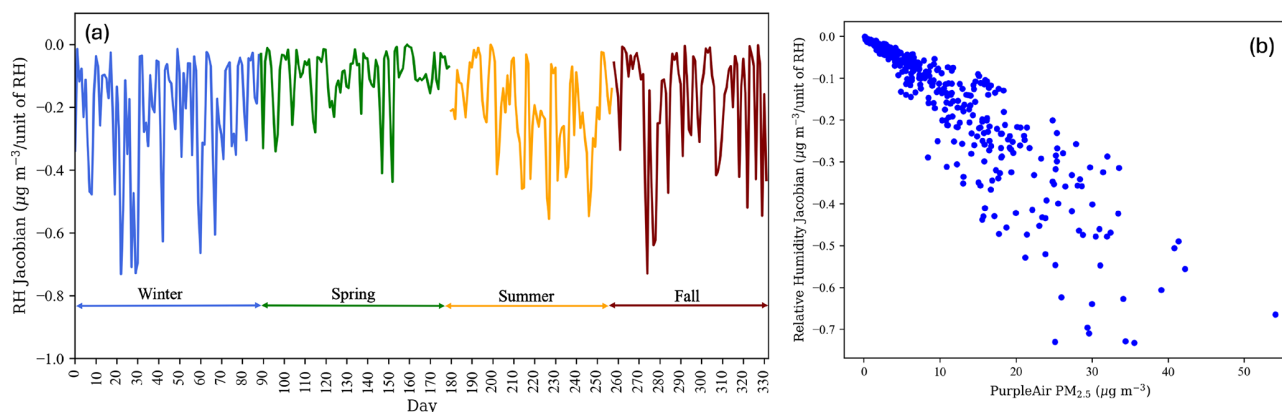


Figure 3. (a) A time series of the sensitivity of the forward model to the relative humidity (RH). The y axis depicts the change in the corrected PM_{2.5} value for every 1 % change in RH. (b) The correlation between the RH Jacobian and the PM_{2.5} measured by PurpleAir, with each point representing 1 d.

Table 3. MAE of all raw, physically corrected, and statistically corrected measurements along with the R^2 value of the physical and statistical corrections.

Season	Raw MAE ($\mu\text{g m}^{-3}$)	Raw R^2	Physical MAE ($\mu\text{g m}^{-3}$)	Physical R^2	Statistical MAE ($\mu\text{g m}^{-3}$)	Statistical R^2
Spring	3.48	0.61	1.38	0.78	1.05	0.87
Summer	6.95	0.42	1.98	0.72	1.74	0.78
Fall	5.00	0.40	1.59	0.61	1.50	0.66
Winter	6.90	0.46	1.78	0.75	1.34	0.75

in Sect. 3.1.3 that the forward model (the corrected PM_{2.5}) is sensitive to relative humidity, our retrieved value for bulk hygroscopicity is not very sensitive to relative humidity. The results show that our method may have the potential to estimate bulk hygroscopicity.

4 Limitations of our approach

One limitation of our method is that the reference instrument is not co-located with our PurpleAir sensor. From observations of the spatial spread of PM_{2.5} from the PurpleAir website, we notice that regions in close proximity to the London site used for this study 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 correction 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 what effect this limitation could have had on our study.

Another limitation is with respect to the correction of daily-averaged data sets. The main purpose of this correction 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 h, and averaging greatly smooths these variations. Therefore it is undetermined whether the daily-averaged data sets fully encapsulate the model of hygroscopic growth in the presence of humidity. Moreover, this study does not include any extreme events, such as wildfires, which would increase PM_{2.5} significantly, so it is not known how our method would be affected by unusually high concentrations.

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

5 Conclusions

We applied a physical model based on the hygroscopic growth of particulates to correct PM_{2.5} measurements from a PurpleAir sensor and showed that it is possible to estimate hygroscopic growth as part of the sensor calibration using the optimal estimation method. We corrected daily-averaged data for 1 year, split into four seasons. The physical correction reduced average daily MAE and bias from 5.58 to 1.68 $\mu\text{g m}^{-3}$ and from 4.75 to $-0.23 \mu\text{g m}^{-3}$, respectively. The physical model tended to overcorrect data points, with daily-averaged relative humidity above approximately 65 %.

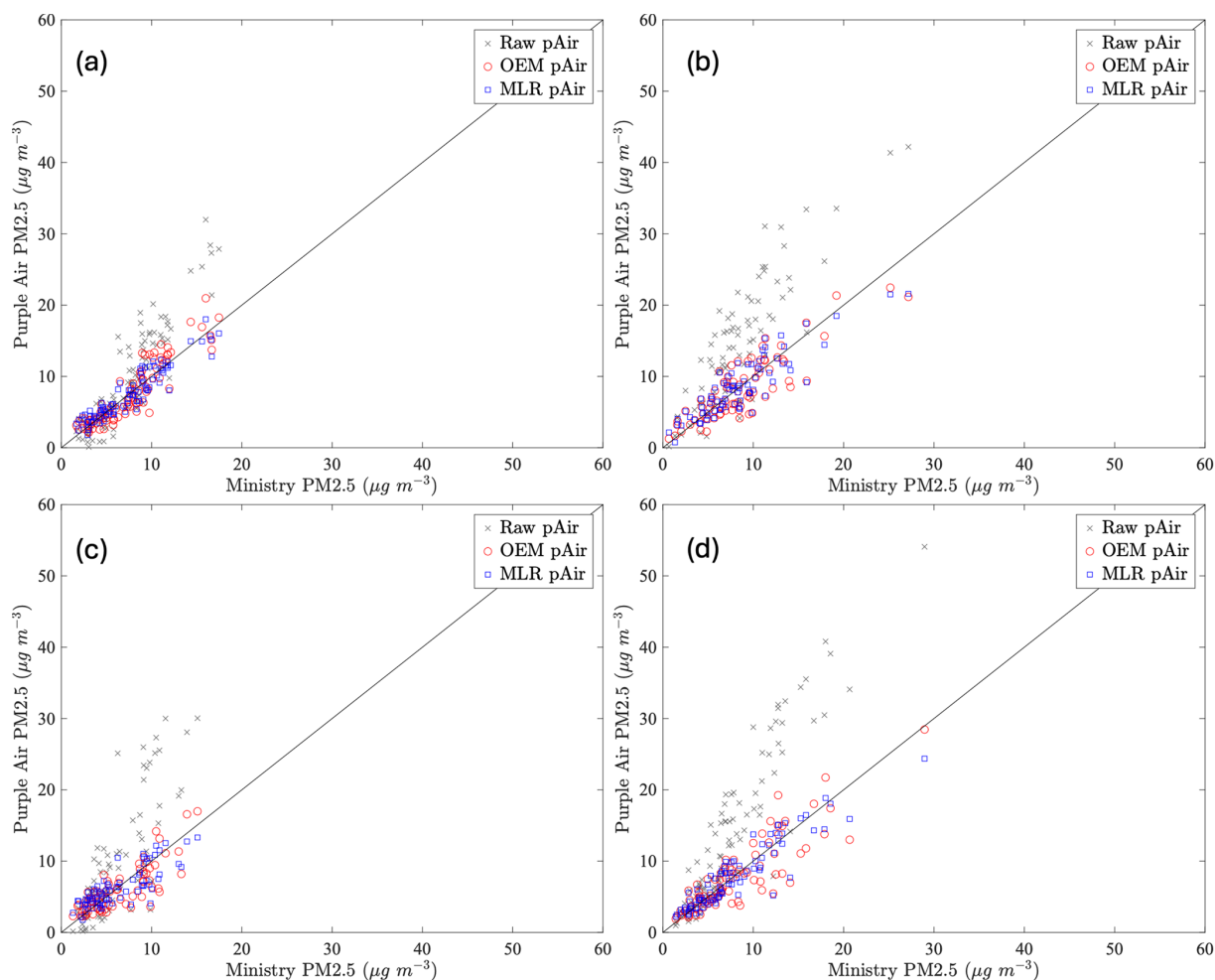


Figure 4. Daily-averaged $\text{PM}_{2.5}$ as measured by MECP vs. PurpleAir for (a) spring, (b) summer, (c) fall, and (d) winter. Data correction using the optimal estimation method (OEM) is shown in red, raw data are shown in grey, statistically corrected data using a multiple linear regression (MLR) are shown in blue, and the 1 : 1 line is shown in black. The horizontal axes are $\text{PM}_{2.5}$ readings by the MECP reference sensor, and the vertical axes are averaged $\text{PM}_{2.5}$ readings from the PurpleAir sensor. The location of each point in the plots signifies 1 d of $\text{PM}_{2.5}$ measurements by the MECP sensor and the PurpleAir sensor measurement for the same day averaged from 2 min readings.

This relative humidity bias was not seen in the statistical correction, which reduced the average bias to $0 \mu\text{g m}^{-3}$ (due to the statistical nature of linear regression) and average MAE to $1.46 \mu\text{g m}^{-3}$.

The physical correction did not perform quite as well as the statistical correction, 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 ranging from 0.33 to 0.4. Furthermore, our method is extremely fast computationally, making it ideal to apply to “real-time” situations, such as air quality maps like the hourly $\text{PM}_{2.5}$ University of Northern British Columbia (UNBC)/Environment and

Climate Change Canada (ECCC) map by Nilson and Jackson (2025) (<https://aqmap.ca/aqmap>, last access: 27 February 2025).

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 $\text{PM}_{2.5}$ at high levels of relative humidity. Finally, it should be noted that the OEM could, in principle, retrieve the individual values and bulk values of the hygroscopicity as in the current retrieval.

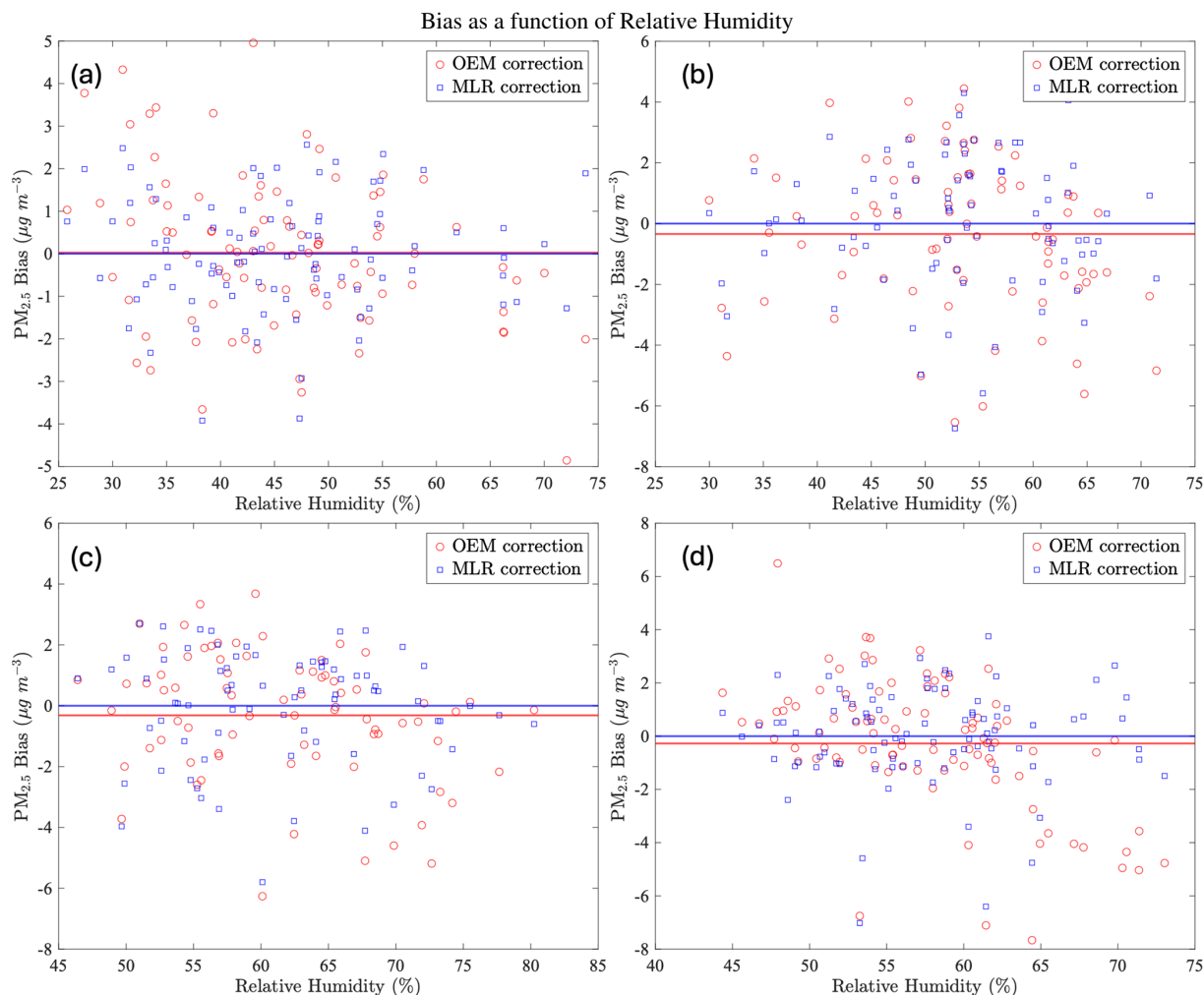


Figure 5. The bias in the PM_{2.5} measurements, corrected using the optimal estimation method (OEM) and shown in red, and using a multiple linear regression (MLR) shown in blue, as a function of relative humidity for (a) spring, (b) summer, (c) fall, and (d) winter. The horizontal red line shows the average bias from the OEM correction, and the horizontal blue line is the average bias from the MLR correction.

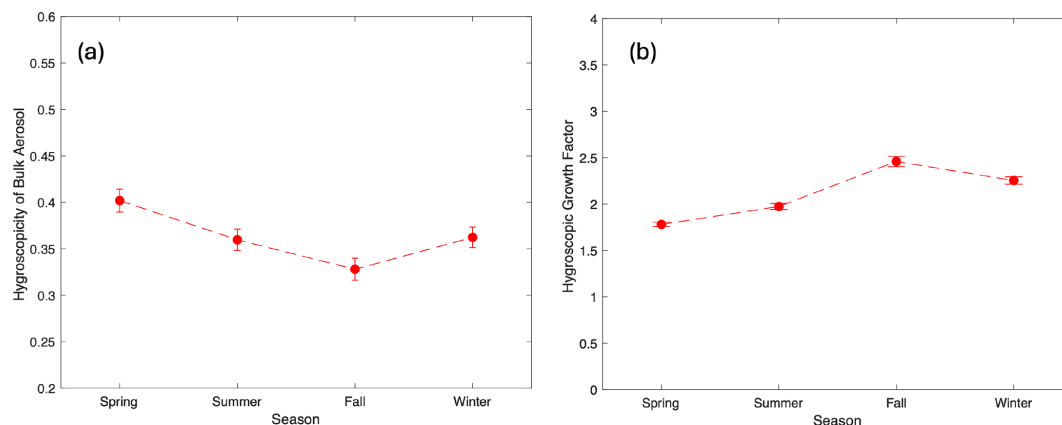


Figure 6. (a) Bulk hygroscopicity and (b) the hygroscopic growth factor retrieved through the OEM for each season. Uncertainties are based on the assumption that relative humidity is known within a 2.5 % uncertainty.

Appendix A: Humidity correction

PurpleAir sensors are known to have high-temperature readings and low-relative-humidity readings due to internal heating (PurpleAir, 2021). A simple correction factor of adding 21 % to all PurpleAir measurements sufficed to correct the PurpleAir relative humidity for our purposes. The raw and corrected relative humidity readings across the year of data used in this study are shown in Fig. A1 along with readings from the official ECCC weather station.

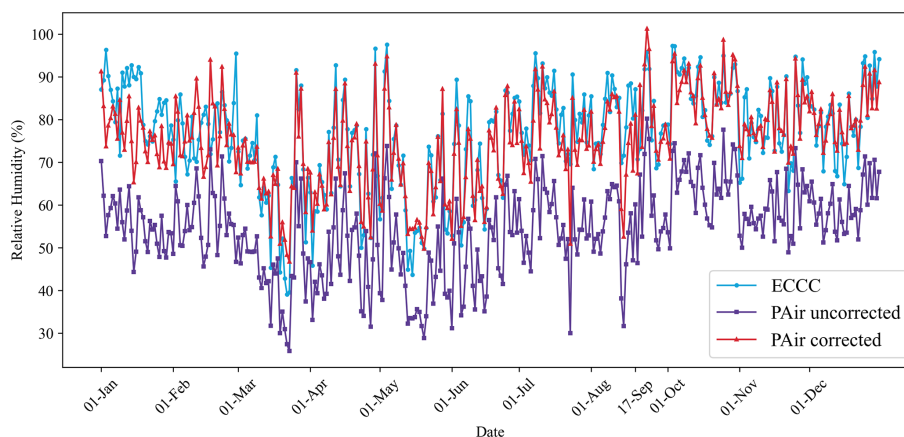


Figure A1. A time series of the relative humidity measured by the PurpleAir sensor before and after correction compared to the ECCC reference measurement taken at London International Airport.

Appendix B: Statistical correction

The forward-model equation for the statistical calibration is as follows:

$$\text{PM}_{2.5,\text{corr}} = a\text{PM}_{2.5,\text{meas}} + bH + c, \quad (\text{B1})$$

where a , b , and c are constants. Their values for each data set are shown in Table B1.

Table B1. Statistical model coefficients and errors.

Season	a	b ($\mu\text{g m}^{-3}$)	c ($\mu\text{g m}^{-3}$)
Spring	0.46 ± 0.02	-0.07 ± 1.34	6.25 ± 0.95
Summer	0.49 ± 0.03	-0.10 ± 2.77	6.65 ± 1.96
Fall	0.33 ± 0.03	-0.06 ± 2.96	6.85 ± 2.40
Winter	0.42 ± 0.02	-0.014 ± 3.16	2.68 ± 2.44

Code availability. The OEM retrieval used for this study is part of the ARTS package and can be downloaded at <https://github.com/atmtools> (Eriksson et al., 2011).

Data availability. The data set used is available on the Zenodo database at <https://doi.org/10.5281/zenodo.14146969> (Sica, 2024).

Author contributions. JP was responsible for the initial assembly and processing of the PurpleAir and Ministry of the Environment, Conservation and Parks measurements and wrote an initial report on the work. ET performed much of the reprocessing of the data and contributed to the paper. RJS supervised both students, assisted with the implementation and coding of the OEM, and wrote the initial draft of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We thank the reviewers for their many insightful comments, which improved our explanations and presentation of the results.

Financial support. This research has been supported by the NSERC Discovery Grant (grant no. RGPIN-2018-04999) and the Canadian Space Agency FAST grant (grant no. 2019-TORA07).

Review statement. This paper was edited by Albert Presto and reviewed by two anonymous referees.

References

- Akpootu, D. and Gana, N. N.: The Effect of Relative Humidity on the Hygroscopic Growth Factor and Bulk Hygroscopicity of water Soluble Aerosols, *Int. J. Eng. Sci.*, 2, 48–57, 2013.
- Ardon-Dryer, K., Dryer, Y., Williams, J. N., and Moghimi, N.: Measurements of PM_{2.5} with PurpleAir under atmospheric conditions, *Atmos. Meas. Tech.*, 13, 5441–5458, <https://doi.org/10.5194/amt-13-5441-2020>, 2020.
- Barkjohn, K. K., Gantt, B., and Clements, A. L.: Development and application of a United States-wide correction for PM_{2.5} data collected with the PurpleAir sensor, *Atmos. Meas. Tech.*, 14, 4617–4637, <https://doi.org/10.5194/amt-14-4617-2021>, 2021.
- Barkjohn, K. K., Holder, A. L., Frederick, S. G., and Clements, A. L.: Correction and Accuracy of PurpleAir PM_{2.5} Measurements for Extreme Wildfire Smoke, *Sensors*, 22, 9669, <https://doi.org/10.3390/s22249669>, 2022.
- Bell, M., Dominici, F., Ebisu, K., Zeger, S., and Samet, J.: Spatial and Temporal Variation in PM_{2.5} Chemical Composition in the United States for Health Effects Studies, *Environmental Health Perspectives*, 115, 989–995, 2007.
- Bi, J., Wildani, A., Chang, H. H., and Liu, Y.: Incorporating Low-Cost Sensor Measurements into High-Resolution PM_{2.5} Modeling at a Large Spatial Scale, *Environ. Sci. Technol.*, 54, 2152–2162, 2020.
- Cerully, K. M., Bougiatioti, A., Hite Jr., J. R., 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, *Atmos. Chem. Phys.*, 15, 8679–8694, <https://doi.org/10.5194/acp-15-8679-2015>, 2015.
- Eriksson, P., Jiménez, C., and Buehler, S. A.: Qpack, a general tool for instrument simulation and retrieval work, *J. Quant. Spectrosc. Ra.*, 91, 47–64, 2004.
- Eriksson, P., Buehler, S. A., Davis, C. P., Emde, C., and Lemke, O.: ARTS, the atmospheric radiative transfer simulator, version 2, *J. Quant. Spectrosc. Ra.*, 112, 1551–1558, <https://doi.org/10.1016/j.jqsrt.2011.03.001>, 2011 (code available at: <https://github.com/atmtools>, last access: 27 February 2025).
- Farmer, D. K., Cappa, C. D., and Kreidenweis, S. M.: Atmospheric Processes and Their Controlling Influence on Cloud Condensation Nuclei Activity, *Chem. Rev.*, 115, 4199–4217, 2015.
- Kreidenweis, S. M. and Asa-Awuku, A.: Aerosol Hygroscopicity: Particle Water Content and Its Role in Atmospheric Processes, *Treatise on Geochemistry: Second Edition*, 5, 331–361, <https://doi.org/10.1016/B978-0-08-095975-7.00418-6>, 2013.
- Magi, B. I., Cupini, C., Francis, J., Green, M., and Hauser, C.: Evaluation of PM_{2.5} measured in an urban setting using a low-cost optical particle counter and a Federal Equivalent Method Beta Attenuation Monitor, *Aerosol Sci. Technol.*, 54, 147–159, 2019.
- Malings, C., Tanzer, R., Hauriuk, 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 Sci. Technol.*, 24, 160–174, 2019.
- Mathieu-Campbell, M. E., Guo, C., Grieshop, A. P., and Richmond-Bryant, J.: Calibration of PurpleAir low-cost particulate matter sensors: model development for air quality under high relative humidity conditions, *Atmos. Meas. Tech.*, 17, 6735–6749, <https://doi.org/10.5194/amt-17-6735-2024>, 2024.
- McFiggans, G., Artaxo, P., Baltensperger, U., Coe, H., Facchini, M. C., Feingold, G., Fuzzi, S., Gysel, M., Laaksonen, A., Lohmann, U., Mentel, T. F., Murphy, D. M., O'Dowd, C. D., Snider, J. R., and Weingartner, E.: The effect of physical and chemical aerosol properties on warm cloud droplet activation, *Atmos. Chem. Phys.*, 6, 2593–2649, <https://doi.org/10.5194/acp-6-2593-2006>, 2006.
- MECP: Fine Particulate Matter, <http://www.airqualityontario.com/science/pollutants/particulates.php> (last access: 27 February 2025), 2021.
- Nilson, B. and Jackson, P.: AQmap, version 3.5.0, <https://aqmap.ca/aqmap> (last access: 27 February 2025), 2025.
- Nilson, B., Jackson, P. L., Schiller, C. L., and Parsons, M. T.: Development and evaluation of correction models for a low-cost fine particulate matter monitor, *Atmos. Meas. Tech.*, 15, 3315–3328, <https://doi.org/10.5194/amt-15-3315-2022>, 2022.
- Ouimette, J. R., Malm, W. C., Schichtel, B. A., Sheridan, P. J., Andrews, E., Ogren, J. A., and Arnott, W. P.: Evaluating the PurpleAir monitor as an aerosol light scattering instrument, *Atmos. Meas. Tech.*, 15, 655–676, <https://doi.org/10.5194/amt-15-655-2022>, 2022.
- Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmos. Chem. Phys.*, 7, 1961–1971, <https://doi.org/10.5194/acp-7-1961-2007>, 2007.
- PurpleAir: PurpleAir Map, air quality Map, <https://map.purpleair.com> (last access: 27 February 2025), 2016.
- PurpleAir: PurpleAir FAQ, <https://www2.purpleair.com/community/faq> (last access: 27 February 2025), 2021.
- Reutter, P., Su, H., Trentmann, J., Simmel, M., Rose, D., Gunthe, S. S., Wernli, H., Andreae, M. O., and Pöschl, U.: Aerosol- and updraft-limited regimes of cloud droplet formation: influence of particle number, size and hygroscopicity on the activation of cloud condensation nuclei (CCN), *Atmos. Chem. Phys.*, 9, 7067–7080, <https://doi.org/10.5194/acp-9-7067-2009>, 2009.
- Rodgers, C.: Inverse Methods for Atmospheric Soundings, World Scientific Publishing Co. Pte. Ltd., Singapore, <https://doi.org/10.1142/3171>, 2000.
- Sica, R.: Dataset for PurpleAir OEM retrieval, Zenodo [data set], <https://doi.org/10.5281/zenodo.14146970>, 2024.
- Su, H., Rose, D., Cheng, Y. F., Gunthe, S. S., Massling, A., Stock, M., Wiedensohler, A., Andreae, M. O., and Pöschl, U.: Hygroscopicity distribution concept for measurement data analysis and modeling of aerosol particle mixing state with regard to hygro-

- scopic growth and CCN activation, *Atmos. Chem. Phys.*, 10, 7489–7503, <https://doi.org/10.5194/acp-10-7489-2010>, 2010.
- Tang, M., Cziczo, D. J., and Grassian, V. H.: Interactions of Water with Mineral Dust Aerosol: Water Adsorption, Hygroscopicity, Cloud Condensation, and Ice Nucleation, *Chem. Rev.*, 116, 4205–4259, <https://doi.org/10.1021/acs.chemrev.5b00529>, 2016.
- 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, *Atmos. Environ.*, 220, 1352–2310, 2020.