Journal of Aerosol Science 159 (2022) 105872

Contents lists available at ScienceDirect

Journal of Aerosol Science

journal homepage: www.elsevier.com/locate/jaerosci

Tutorial: Guidelines for implementing low-cost sensor networks for aerosol monitoring

Naomi Zimmerman

Department of Mechanical Engineering, University of British Columbia, Vancouver, Canada, V6T 1Z4

ARTICLE INFO

Editor: Dr. Chris Hogan

Keywords:
Low-cost sensors
Aerosol
$PM_{2.5}$
Performance assessment
Calibration
Citizen science

ABSTRACT

Over the past decade, there has been exponential growth in low-cost air pollution sensing technology. While low-cost sensors can provide a path towards more accessible air quality measurement, there are several guidelines that should be considered prior to deploying a low-cost sensor network. In this tutorial guide, we focus on low-cost aerosol sensors (in this case, $PM_{2.5}$). The tutorial reviews key guidelines for implementing a low-cost PM sensor network. This article is also associated with a companion web-tutorial (https://nzimmerman-ubc.github.io/lcs-PM-demo/) on downloading and assessing sample low-cost PM sensor data from the PurpleAir network using the new U.S. EPA Fine Particle Sensor Base Testing Guidelines. While this tutorial does not cover every single consideration a researcher or citizen scientist might undertake, it covers the key areas of evaluation and calibration, siting, data reporting, and post-processing. The aim of the tutorial is to improve outcomes for researchers using low-cost PM sensor networks and develop a broader community of practice.

1. Introduction

Air pollutants can be broadly classified into aerosols and gases. Aerosols, which are any solid or liquid suspended in a gas, can be further classified based on their size or composition. In this tutorial, we will focus on a size-based aerosol classification: $PM_{2.5}$, or particles with an aerodynamic diameter <2.5 μm. $PM_{2.5}$ is typically reported and regulated as a mass-based concentration in units of micrograms per meter cubed of air ($\mu g\,m^{-3}$). Sources of $PM_{2.5}$ include combustion processes (traffic, wood burning, coal), atmospheric reactions, and mechanical processes, such as road dust, sea spray, and erosion.

Air pollution studies at both the global and city scale have repeatedly shown the substantial health effects attributable to exposure to $PM_{2.5}$. These health effects include but are not limited to premature mortality, cardiovascular disease, lung cancer and asthma (Brook et al., 2010; Pope et al., 2009; Raaschou-Nielsen et al., 2013). As such, monitoring and reduction of $PM_{2.5}$ is an important public health priority.

Traditionally, $PM_{2.5}$ monitoring has relied on expensive government-managed regulatory monitoring sites often located far from individual sources. Research studies have complemented these efforts, but have also frequently relied on spatially-intensive but temporally-limited sampling campaigns to assess population-scale exposure (Ito et al., 2016; Wang et al., 2013).

Air pollutant concentrations, including $PM_{2.5}$, can exhibit significant spatiotemporal variability depending on local sources and features of the built environment. These variations may not be well captured by the existing monitoring regime (Apte et al., 2017; Tan et al., 2014). Improving our understanding of these variations is expected to improve our ability to characterize and reduce population exposure to air pollution (Zimmerman et al., 2020). Furthermore, increasing the spatiotemporal resolution of our measurements may support improved air pollution modeling and mapping (Jain et al., 2021), which can be used in turn to predict

E-mail address: nzimmerman@mech.ubc.ca.

https://doi.org/10.1016/j.jaerosci.2021.105872
Received 5 July 2021; Received in revised form 31 August 2021; Accepted 31 August 2021
Available online 9 September 2021
0021-8502/

concentrations in areas where limited to no monitoring exists. As such, there has been growing interest in new, lower-cost ways of measuring $PM_{2.5}$ to achieve these goals.

$PM_{2.5}$ measurement techniques can be broadly classified into optical-based methods and gravimetric methods. Gravimetric analysis relies on using physical weighing-based techniques to determine the mass collected on a tapered element oscillating micro-balance or a filter during a sampling period at a given flow rate, resulting in a time-weighted average mass concentration. These methods are the gold-standard for regulatory monitoring stations (other regulatory monitoring techniques include beta ray attenuation). Optical methods include nephelometers and optical particle counters (OPCs). These techniques rely on light scattering to determine concentration; the aerosols in the sample flow pass through a beam of infrared or visible light and the intensity of the light scattered is measured by a photodetector (Morawska et al., 2018). In nephelometers, the total light scattered by the ensemble of particles is correlated with mass measurements made by a reference monitor; it is not possible to directly reconstruct a size distribution. OPCs measure light scattered by individual particles; each pulse is assigned a size bin based on the total light intensity. This results in a size distribution that is converted to mass concentration once the entire distribution is measured (Hagan & Kroll, 2020).

While optical methods are much faster than gravimetric measurements, light scattering is sensitive to particle size distribution, shape, and composition (refractive index and absorption) (Hinds, 1999). These instruments can also be "blind" to particles below a certain diameter; most low-cost PM sensors cannot detect particles smaller than 300 nm and even more expensive optical particle counters are typically limited to particles 100 nm or larger (Hagan & Kroll, 2020).

Over the past decade, there has been increasing interest in the development of small, portable, and low-cost sensors for measuring particulate matter, and $PM_{2.5}$ concentrations specifically. The push for sensors of this nature originated in the research and occupational hygiene domain as interest in personal exposure assessments grew (Koehler & Peters, 2015). Commercially-available products for researchers and citizen scientists to purchase became widely available in 2012–2016 (Dye, 2017). These sensors typically rely on optical methods for measuring $PM_{2.5}$, making them sensitive to changes in particle size distribution and changes in particle composition which may affect particle density and aerosol optical properties (Giordano et al., 2021; Hagan & Kroll, 2020). These sensors typically lack some of the features found in regulatory-grade instruments such a sample conditioning; changes in ambient relative humidity can affect hygroscopic growth of particles (Crilley et al., 2018; Di Antonio et al., 2018; Malings et al., 2020; Zheng et al., 2018). While there are hundreds of companies with commercially-available low-cost PM sensors, many use the same internal components such as the Plantower PMS series (a nephelometer) or Alphasense OPC-N series (an OPC). As such, while implementing a low-cost PM sensor network may be overwhelming due to the choices available and companies that exist, many considerations and guidelines can still be generalized.

In this tutorial, the key guidelines for implementing a low-cost PM sensor network are reviewed and discussed. This article is also associated with a companion web-tutorial described in Section 4 (https://nzimmerman-ubc.github.io/lcs-PM-demo/) on downloading and assessing sample low-cost PM sensor data from the PurpleAir network using the new U.S. EPA Fine Particle Sensor Base Testing Guidelines (Duvall et al., 2021). While this tutorial does not cover every single consideration a researcher or citizen scientist might undertake, it covers the key areas of evaluation and calibration, siting, data reporting, and post-processing. By reading this tutorial article, I aim to improve outcomes for researchers using low-cost PM sensor networks and develop a broader community of practice.

2. Guidelines Part 1: Evaluation and calibration

A key consideration for low-cost sensor networks is that the sensors are well-calibrated and their predictive performance is well-characterized. We expect that low-cost sensors will not achieve the same accuracy and precision as their reference-grade instrument counterparts, however, there is a range of performance relative to other low-cost sensors or external standards that we might expect to achieve. This section describes approaches for evaluating low-cost sensor performance (accuracy, precision, data completeness, drift, time resolution and noise), common environmental artifacts that influence sensor performance, and some general guidelines on developing additional calibration algorithms.

2.1. Evaluation methods

When evaluating low-cost $PM_{2.5}$ sensors, there are two primary methods: (1) co-location with regulatory-grade monitors and (2) laboratory-based evaluation. The U.S. EPA refers to these as "base" testing and "enhanced" testing, respectively (Duvall et al., 2021).

During co-location evaluation ("base testing"), the low-cost sensors under consideration are co-located with a Federal Reference Method (FRM) or Federal Equivalent Method (FEM) approved instrument. Each legislative area typically maintains their own lists of FRM and FEM approved monitors; the U.S. EPA routinely updates approved monitoring equipment and methods for achieving designation (U.S. EPA, 2021a). The U.S. EPA recommends testing a minimum of three sensors simultaneously for at least 30 days. During this 30 day deployment, there should be a 75% data completeness threshold. It is also recommended that concentrations exceed 25 $\mu g\, m^{-3}$ for at least one day of the co-location period.

During laboratory evaluation ("enhanced testing"), the low-cost sensors are exposed to known concentrations of generated aerosols within an environmental chamber with controlled temperature and relative humidity. Ideally, the environmental conditions in the laboratory should mimic those to be encountered in the planned sampling environment. In addition to the effects of temperature and relative humidity, considerations for a laboratory evaluation include the range of concentrations tested (including accuracy at high concentrations), the composition of the aerosol, the size distribution of the aerosol population, and drift over time. Laboratory evaluations are typically shorter; the U.S. EPA recommends collecting a minimum of 20 pairs of time-matched sensor and FRM/FEM data points or three consecutive hours of steady-state data.

For laboratory evaluation, there is also the question of aerosol source. While the U.S. EPA makes no specific recommendation on which test aerosols to use, published studies have relied on nebulized potassium chloride aerosol (Papapostolou et al., 2017); dust (Austin et al., 2015; Li et al., 2020; Papapostolou et al., 2017; Sayahi et al., 2019; Wang et al., 2020); atomized sodium chloride, sucrose, and ammonium nitrate aqueous solutions (Li et al., 2020; Sayahi et al., 2019; Wang et al., 2015); polystyrene latex (PSL) spheres (Austin et al., 2015); incense (Hapidin et al., 2020; Li et al., 2020; Wang et al., 2020) and smoke from biomass burning (Sayahi et al., 2019). Ideally, the test aerosol should represent the target deployment environment; for example PSL spheres may not be a good choice since they are not hygroscopic and may mask the effects of humidity on accuracy (Duvall et al., 2021). The design of an environmental chamber for controlled aerosol experiments is outside the scope of this tutorial, but interested readers may refer to the publications cited in this section as well as the U.S. EPA Enhanced Testing Guidelines (Duvall et al., 2021) for more details.

2.2. Costs for performance evaluation

In general, costs to conduct co-location-based evaluations (“base” testing) will be less than conducting laboratory evaluation (i.e., “enhanced testing”). This is predicated on the assumption that field calibrations can be conducted in partnership with a local municipality who is operating a regulatory monitoring station. If you are unsure of where to locate stations with reference-grade equipment, you can begin by using the OpenAQ map feature (OpenAQ, 2020). Many municipalities will allow you to co-locate your sensors for free (or as in-kind support for a grant), but may require you to sign liability waivers or other service agreements, which can take time to process. If you are aiming to partner with a municipality, it is recommended that you begin this engagement as early as possible (ideally 3–6 months in advance) to minimize delays. For this type of partnership, materials costs can be limited to any supplies you need to install the sensors (e.g., zip ties, extension cords, etc.) and are typically less than \$1000 USD.

For “enhanced” performance assessments, one should ideally look for sensor companies that have sent their sensors for evaluation at an external validation institute (see: Section 2.7), as costs for most groups to conduct these evaluations can be prohibitive. Using the materials and instrumentation listed in Papapostolou et al. (2017), the materials costs to construct an appropriate environmental chamber capable of generating different aerosols and measuring PM mass and size distribution can be conservatively estimated to exceed \$200,000 USD. Conversely, sensor manufacturers, vendors, developers or integrators can often send their sensors for laboratory evaluation free of charge (e.g., South Coast AQMD (2021)).

2.3. Performance metrics

After conducting the laboratory or co-location tests, several key performance metrics should be reported. This tutorial focuses on metrics from sensor co-location. The U.S. EPA $PM_{2.5}$ Base Testing Report recommends reporting on sensor accuracy (i.e, how well the sensor compares to FRM/FEM monitors) and sensor–sensor precision (i.e., how consistent different sensors are); these are discussed in Sections 2.3.1 and 2.3.2, respectively. It is worth noting that some of the performance metrics (namely RMSE) can be skewed by high concentration events (such as wildfires). In the demonstration in Section 4, we will illustrate the value of separating high concentrations ($\geq$25 $\mu g\,m^{-3}$) from more typical concentrations. Other parameters one might consider reporting include mean normalized bias (MNB) (Giordano et al., 2021), detection range (minimum and maximum a sensor is capable of measuring), sensor detection limit, and response time (Morawska et al., 2018).

For each of these metrics reported in Sections 2.3.1–2.3.2, the equations for calculating these parameters are provided. For all equations in these sections we will assume that x is the FRM/FEM data and y is the sensor data. If the equation considers multiple low-cost sensors, this is denoted with variable M.

2.3.1. Sensor accuracy metrics

The U.S. EPA recommends calculating the following sensor accuracy metrics (Duvall et al., 2021):

- Coefficient of Determination (R^2) (Eq. (1))
- Slope (Eq. (2))
- Intercept (Eq. (3))
- Root Mean Square Error (RMSE) (Eqs. (4)–(5))
- Normalized Root Mean Square Error (NRMSE) (Eq. (6))

The coefficient of determination is a metric that assess how differences in one variable can be explained by the difference in a second variable (Eq. (1)). The U.S. EPA recommends that R^2 should exceed 0.70.

$$R^2 = 1 - \frac{\sum(y_i - x_i)^2}{\sum(y_i - \bar{y})^2} \tag{1}$$

Slope is calculated by finding the ratio of the "vertical change" (rise) to the "horizontal change" (run) between (any) two distinct points on a line. For a simple linear regression between the FRM/FEM data (x) and the sensor data (y), it is calculated by Eq. (2). The U.S. EPA recommends a target slope of 1.0 ± 0.35.

$$m = \frac{\sum (y_i - \bar{y}) \times (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \tag{2}$$

The intercept is the value at which the fitted line from a simple linear regression crosses the y-axis. Eq. (3) provides the intercept calculation for a simple linear regression between the FRM/FEM data (x) and the sensor data (y). The U.S. EPA recommends a target intercept range of −5 to +5.

$$b = \bar{y} - m \times \bar{x} \tag{3}$$

The RMSE is measure of the differences between the sensor data (y) and the FRM/FEM monitor data (x). It can be calculated for a single sensor (Eq. (4)) or in aggregate for multiple, M, sensors (Eq. (5)). The U.S. EPA recommends a target RMSE of ≤7 $\mu g\, m^{-3}$.

$$\mathrm{RMSE}_{\mathrm{single}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - x_i)^2} \tag{4}$$

If you wish to calculate one average RMSE across multiple (M) sensors, use Eq. (5).

$$\mathrm{RMSE}_{\mathrm{multiple}} = \sqrt{\frac{1}{N \times M} \sum_{j=1}^{M} \sum_{i=1}^{N} (y_{ij} - x_i)^2} \tag{5}$$

The NRMSE helps account for periods when ambient concentrations are high (which may skew the RMSE). It is calculated by normalizing the RMSE by the average concentration measured by the FRM/FEM monitor (Eq. (6)). The U.S. EPA recommends a target NRMSE of ≤ 30%.

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\bar{x}} \times 100 \tag{6}$$

2.3.2. Sensor precision metrics

The U.S. EPA recommends calculating the following sensor accuracy metrics:

- Standard Deviation (Eq. (7))
- Coefficient of Variation (Eq. (8))

The standard deviation (SD) metric is calculated as the standard deviation for each of the N simultaneous sensor $PM_{2.5}$ concentration measurements by multiple (M) sensors. It is calculated using data from all co-deployed low-cost sensors using Eq. (7). The U.S. EPA recommends a target SD of ≤ 5 $\mu g\, m^{-3}$.

$$SD = \sqrt{\frac{1}{(N \times M) - 1} \sum_{j=1}^{M} \sum_{d=1}^{N} (y_{dj} - \bar{y_d})^2} \tag{7}$$

In Eq. (7), M is the number of identical sensors operated simultaneously, N is the number of periods during which all instruments are operating and reporting valid data, y_{dj} is the average concentration for day or hour d and sensor j, and $\bar{y_d}$ is the average concentration across all sensors for day or hour d.

The coefficient of variation (CV) can subsequently be calculated after calculating the standard deviation. The CV is the relative standard deviation (SD) divided by the deployment averaged sensor $PM_{2.5}$ concentration across the field test (Eq. (8)). The U.S. EPA recommends a target RMSE of ≤ 30%.

$$CV = \frac{SD}{\bar{y}} \times 100 \tag{8}$$

2.4. Data completeness

In general, the target for time online of a low-cost sensor should be >75% of the time (Duvall et al., 2021). It is also recommended that you consider a system with both cloud-based and local storage options (e.g., cellular connectivity posting data to web server as well as SD card backup) as certain locations may have poor cellular data coverage. Specific examples of locations where cellular connectivity may be poor include basements, locations obstructed by trees, and rural locations. Additionally, it is not uncommon for low-cost sensors to have older modem technology internally (e.g., 3G) which may also result in a poorer signal.

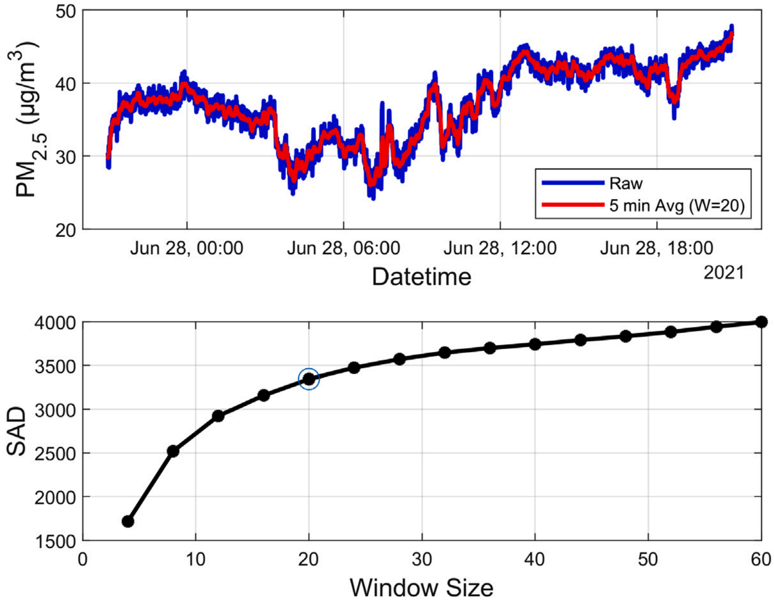

Fig. 1. Using the sum of absolute difference (SAD) to determine window size. Here we can see the SAD starts to plateau around a window size of 20. The raw data is 15 s resolution, so a window size of 20 corresponds to a 5 min moving average.

2.5. Drift

Drift in low-cost PM sensors is generally believed to result from either degradation of electrical components within the sensor or from dust accumulation in the sensor (Bulot et al., 2019). Bulot et al. (2019) suggest that sensors facing downward may help reduce the build-up of dust inside the sensors.

US EPA recommends a 60-day test for drift (Duvall et al., 2021). If laboratory/“enhanced” testing is possible, the U.S. EPA recommends conducting two sets of tests under steady-state conditions 60 days apart. During the 60 day gap, the sensors should be operated outdoors measuring ambient air. If no laboratory-based evaluation is possible, you may consider comparing agreement with FRM/FEM monitors via co-location using a longer co-location window (60+ days) or via co-location before and after 60 days of operation. If there is substantial drift, this will be reflected in changing slopes before and after deployment; you may also consider plotting agreement between sensors and FRM/FEM monitors and color coding the data points by time to assess drift or by plotting normalized concentrations as a function of time.

2.6. Time resolution and noise

For calculation performance metrics, the U.S. EPA recommends comparing both 24 h and 1 h averages to FRM/FEM monitors (Duvall et al., 2021). They place extra emphasis on 24 h averages since FRM and FEM monitors are expected to be the approximately the same for a 24-h average, whereas for hourly averages there may be more reference instrument variability. Essentially, assessing performance metrics on a 24-h average expands reference monitor eligibility to both FRM and FEM monitors and as such generates more co-location options. However, most sensor data will be used at 1 h resolution or faster, so it is also recommended to evaluate them across time resolutions you plan to use in your deployment.

When choosing a minimum data reporting interval for low-cost sensors there is generally a trade-off between time resolution and noise and error (i.e., data at faster intervals is noisier and/or less accurate). To reduce instrument signal noisiness and down-average the data, a common approach applying a moving average. There is no definitive method for choosing a minimum averaging window for your moving average, but one option is calculating the sum of absolute differences (SAD) between the smoothed signal (S) and the raw signal (R) (Eq. (9)) for different window sizes, and visually inspecting the results for a plateau (Wiktorski & Królak, 2020). For some sample data collected by a SENSIT RAMP (which uses a Plantower PM sensor) on the UBC campus, we see that the SAD begins to stabilize around a window size of 20 (Fig. 1); since the raw reporting interval is 15 s, a window size of 20 corresponds to a 5-min moving average. From this we would conclude that we should always *at a minimum* average our data to 5-min moving averages.

$$SAD = \sum_{i=1}^{N} |S_i - R_i| \tag{9}$$

2.7. Validation institutes and programs

Conducting all of the performance metrics described in Section 2.3 can be time, labour, and cost intensive. While the general recommendation remains that these tests be conducted in an environment that matches your deployment environment as closely as possible (e.g., if you want to use low-cost PM sensors in Vancouver, conduct base/co-location testing in Vancouver), when choosing or comparing low-cost PM sensors there are also several established validation institutes that can help inform your decision. Some examples of these include the Air Quality Sensor Performance Evaluation Center (AQ-SPEC), run by the California South Coast Air Quality Management District located in Southern California (Feenstra et al., 2019; Papapostolou et al., 2017), EuNetAir in Europe (Borrego et al., 2018), and the U.S. EPA Air Sensor Toolbox (U.S. EPA, 2021b). These initiatives often report at a minimum the R^2 and some assessment of measurement error (RMSE or mean absolute error, MAE). If you are not sure which sensor to purchase, these validation institutes are helpful to narrow down candidate sensors.

2.8. Known influences

Any environmental conditions that affect an aerosol's optical properties may influence sensor accuracy. One of the most commonly observed influencing factors on low-cost PM sensor accuracy is relative humidity (RH) (Hagan & Kroll, 2020; Jayaratne et al., 2018; Malings et al., 2020; Wang et al., 2015). If a particle is hygroscopic, it will uptake more water at higher RH. This in turn results in larger particles (Petters & Kreidenweis, 2007), which scatter more light and can result in overestimated mass concentrations. Furthermore, while explored in detail in fewer studies, this water uptake can change the refractive index and density of the aerosol, making corrections more complex (Hagan & Kroll, 2020). To avoid these issues in FRM/FEM monitors, the U.S. EPA requires sample conditioning to an RH range of 30%–40% (Chow & Watson, 2008). Since the vast majority of low-cost PM sensors do not include any sample conditioning, this is a primary source of measurement error for low-cost PM sensors. To account for this, many low-cost PM sensor calibrations include parameters influencing water uptake (e.g., RH, temperature, dew point) in their calibration models (see Section 2.9 for more discussion on this).

Other aerosol properties that may vary during a sensor calibration or deployment and influence accuracy include the refractive index (RI) of the aerosol (which will affect the optical properties) and the aerosol size distribution (Hagan & Kroll, 2020). The main conclusions from Hagan and Kroll (2020) are that if you calibrate your sensor with a source type that has substantially different RI than your sampling environment you may have significant over-estimation of PM mass if the RI of sampled aerosol exceeds calibration material and vice versa. This effect was found to be more pronounced in OPCs than in nephelometers, since OPCs assign a pulse signals to one bin and the bins are large; potential bin misclassification due to RI mismatch can lead to outsized errors. Conversely, low-cost nephelometers tend to fail if the particle size distribution changes drastically during sampling since they are calibrated to a single size distribution.

2.9. Calibration methods

If the raw sensor performance (as assessed in Section 2.3) is deemed unsatisfactory, you might consider constructing a calibration model to correct your data. Calibration model building is not the primary focus of this tutorial and has been discussed at length in a number of other papers and review articles (Cavaliere et al., 2018; Crilley et al., 2018; Di Antonio et al., 2018; Giordano et al., 2021; Maag et al., 2018; Malings et al., 2020; Patra et al., 2021; Wang et al., 2015; Zheng et al., 2018).

Briefly, post-processing calibration algorithms can be generally classified under three categories: (1) mechanistic models with corrections for hygroscopicity using κ-Köhler theory (Crilley et al., 2018; Di Antonio et al., 2018; Malings et al., 2020); (2) empirical models using nonlinear or multiple linear regression equations (Cavaliere et al., 2018; Malings et al., 2020; Zheng et al., 2018); or (3) machine learning algorithms (Kumar & Sahu, 2021; Patra et al., 2021; Wang et al., 2019).

In general, a mechanistic calibration model should be preferred over an empirical model, as it can account for influences of particle composition or hygroscopicity (Giordano et al., 2021). Developing such corrections may be possible with an "enhanced" testing set-up as described by the U.S. EPA where aerosol composition and size distributions may be more easily varied. In practice, however, conducting the experiments necessary for a mechanistic model is often unfeasible and/or too expensive and empirical or machine learning models must be used. Such models can be developed using the co-location or U.S. EPA "base" testing approach, with a goal of co-location location, meteorological range, and expected aerosol sources (e.g., traffic) closely matching the target deployment environment. In terms of preference for empirical vs machine learning models, we suggest a parsimonious approach; model simplicity and interpretability should be prioritized and machine learning models should not be used unless empirical models have been shown to fail. Input variables into empirical and machine learning models may include temperature, RH and dew point as well as any other simultaneously measured pollutants (e.g., criteria air contaminants). Inclusion of these variables may help resolve some of the impacts of RH and particle water uptake or impacts of changing source on RI (if other criteria air contaminants are present in characteristic ratios for different sources). Additionally, empirical models may also be built for different concentration ranges or dominant sources, since it has been shown that for wildfire smoke (which creates high PM concentrations with different aerosol composition than typical ambient air), different empirical model coefficients are needed (Holder et al., 2020). One advantage of machine learning models is that they are less susceptible to model over-fitting; however, to choose variables, it is recommended that one starts with fewer variables and assesses whether R^2 increases by more than 1% on the withheld testing data, as is common in other model building exercises, such as land use regression (Jain et al., 2021).

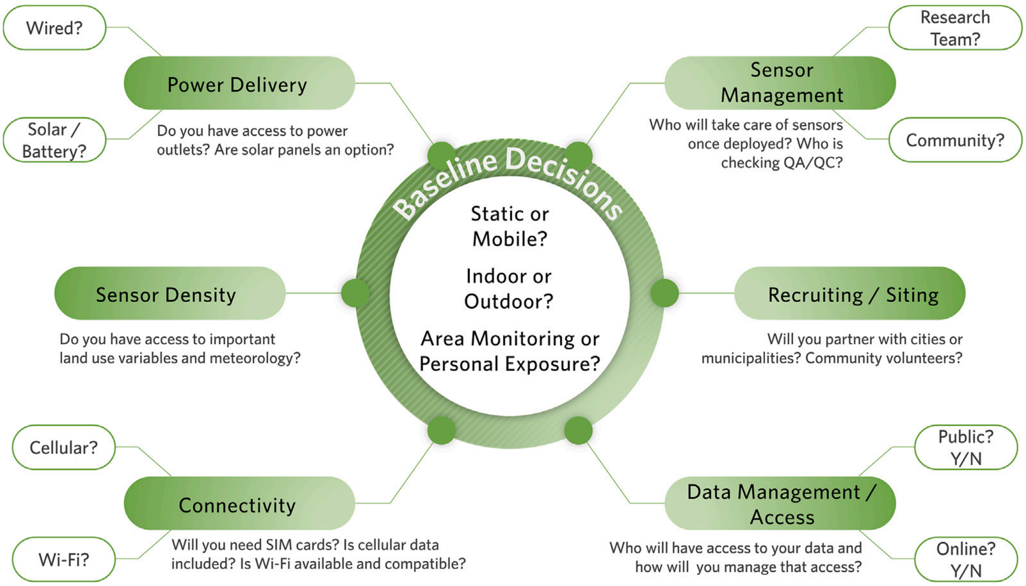

Fig. 2. A summary of key decisions that are required when developing your sensor network.

In the demonstration section of this article (Section 4), we will demonstrate building an empirical calibration model with sensor signal, temperature, RH and dew point as our input variables. While many low-cost sensors measure temperature and RH internally, these are not always reliable (if they are affected by internal heating or conditions within the sensor), and so models should be built using external meteorological stations wherever possible. If dew point (DP, °C) is not directly measured, it can be estimated using Eq. (10), where RH is in % and T is temperature in °C.

$$DP = 243.04 \times \left[\frac{\ln \frac{RH}{100} + \frac{17.625 \times T}{243.04 + T}}{17.625 - \ln \frac{RH}{100} - \frac{17.625 \times T}{243.04 + T}} \right] \tag{10}$$

Typically in the calibration model building process, many models are constructed and compared to determine which model achieves optimal performance. One useful tool for comparing candidate calibration models is the Taylor diagram (Jolliff et al., 2009). A Taylor diagram provides a quick visual snapshot of three model performance metrics: (1) standard deviation, as measured based on radial distance from the origin to the marker on the plot; (2) correlation coefficient (Pearson r); and (3) Centered RMSE (which is the RMSE corrected for bias). A sample Taylor diagram based on the demonstration in Section 4 is provided in Fig. 7.

3. Guidelines Part 2: Logistics and sensor selection

When choosing where and how low-cost PM sensors will be deployed, there are a number of questions that should be addressed including: what is my measurement environment (indoor, outdoor or both); do I want static or mobile monitoring; and I interested in area monitoring or in personal exposure? After making these baseline decisions, there are a number of key considerations for deploying a low-cost sensor network including power delivery, sensor density, connectivity, data management and access, and sensor management (Fig. 2). Each of these concepts are addressed in this Section.

3.1. Siting and sensor density

3.1.1. Power delivery and site logistics

There are several considerations when choosing where to site your low-cost sensors during field deployments. For static deployments, there are practical considerations, such as access to electrical power outlets, suitable mounting structures (e.g., light poles), and sensor security/safety. In general, sensors with batteries and solar panels will provide more flexibility in instrument siting in outdoor contexts, as access to electrical outlets can often be a limiting factor. If relying on solar panels, you will need to ensure that they are appropriately sized and sited to maintain charge in the minimum sunlight conditions you expect to encounter. Indoors, it is generally safe to assume that wired/plug-in access is available, unless you are monitoring in a remote location with limited access to grid electricity. For mobile air quality monitoring (e.g., wearables or mounted on bicycles or vehicles), battery-powered rechargeable units are best, as it is unlikely the monitors will be used continuously.

3.1.2. Sensor density and location recruitment

Other considerations include identifying the deployment goals. If the goals are to characterize known air pollution sources, sensors could be placed on a transect following the prevailing wind direction passing through the known source location. If the goal is to characterize population exposure, sensor locations may be chosen to represent a range of socioeconomic conditions based on census data (if focus is environmental justice), be placed near vulnerable populations (e.g., daycares, senior care facilities, schools), or be placed in known gathering spots (e.g., parks). If the goal is area monitoring, then sensors should be stratified to cover the geographical area of interest as well as stratify land use features such as vehicle density, commercial area, and population density. A common denominator across all cases is access to data sets used to inform these choices. This might include historical meteorological data (for prevailing wind direction), geospatial data (for land use characteristics), and census data. Most University libraries maintain geospatial databases, as do most municipalities. For example, to develop candidate locations in Metro Vancouver, one might consult the Metro Vancouver Open Data Catalogue (Metro Vancouver, 2016) for land use data, the Canadian Urban Environmental Health Research Consortium (CANUE, 2021) for socioeconomic status data, and the BC Ministry of Forests, Lands, Natural Resource Operations and Rural Development GeoBC Data Catalogue (GeoBC, 2021) for historical meteorological data.

Once you have identified candidate locations based on the above criteria, the next step is deciding which sites can be used based on volunteers and/or collaborations with relevant partners (e.g., cities who may let you mount sensors on light posts). While many low-cost PM sensor siting decisions are limited to having willing sensor hosts (e.g., via recruitment of citizen scientists/crowd-sourcing), there are also mathematical methods to help optimize sensor placement. One technique growing in popularity is constructing a Gaussian Process (GP) model of air pollutant concentrations based on existing monitoring data and then using the GP model to inform new site selection (Krause et al., 2008). This is done by taking the covariance matrix from the trained GP model and combining it with the number of sensors available to place, coordinates for candidate locations and coordinates for unsuitable locations (e.g., over water). The algorithm proposed by Krause et al. (2008) suggests placing sensors at locations that most significantly reduce the collective posterior uncertainty of the GP. They refer to this method as the Optimal Experimental Design (OED) algorithm. For a demonstration of this applied to an air quality data set in London, see Boghozian (2021).

3.2. Connectivity and data management

3.2.1. Connectivity

If large numbers of sensors (100+) are to be deployed, questions around network connectivity and managing a large number of connections emerge (Motlagh et al., 2020). Important considerations when choosing a low-cost PM sensor are what modem the sensor is equipped with (e.g., LTE-4G, NB-IoT), and if it can be potentially upgraded to a 5G modem that may provide the low-latency needed to process the large amounts data in real-time. The implementation of 5G would also enable the utilization of edge computing, which could enable real-time application of complex calibration algorithms, support computations for faster decision making, or communicate risks in real-time to people nearby (Han et al., 2019). If as a sensor user you expect your network to grow, keeping these considerations in mind early on may support a more robust low-cost sensor network as you expand.

Conversely, for deployments that are either largely indoor or at static locations with reliable internet access (e.g., schools), a sensor that relies on Wi-Fi may be preferred. This can eliminate two challenges of cellular network connectivity: (1) poor signal in certain indoor locations (e.g., near concrete or brick) and (2) difficulty accessing SIM cards and data plans for the sensors (unless included with the sensor).

3.2.2. APIs and data reporting

Accessing your sensor data or providing sensor access to others is another key consideration. While some low-cost sensor providers provide access and analysis of low-cost PM sensor data via paid private web portals using a software-as-a-service (SaaS) model, others provide public access and Application Programming Interface (API) keys (e.g., PurpleAir, Air Quality Egg). Feenstra et al. (2020) state that low-cost PM data can *only* be considered open access when the data is available through a stable and consistent API. The value of open access data is that it allows the freedom for researchers and data scientists to develop custom applications to display and report data in their own useful and/or meaningful ways.

When storing and sharing air quality data from low-cost PM sensors, choosing public vs. private data reporting is research question dependent (i.e., if any data is sensitive, password protection may be preferred). The net impact of a given project or product may be higher with a public-facing interface, as others may glean useful insights from your data beyond its initial purpose for collection. For example, in Section 4, we will demonstrate the usefulness of the public PurpleAir data by using free open-source software to run our tutorial on assessing sensor performance using the new U.S. EPA Fine Particle Sensor Performance Guidelines (Duvall et al., 2021).

If you choose to develop your own public-facing application or map of your low-cost PM data, Feenstra et al. (2020) recommend that if you choose to display both data from both regulatory-grade instruments and low-cost PM sensors, that the source and type of data displayed be readily apparent to minimize confusion for end users if there is mismatch or inaccuracy in your sensor data. You might also consider hosting your low-cost PM sensor data on an existing application, such as OpenAQ which now accepts low-cost sensor data (OpenAQ, 2020). To make your data compatible with an external application such as OpenAQ, consider formatting your data according to OpenAQ's standards (OpenAQ, 2019). These standards require a JSON format with specific units for pollutants and also requires that dates are stored as a Javascript date containing both the local time and the UTC time.

3.3. Sensor management

Once the locations for the sensors have been established and the sensors have been mounted, there is also the question of sensor management. Ideally this question should be addressed before the network is deployed. A review by Morawska et al. (2018) found that in-depth expertise is typically required to manage sensor networks, interpret the data, and identify relevant measurement artifacts. As such, one should plan not just for deploying a sensor network but also conducting ongoing maintenance. For community-based projects, this entails outlining clear management frameworks for any volunteers (Will they be expected to check their own sensors? How will you get in touch if a sensor needs replacement?). For research-community partnerships this also entails being transparent in how long it may take to process and report data back (typically 12–18 months for the standard graduate-student led project). If QA/QC tasks can be automated via scheduled scripts (data download, checking that monitors are online), that is also strongly recommended. Campbell et al. (2013) review common QA/QC considerations for environmental sensors; most of their considerations are also applicable to low-cost sensor networks.

3.4. Interactive guidance tool

To support community members with their own sensor selections, the companion website for this article (https://nzimmerman-ubc.github.io/lcs-PM-demo/) includes a tab labeled "Sensor Guidelines Quiz" where you can answer 3 questions about the type of monitoring you are interested in doing, and it will provide some general recommendations.

4. Demonstration: Evaluating three PurpleAir monitors

To illustrate how to work with some sample low-cost PM sensor data, we will take advantage of a large, established network: PurpleAir (PurpleAir, 2021). The tutorial will study some sensors deployed in the Greater Vancouver Area, but this code could be easily adopted to different PurpleAir sensors by adjusting the sensor identifier.

Using the PurpleAir network in this tutorial is convenient, since the R Package "AirSensor" (Callahan et al., 2020) includes easily implementable tools for downloading PurpleAir data from their network directly into R. The development of the AirSensor package is also detailed by Feenstra et al. (2020). In this tutorial, we will download the averaged A–B channels of the PurpleAir data for assessment.

Metro Vancouver, the local regulatory agency, has two PurpleAir monitors co-located with their 'Burnaby South' Regulatory Monitoring Site (Metro Vancouver, 2021). This is a good site for our tutorial, as there is both regulatory data and low-cost sensor data available. The U.S. EPA recommends that when assessing sensor performance, a minimum of three co-located sensors be assessed (Duvall et al., 2021). While this is not exactly possible here, for demonstration purposes only we will also scrape data from an additional PurpleAir monitor located approximately 1 km east of the Metro Vancouver Burnaby South Regulatory Monitoring Station for our analysis. We expect data at this site ('Burnaby3') to roughly match the data at the other two sites ('Burnaby1' and 'Burnaby2') as this is an urban background site in the Greater Vancouver Area with few local sources (3-year average $PM_{2.5}$ 2016–2018 of 5.7 $\mu g\, m^{-3}$ Environmental Reporting BC, 2021). We will download data from September 1 2020–December 31 2020 (4 months — including a wildfire smoke episode in September 2020).

When the U.S. EPA released their performance testing protocols for fine particulate matter sensors, they also released a "Fillable Reporting Template". The aim of the U.S. EPA protocol was an attempt to standardize the way sensor performance is reported. To that end, this part of the tutorial builds all of the relevant plots and tables required in the Fillable Report, with all code provided on the companion web tutorial. We will use our three Burnaby South PurpleAir monitors and the co-located FEM data for this purpose.

This analysis is not extensive; there are numerous details and caveats mentioned within the full U.S. EPA protocol (Duvall et al., 2021). It is strongly recommended that in addition to this tutorial, any low-cost PM sensor users read the U.S. EPA report in full. Additionally, this tutorial is only applicable to "Base Testing" (i.e., field tests). As mentioned in Section 2.3, the U.S. EPA also recommends "Enhanced Testing" based on lab-based analyses. This is not covered here.

4.1. Software requirements

This demonstration was developed in R version 4.1.0 (R Core Team, 2020) using the RStudio IDE (RStudio Team, 2020). The tutorial uses several R packages that are all cited in the companion web tutorial accessible at https://nzimmerman-ubc.github.io/lcs-PM-demo/. The tutorial also uses many packages from the "tidyverse" ecosystem (Wickham, 2019) and follows their general coding philosophy.

4.2. Time series and scatter plots

Our first assessment is a visual assessment of the data quality. This will be achieved through time series and scatter plots of $PM_{2.5}$ concentrations by both the FEM and PurpleAir monitors for both 1 h and 24 h averages (Fig. 3).

From the time series plots, we clearly see a period where the PurpleAir sensors are overestimating $PM_{2.5}$ concentrations (early October). The scale on the y-axis is also influenced by a wildfire smoke episode which occurred in mid-September 2020. From these scatter plots zoomed in to 30 $\mu g\, m^{-3}$, we can see that the PurpleAir monitors tend to slightly overestimate concentrations (data falls above 1:1 line). We will explicitly calculate the slope in Section 4.3.

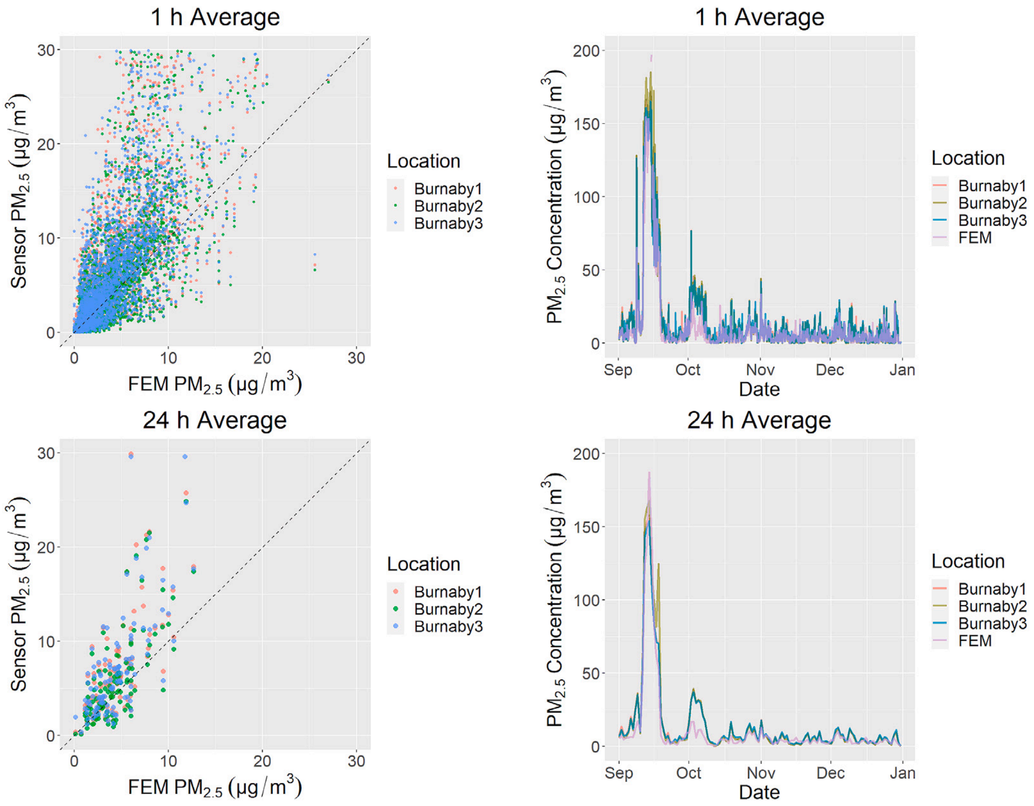

Fig. 3. FEM data compared to three PurpleAir monitors (Burnaby 1–3). Comparisons are shown both as a scatter plot (left) and a time series plot (right).

Table 1
Summary of accuracy performance metrics for three PurpleAir monitors in Vancouver (Burnaby1-3) for both 1 h and 24 h averages. Metrics include R^2, slope (m), intercept (b), RMSE ($\mu g\,m^{-3}$) and NRMSE (%).

Sensor ID	1 h					24 h				
	R^2	m	b	RMSE	NRMSE	R^2	m	b	RMSE	NRMSE
Burnaby1	0.91	1.00	3.43	8.77	86.03	0.94	0.96	3.66	7.64	67.29
Burnaby2	0.71	1.12	2.45	18.55	181.98	0.91	1.06	2.80	10.13	89.28
Burnaby3	0.90	0.98	3.59	8.72	85.54	0.93	0.93	3.83	7.65	67.42
Mean	**0.84**	**1.03**	**3.16**	**12.02**	**117.85**	**0.93**	**0.98**	**3.43**	**8.47**	**74.66**

4.3. Performance assessment

4.3.1. Sensor-FRM/FEM accuracy

We will calculate accuracy metrics for both 1 h and 24 h averages, as recommended. We will use built-in linear regression functions in R to calculate slope, intercept, and R^2. The RMSE and NRMSE are calculated manually as per Eqs. (4)–(6). Performance is summarized in box plots (Fig. 4) and as tabular statistics (Table 1).

From Table 1, we can see that the performance of sensor 'Burnaby2' is worse than the other two sensors. Based on this performance assessment, it is recommended to assess the health of this specific monitor as it has lower R^2 and more than double the RMSE and NRMSE of its counterparts. A full performance summary following all analyses is provided at the end of Section 4 in Table 4.

4.3.2. Sensor–sensor precision

As with the accuracy metrics, we will calculate sensor–sensor precision metrics for both 1 h and 24 h averages. The SD and CV (%) are calculated manually as per Eqs. (7)–(8). Since only one value is reported for SD and CV (%), performance is summarized as tabular statistics only (Table 2).

From this assessment, the standard deviation is higher than the target of 5 $\mu g\,m^{-3}$, however, the CV (%) is fairly low, suggesting fairly good sensor–sensor precision. The high SD was likely influenced by the wildfire period when concentrations exceeded 150 $\mu g\,m^{-3}$.

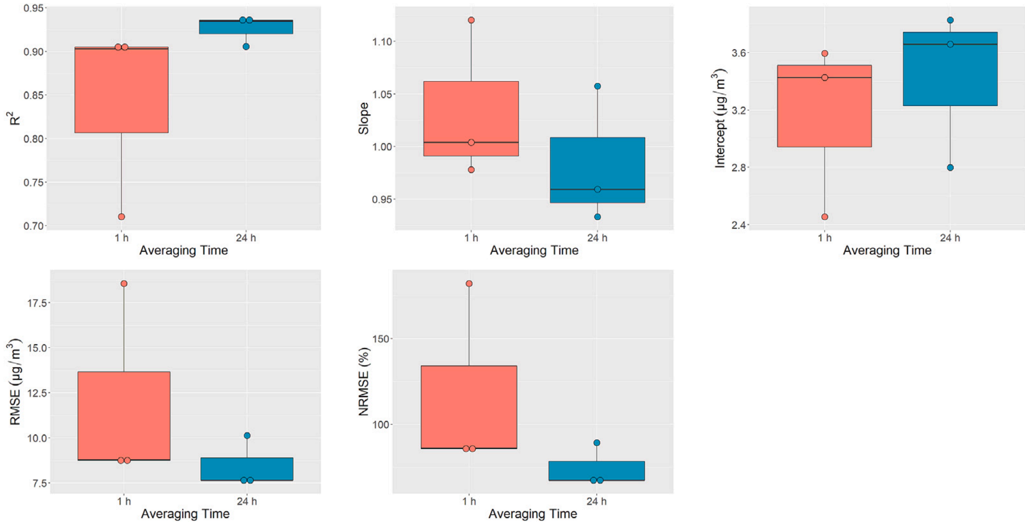

Fig. 4. Accuracy performance metrics for three PurpleAir monitors. The box plots represent the variability across the three PurpleAir monitors (whiskers = range calculated, solid line = median, box edges = 25th–75th percentile).

Table 2
Standard deviation (SD) and coefficient of variation (%) between the three PurpleAir monitors.

1 h		24 h	
SD ($\mu g\,m^{-3}$)	CV (%)	SD ($\mu g\,m^{-3}$)	CV (%)
51.67	7.59	18.42	2.69

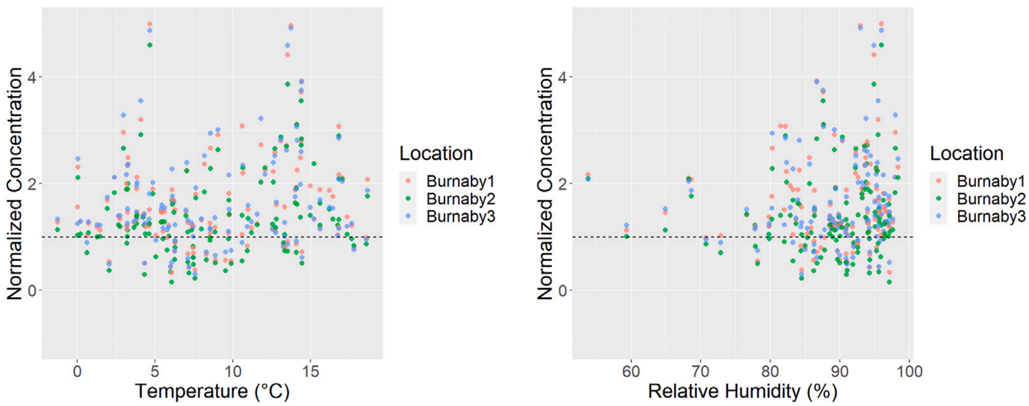

Fig. 5. Assessing the influence of temperature and relative humidity (RH) on normalized concentrations (PurpleAir/FEM) from monitors (Burnaby 1–3). If normalized concentration is >1, it means the PurpleAir is overestimating relative to the FEM and vice versa. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.4. Meteorological influence

In this section, we plot the normalized $PM_{2.5}$ concentration (defined as Sensor $PM_{2.5}$/FEM $PM_{2.5}$) for each 24 h period against the monitored temperature and relative humidity. We will also show scatter plots of FEM vs Sensor $PM_{2.5}$ as colored by RH. To do this, we must first join weather data to our sensor data and then calculate normalized concentrations (Fig. 5). As mentioned in Section 2.9, temperature and relative humidity sensors inside low-cost sensor units are often influenced by their internal environment and thus for these analyses external meteorological data should be used. In this tutorial, the data from the nearest meteorological station in Delta, BC is used. The web tutorial provides an overview of R packages capable of extracting the closest weather station data.

From Fig. 5, we do not see any obvious temperature effect, similar to the conclusions in Giordano et al. (2021); normalized concentrations appear to be fairly evenly distributed across the temperature range. For RH, we see a clear influence, with normalized

Table 3
Summary of accuracy performance metrics for three PurpleAir monitors in Vancouver (Burnaby1-3) for both 1 h and 24 h averages split into concentrations <25 and ≥25 μg m^{-3}. Metrics include R^2, slope (m), intercept (b), RMSE (μg m^{-3}) and NRMSE (%).

<25 μg m^{-3}	1 h					24 h				
Sensor ID	R^2	m	b	RMSE	NRMSE	R^2	m	b	RMSE	NRMSE
Burnaby1	0.56	1.57	0.44	6.72	143.22	0.69	1.92	−1.26	6.12	125.43
Burnaby2	0.56	1.59	−0.39	6.51	138.80	0.68	1.94	−2.09	5.95	121.92
Burnaby3	0.56	1.55	0.59	6.62	141.15	0.69	1.86	−0.90	5.96	122.13
Mean	**0.56**	**1.57**	**0.21**	**6.62**	**141.06**	**0.69**	**1.91**	**−1.42**	**6.01**	**123.16**
≥25 μg m^{-3}	1 h					24 h				
Sensor ID	R^2	m	b	RMSE	NRMSE	R^2	m	b	RMSE	NRMSE
Burnaby1	0.81	0.77	29.92	23.65	25.22	0.86	0.76	27.38	18.87	18.22
Burnaby2	0.24	0.72	49.74	70.20	74.87	0.66	0.64	53.85	32.58	31.46
Burnaby3	0.79	0.74	30.82	23.77	25.34	0.85	0.73	28.25	19.67	19.00
Mean	**0.61**	**0.75**	**36.83**	**39.20**	**41.81**	**0.79**	**0.71**	**36.50**	**23.71**	**22.89**

concentrations increasing sharply at RH >80%. This suggests a post-processing calibration algorithm is required; this is discussed in Section 4.8.

4.5. Separating high concentration events

Some of our performance metrics may be skewed by periods with very high concentrations (e.g., wildfires). We will also calculate accuracy performance metrics separately during high concentration episodes and lower concentration periods. Here we will look at performance when splitting based on whether concentrations exceed 25 μg/m^3 (Table 3).

In this scenario, the split analysis provides useful insight: the PurpleAir monitors are tending to overestimate at low concentrations (slopes >1) and underestimate during the wildfire period (slopes <1). This suggests we may consider trying a piecewise calibrations for low and high concentrations, such as what is discussed in the literature (Malings et al., 2020). We will demonstrate this in Section 4.8. Additionally, in Table 3, we see that normalized sensor accuracy (NRMSE) and sensor-FEM correlations (R^2) are substantially improved for high concentration events. It is also clear that during the high concentration events, sensor 'Burnaby2' performed very poorly, however, its performance is more similar to the other Burnaby sensors during low concentrations.

4.6. Other tabular data

In addition to the tabular data already provided within the "Performance Metrics" section, it is also recommended to provide tabular data on uptime and paired valid data set counts. Specifically, it is recommended that you report the range of FRM/FEM monitor concentrations over the duration of the test (μg m^{-3}), sensor uptime, the number of 24 h periods in FRM/FEM monitor measurements with a concentration of ≥ 25 μg m^{-3}, the number of 24 h periods outside of manufacturer-listed temperature and RH criteria, the number of concurrently reported sensor values (1 h and 24 h), the number of paired sensor and FRM/FEM values (1 h and 24 h), and the number of paired 24 h normalized concentration and T and RH values.

These results are only provided in tabular format in the web tutorial for brevity. The uptime for this assessment was an impressive 100%, and all data fell within the manufacturer-listed temperature and RH target criteria. During this assessment there were 8 days of the total 122 day assessment where concentrations were ≥ 25 μg m^{-3}. There were 2895 pairs of 1 h FEM-low-cost sensor data, 2905 pairs of sensor–sensor data; on a 24 h scale there were 122 pairs of FEM-low-cost sensor data points and 122 pairs of sensor and temperature and RH values.

4.7. Performance summary

We have now completed all of the main reporting metrics requested in the U.S. EPA Base Testing $PM_{2.5}$ Report (Table 4). From this analysis, we can conclude that without calibration, these sensors fail to meet some targets (RMSE, NRMSE, SD), but in general are well-correlated to both the regulatory monitoring data and each other, especially during high concentration events that are likely of most concern for human health. We also identified in this analysis that the health of the 'Burnaby2' sensor seems poor and this sensor should be inspected and potentially replaced.

Given that low-cost PM sensors are susceptible to aging and drift (laser power will reduce over time, dust can accumulate on sensor surfaces), and meteorology can change seasonally (affecting parameters such as dominant aerosol source and particle hygroscopicity), it is recommended that such evaluations are repeated 2–4 times a year for continuous deployments (ideally, seasonally). This will help identify sensors due for replacement and help track long-term sensor degradation. Repeat evaluations may consider shorter co-location time periods once baseline performance has been established; consider 2–3 weeks, 2–4 times per year beyond initial performance assessment). Additionally, if sensors are exposed to high concentrations for prolonged periods of time (e.g., if used near-source and exposed to high concentrations of smoke), a new performance evaluation and calibration is recommended prior to re-deployment due to potential accelerated aging of sensors.

Table 4
A full summary of the sensor performance metrics compared to the U.S. EPA Base Testing recommendations, along with some commentary.

Category	Metric	Pass/Fail?	Comments
Accuracy	R^2	Pass	Target is >0.7 and the mean across all sensors was 0.84 (1 h) and 0.93 (24 h).
	Slope	Pass	Target is 1 ± 0.35 and the mean across all sensors was 1.03 (1 h) and 0.98 (24 h).
	Intercept	Pass	Target is −5 to +5 and the mean across all sensors was 3.16 $\mu g\,m^{-3}$ (1 h) and 3.43 $\mu g\,m^{-3}$ (24 h).
	RMSE	Fail	Target is ≤ 7 $\mu g\,m^{-3}$ and the mean across all sensors was 12.02 $\mu g\,m^{-3}$ (1 h) and 8.47 $\mu g\,m^{-3}$ (24 h). This was influenced heavily by sensor "Burnaby2" which seems to have some issues.
	NRMSE	Fail	Target is ≤30% and the mean across all sensors was 117.85% (1 h) and 74.66% (24 h). This was likely influenced by the wildfire period. NRMSE improved for ≥25 $\mu g\,m^{-3}$ to generally <30%.
Precision	SD (%)	Fail	Target is <5 $\mu g\,m^{-3}$ and the mean across all sensors was 51.67 $\mu g\,m^{-3}$ (1 h) and 18.42 $\mu g\,m^{-3}$ (24 h). This was likely skewed by the poorer performance of "Burnaby2" and the high concentrations during the wildfire period.
	CV	Pass	Target is <30%. For both 1 h and 24 h calculations this was <10%.
Other	Uptime	Pass	Uptime was 100%.
	Paired observations	Pass	Acceptable number of paired observations. Minimum of 30 consecutive days.
	High concentrations	Pass	Must be ≥25 for at least one day. 8 days achieved these concentrations.

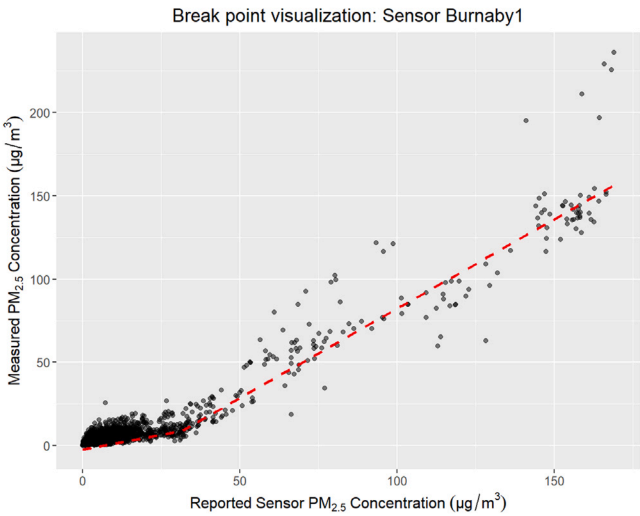

Fig. 6. Break point analysis is one means of fitting different calibration models to different concentration ranges. Here, the break point is at approximately 30 $\mu g\,m^{-3}$.

4.8. Calibration model demonstration

The last thing we will consider is building a calibration model. Many low-cost PM sensors have relied on multiple linear regression calibrations with parameters such as temperature, relative humidity and dew point (e.g., Malings et al. (2020)). Many low-cost sensors measure temperature and relative humidity directly, however, these are not always reliable (if they are affected by internal heating or conditions within the sensor), and so models should be preferably be built using external meteorological stations if possible. Let us consider building a calibration model for sensor "Burnaby1" at 1 h time resolution and look at some techniques for assessing model performance.

Before we begin, we will need to extract the relevant data for model building and separate it into testing and training data sets. A typical testing–training split is 80:20. For a full discussion of data splitting techniques, including k-folds cross-validation, see Giordano et al. (2021). Purely for demonstration purposes, we will consider every linear combination of the sensor signal, the relative humidity and the dew point in our model building. We will exclude temperature, as we previously determined that temperature was not correlated with sensor-FEM agreement (Section 4.4). Given the results of separating the high concentration and low concentration and the change in slope (Section 4.5), we will also try three piecewise regression models. All of the candidate models we are considering for this demonstration are provided in Table 5. We will duplicate models A, B and D using the piecewise regression approach. As described in the web tutorial, we have determined an appropriate segment split around a concentration of 30 $\mu g\,m^{-3}$. This is visualized in Fig. 6.

From Fig. 7, we can see that even without any corrections, the PurpleAir monitors are well correlated to the FEM monitors (model D). The segmented models (E-G) have notably better performance than the linear models across all concentration ranges,

Table 5
Empirical models tested as part of this demonstration. Note that models E through G are segmented models with a split at ≈ 30 $\mu g\,m^{-3}$.

Model	Formula
A	FEM = $\beta_0 + \beta_1 \times$ (sensor signal) $+ \beta_2 \times RH + \beta_3 \times T_{dew}$
B	FEM = $\beta_0 + \beta_1 \times$ (sensor signal) $+ \beta_2 \times RH$
C	FEM = $\beta_0 + \beta_1 \times$ (sensor signal) $+ \beta_2 \times T_{dew}$
D	FEM = $\beta_0 + \beta_1 \times$ (sensor signal)
E*	FEM = $\beta_0 + \beta_1 \times$ (sensor signal)
F*	FEM = $\beta_0 + \beta_1 \times$ (sensor signal) $+ \beta_2 \times RH$
G*	FEM = $\beta_0 + \beta_1 \times$ (sensor signal) $+ \beta_2 \times RH + \beta_3 \times T_{dew}$

A note: There are numerous options for calibration, but this will not be the focus of the tutorial. The primary goal is to illustrate how model performance can be assessed. As such, do not consider these models as your only option just some example models to look at performance metrics using a Taylor Diagram (Fig. 7).

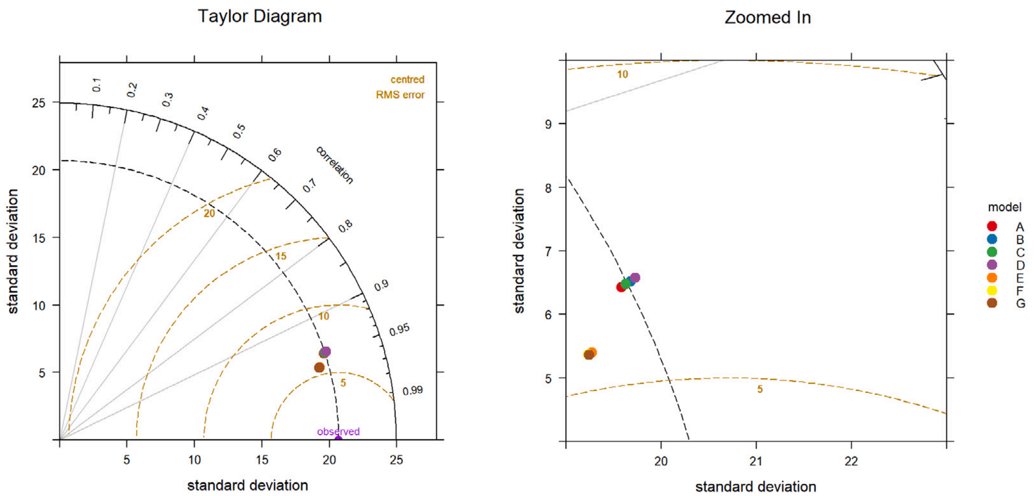

Fig. 7. Taylor diagram comparing all of our different candidate calibration models. The light gray lines correspond to the Pearson R correlation coefficient, the radial gold lines represent the centered RMSE, and the black dashed line represents the standard deviation of the observations; here all models are to the left of the black dashed line indicating smaller standard deviation than the reference measurements. To help differentiate the models, a zoomed in version is provided on the right panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as was expected based on the our analysis in Section 4.5. Looking at the tabular data (see: web tutorial), we determined that the optimal model was model F, since it had the highest R^2 and lowest standard error. While Model G had the same performance metrics as Model F (they overlap on Fig. 7), the parsimonious choice is Model F, since adding dew point did not improve the model.

5. Conclusions

This guideline tutorial was intended to serve two purposes: (1) to distill some common considerations when deploying low-cost PM sensors, ranging from known sensor influences to network logistics, into one concise document, and (2) to provide an open-access tutorial demonstration how one might assess $PM_{2.5}$ sensor performance using the U.S. EPA "Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors" guidance. Note that this demonstration performance evaluation was conducted only for a 4-month period, and that for ongoing deployments, routine sensor performance and calibration model evaluations should be conducted (ideally with changing environmental conditions, such as seasonal assessments). These routine assessments will help address sensor drift, as all low-cost sensors have a limited service life.

This paper is not a formal literature review, and as such it is strongly recommended that readers and sensor users continue to seek out other published documents when reporting on low-cost PM sensor data. The body of literature in this field has grown exponentially over the past 5 years and it would have been impossible to cite every relevant paper in this document, nor was that the goal.

In general, it is my view that with careful planning, transparent assessments of sensor accuracy, thoughtfully-developed calibration models, and open-access data reporting structures, that low-cost PM sensors can be a critical and important tool for improving air quality, engaging with citizen scientists, and promoting environmental justice. It is my hope that this tutorial will provide some inspiration on the means of achieving these goals.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program. The companion web tutorial was constructed entirely in the open-source coding language R, and I wish to acknowledge the amazing community of practitioners who have developed and maintained the R packages that made this analysis possible (full citations to R packages used found in the web tutorial). I also wish to acknowledge the hard-working students in my lab, whose questions and interest in low-cost sensors inspired this tutorial.

About this article

This article is an Editor-Invited Tutorial Article. Tutorial articles, established to commemorate the 50th Anniversary of the Journal of Aerosol Science in 2020, are intended to serve as educational resources for the aerosol research community on state-of-the-art experimental, theoretical, and numerical techniques in aerosol science.

References

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C. H., & Hamburg, S. P. (2017). High-resolution air pollution mapping with google street view cars: Exploiting big data. *Environmental Science and Technology*, *51*(12), 6999–7008. http://dx.doi.org/10.1021/acs.est.7b00891.

Austin, E., Novosselov, I., Seto, E., & Yost, M. G. (2015). Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor. *PLOS ONE*, *10*(9), Article e0137789, URL: https://doi.org/10.1371/journal.pone.0137789.

Boghozian, A. J. (2021). *An exercise in selecting low-cost air quality sensor placements within an urban environment* (Ph.D. thesis), Massachusetts Institute of Technology, URL: http://dspace.mit.edu/handle/1721.1/7582.

Borrego, C., Ginja, J., Coutinho, M., Ribeiro, C., Karatzas, K., Sioumis, T., Katsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Salvato, M., Smith, P., André, N., Gérard, P., Francis, L. A., Castell, N., Schneider, P., Viana, M., Minguillón, M. C., ... Penza, M. (2018). Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise – Part II. *Atmospheric Enviroment*, *193*, 127–142. http://dx.doi.org/10.1016/j.atmosenv.2018.08.028, URL: https://www.sciencedirect.com/science/article/pii/S1352231018305430.

Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Whitsel, L., & Kaufman, J. D. (2010). Particulate matter air pollution and cardiovascular disease. *Circulation*, *121*(21), 2331 LP–2378.

Bulot, F. M. J., Johnston, S. J., Basford, P. J., Easton, N. H. C., Apetroaie-Cristea, M., Foster, G. L., Morris, A. K. R., Cox, S. J., & Loxham, M. (2019). Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Scientific Reports*, *9*(1), 7497. http://dx.doi.org/10.1038/s41598-019-43716-3.

Callahan, J., Martin, H., Wilson, K., Brasel, T., & Miller, H. (2020). AirSensor: Process and display data from air quality sensors. URL: https://github.com/MazamaScience/AirSensor.

Campbell, J. L., Rustad, L. E., Porter, J. H., Taylor, J. R., Dereszynski, E. W., Shanley, J. B., Gries, C., Henshaw, D. L., Martin, M. E., Sheldon, W. M., & Boose, E. R. (2013). Quantity is nothing without quality: Automated QA/QC for streaming environmental sensor data. *BioScience*, *63*(7), 574–585. http://dx.doi.org/10.1525/bio.2013.63.7.10, arXiv:https://academic.oup.com/bioscience/article-pdf/63/7/574/19406652/63-7-574.pdf.

CANUE (2021). CANUE data portal. URL: https://canue.ca/data/.

Cavaliere, A., Carotenuto, F., Di Gennaro, F., Gioli, B., Gualtieri, G., Martelli, F., Matese, A., Toscano, P., Vagnoli, C., & Zaldei, A. (2018). Development of low-cost air quality stations for next generation monitoring networks: Calibration and validation of PM2.5 and PM10 sensors. http://dx.doi.org/10.3390/s18092843.

Chow, J. C., & Watson, J. G. (2008). New directions: Beyond compliance air quality measurements. *Atmospheric Enviroment*, *42*(20), 5166–5168. http://dx.doi.org/10.1016/j.atmosenv.2008.05.004, URL: https://www.sciencedirect.com/science/article/pii/S1352231008003889.

Crilley, L. R., Shaw, M., Pound, R., Kramer, L. J., Price, R., Young, S., Lewis, A. C., & Pope, F. D. (2018). Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmospheric Measurement Techniques*, *11*(2), 709–720. http://dx.doi.org/10.5194/amt-11-709-2018, URL: https://amt.copernicus.org/articles/11/709/2018/.

Di Antonio, A., Popoola, O. A. M., Ouyang, B., Saffell, J., & Jones, R. L. (2018). Developing a relative humidity correction for low-cost sensors measuring ambient particulate matter. http://dx.doi.org/10.3390/s18092790.

Duvall, R., Clements, A., Hagler, G., Kamal, A., Kilaru, V. J., Goodman, L., Frederick, S., Jonhson Barkjon, K., VonWold, I., Greene, D., & Dye, T. (2021). *Performance testing protocols, metrics, and target values for fine particulate matter air sensors: Use in ambient, outdoor, fixed site, non-regulatory supplemental and informational monitoring applications*: *Technical Report*, Washington, D.C.: U.S. EPA Office of Research and Development, URL: https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=350785&Lab=CEMM.

Dye, T. (2017). A brief history of air quality sensors. URL: https://www.sutori.com/story/a-brief-history-of-air-quality-sensors--jMm2LKEQZDWmXvsDMv2MP5LV%0A.

Environmental Reporting BC (2021). *Status of fine particulate matter in B.C. (2016-2018)*. *State of environment reporting*: *Technical Report*, (p. 16). British Columbia, Canada: Ministry of Environment and Climate Change Strategy, URL: https://www.env.gov.bc.ca/soe/indicators/air/fine-pm.html.

Feenstra, B., Collier-Oxandale, A., Papapostolou, V., Cocker, D., & Polidori, A. (2020). The AirSensor open-source R-package and DataViewer web application for interpreting community data collected by low-cost sensor networks. *Environmental Modelling & Software*, *134*, Article 104832. http://dx.doi.org/10.1016/j.envsoft.2020.104832, URL: https://www.sciencedirect.com/science/article/pii/S1364815220308896.

Feenstra, B., Papapostolou, V., Hasheminassab, S., Zhang, H., Boghossian, B. D., Cocker, D., & Polidori, A. (2019). Performance evaluation of twelve low-cost PM2.5 sensors at an ambient air monitoring site. *Atmospheric Enviroment*, *216*, Article 116946. http://dx.doi.org/10.1016/j.atmosenv.2019.116946, URL: https://www.sciencedirect.com/science/article/pii/S1352231019305850.

GeoBC (2021). Meteorological monitoring sites. URL: https://catalogue.data.gov.bc.ca/dataset/meteorological-monitoring-sites.

Giordano, M. R., Malings, C., Pandis, S. N., Presto, A. A., McNeill, V., Westervelt, D. M., Beekmann, M., & Subramanian, R. (2021). From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors. *Journal of Aerosol Science*, Article 105833. http://dx.doi.org/10.1016/j.jaerosci.2021.105833, URL: https://www.sciencedirect.com/science/article/pii/S0021850221005644.

Hagan, D. H., & Kroll, J. H. (2020). Assessing the accuracy of low-cost optical particle sensors using a physics-based approach. *Atmospheric Measurement Techniques*, *13*(11), 6343–6355. http://dx.doi.org/10.5194/amt-13-6343-2020, URL: https://amt.copernicus.org/articles/13/6343/2020/.

Han, Y., Park, B., & Jeong, J. (2019). A novel architecture of air pollution measurement platform using 5G and blockchain for industrial IoT applications. *Procedia Computer Science*, *155*, 728–733. http://dx.doi.org/10.1016/j.procs.2019.08.105, URL: https://www.sciencedirect.com/science/article/pii/S187705091931021X.

Hapidin, D. A., Munir, M. M., Suprijadi, & Khairurrijal, K. (2020). Development of a new personal air filter test system using a low-cost particulate matter (PM) sensor. *Aerosol Science and Technology*, *54*(2), 203–216. http://dx.doi.org/10.1080/02786826.2019.1678734.

Hinds, W. C. (1999). *Aerosol technology: Properties, behavior, and measurement of airborne particles* (2nd ed.). Wiley.

Holder, A. L., Mebust, A. K., Maghran, L. A., McGown, M. R., Stewart, K. E., Vallano, D. M., Elleman, R. A., & Baker, K. R. (2020). Field evaluation of low-cost particulate matter sensors for measuring wildfire smoke. http://dx.doi.org/10.3390/s20174796.

Ito, K., Johnson, S., Kheirbek, I., Clougherty, J., Pezeshki, G., Ross, Z., Eisl, H., & Matte, T. D. (2016). Intraurban variation of fine particle elemental concentrations in new york city. *Environmental Science and Technology*, *50*(14), 7517–7526. http://dx.doi.org/10.1021/acs.est.6b00599.

Jain, S., Presto, A. A., & Zimmerman, N. (2021). Spatial modeling of daily PM2.5, NO2, and CO concentrations measured by a low-cost sensor network: Comparison of linear, machine learning, and hybrid land use models. *Environmental Science and Technology*, http://dx.doi.org/10.1021/acs.est.1c02653.

Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., & Morawska, L. (2018). The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, *11*(8), 4883–4890. http://dx.doi.org/10.5194/amt-11-4883-2018, URL: https://amt.copernicus.org/articles/11/4883/2018/.

Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A. M., Helber, R., & Arnone, R. A. (2009). Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, *76*(1–2), 64–82. http://dx.doi.org/10.1016/j.jmarsys.2008.05.014.

Koehler, K. A., & Peters, T. M. (2015). New methods for personal exposure monitoring for airborne particles. *Current Environmental Health Reports*, *2*(4), 399–411. http://dx.doi.org/10.1007/s40572-015-0070-z.

Krause, A., Singh, A., & Guestrin, C. (2008). Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, *9*(8), 235–284, URL: http://jmlr.org/papers/v9/krause08a.html.

Kumar, V., & Sahu, M. (2021). Evaluation of nine machine learning regression algorithms for calibration of low-cost PM2.5 sensor. *Journal of Aerosol Science*, *157*, Article 105809. http://dx.doi.org/10.1016/j.jaerosci.2021.105809, URL: https://www.sciencedirect.com/science/article/pii/S0021850221005401.

Li, J., Mattewal, S. K., Patel, S., & Biswas, P. (2020). Evaluation of nine low-cost-sensor-based particulate matter monitors. *Aerosol and Air Quality Research*, *20*(2), 254–270. http://dx.doi.org/10.4209/aaqr.2018.12.0485.

Maag, B., Zhou, Z., & Thiele, L. (2018). A survey on sensor calibration in air pollution monitoring deployments. *IEEE Internet of Things Journal*, *5*(6), 4857–4870. http://dx.doi.org/10.1109/JIOT.2018.2853660.

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., & Subramanian, R. (2020). Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation. *Aerosol Science and Technology*, *54*(2), 160–174. http://dx.doi.org/10.1080/02786826.2019.1623863.

Metro Vancouver (2016). Open data catalogue. URL: http://www.metrovancouver.org/data.

Metro Vancouver (2021). AirMap. URL: https://gis.metrovancouver.org/maps/Air.

Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai, F., Christensen, B., Dunbabin, M., Gao, J., Hagler, G. S., Jayaratne, R., Kumar, P., Lau, A. K., Louie, P. K., Mazaheri, M., Ning, Z., Motta, N., ... Williams, R. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?. http://dx.doi.org/10.1016/j.envint.2018.04.018.

Motlagh, N. H., Lagerspetz, E., Nurmi, P., Li, X., Varjonen, S., Mineraud, J., Siekkinen, M., Rebeiro-Hargrave, A., Hussein, T., Petaja, T., Kulmala, M., & Tarkoma, S. (2020). Toward massive scale air quality monitoring. *IEEE Communications Magazine*, *58*(2), 54–59. http://dx.doi.org/10.1109/MCOM.001.1900515.

OpenAQ (2019). Openaq-data-format. URL: https://github.com/openaq/openaq-data-format.

OpenAQ (2020). OpenAQ. URL: https://openaq.org/#/map.

Papapostolou, V., Zhang, H., Feenstra, B. J., & Polidori, A. (2017). Development of an environmental chamber for evaluating the performance of low-cost air quality sensors under controlled conditions. *Atmospheric Enviroment*, *171*, 82–90. http://dx.doi.org/10.1016/j.atmosenv.2017.10.003.

Patra, S. S., Ramsisaria, R., Du, R., Wu, T., & Boor, B. E. (2021). A machine learning field calibration method for improving the performance of low-cost particle sensors. *Building and Environment*, *190*, Article 107457. http://dx.doi.org/10.1016/j.buildenv.2020.107457, URL: https://www.sciencedirect.com/science/article/pii/S0360132320308246.

Petters, M. D., & Kreidenweis, S. M. (2007). A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. *Atmospheric Chemistry and Physics*, *7*(20), 6273–6279. http://dx.doi.org/10.5194/acp-8-6273-2008.

Pope, C. A., Ezzati, M., & Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the united states. *New England Journal of Medicine*, *360*(4), 376–386. http://dx.doi.org/10.1056/NEJMsa0805646.

PurpleAir (2021). No title. URL: https://www.purpleair.com/map.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, URL: https://www.r-project.org/.

Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M. J., Brunekreef, B., Xun, W. W., Katsouyanni, K., Dimakopoulou, K., Sommar, J., Forsberg, B., Modig, L., Oudin, A., Oftedal, B., Schwarze, P. E., ... Hoek, G. (2013). Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *The Lancet Oncology*, *14*(9), 813–822. http://dx.doi.org/10.1016/S1470-2045(13)70279-1.

RStudio Team (2020). *RStudio: Integrated development environment for R*. Boston, MA: RStudio, PBC, URL: http://www.rstudio.com/.

Sayahi, T., Butterfield, A., & Kelly, K. E. (2019). Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environmental Pollution*, *245*, 932–940. http://dx.doi.org/10.1016/j.envpol.2018.11.065, URL: https://www.sciencedirect.com/science/article/pii/S0269749118316129.

South Coast AQMD (2021). Submit a sensor. URL: https://www.aqmd.gov/aq-spec/questions.

Tan, Y., Lipsky, E. M., Saleh, R., Robinson, A. L., & Presto, A. A. (2014). Characterizing the spatial variation of air pollutants and the contributions of high emitting vehicles in Pittsburgh, PA. *Environmental Science and Technology*, *48*(24), 14186–14194. http://dx.doi.org/10.1021/es5034074.

U. S. EPA (2021a). Air monitoring methods - Criteria pollutants. URL: https://www.epa.gov/amtic/air-monitoring-methods-criteria-pollutants.

U. S. EPA (2021b). Air sensors toolbox: Evaluation of emerging air sensor performance. URL: https://www.epa.gov/air-sensor-toolbox/evaluation-emerging-air-sensor-performance.

Wang, Z., Delp, W. W., & Singer, B. C. (2020). Performance of low-cost indoor air quality monitors for PM2.5 and PM10 from residential sources. *Building and Environment*, *171*, Article 106654. http://dx.doi.org/10.1016/j.buildenv.2020.106654, URL: https://www.sciencedirect.com/science/article/pii/S0360132320300123.

Wang, Y., Du, Y., Wang, J., & Li, T. (2019). Calibration of a low-cost PM2.5 monitor using a random forest model. *Environment International*, *133*, Article 105161. http://dx.doi.org/10.1016/j.envint.2019.105161, URL: https://www.sciencedirect.com/science/article/pii/S0160412019322780.

Wang, J.-F., Hu, M.-G., Xu, C.-D., Christakos, G., & Zhao, Y. (2013). Estimation of citywide air pollution in Beijing. *PLOS ONE*, *8*(1), Article e53400, URL: https://doi.org/10.1371/journal.pone.0053400.

Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., & Biswas, P. (2015). Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement. *Aerosol Science and Technology*, *49*(11), 1063–1077. http://dx.doi.org/10.1080/02786826.2015.1100710.

Wickham, H. (2019). Tidyverse: Easily install and load the tidyverse. URL: https://cran.r-project.org/package=tidyverse.

Wiktorski, T., & Królak, A. (2020). Extended approach to sum of absolute differences method for improved identification of periods in biomedical time series. *MethodsX*, *7*, Article 101094. http://dx.doi.org/10.1016/j.mex.2020.101094, URL: https://www.sciencedirect.com/science/article/pii/S2215016120303149.

Zheng, T., Bergin, M. H., Johnson, K. K., Tripathi, S. N., Shirodkar, S., Landis, M. S., Sutaria, R., & Carlson, D. E. (2018). Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments. *Atmospheric Measurement Techniques*, *11*(8), 4823–4846. http://dx.doi.org/10.5194/amt-11-4823-2018, URL: https://amt.copernicus.org/articles/11/4823/2018/.

Zimmerman, N., Li, H., Ellis, A., Hauryliuk, A., Robinson, E., Gu, P., Shah, R., Ye, Q., Snell, L., Subramanian, R., Robinson, A., Apte, J., & Presto, A. (2020). Improving correlations between land use and air pollutant concentrations using wavelet analysis: Insights from a low-cost sensor network. *Aerosol and Air Quality Research*, *20*(2), http://dx.doi.org/10.4209/aaqr.2019.03.0124.

Naomi Zimmerman is an assistant professor in the Department of Mechanical Engineering at the University of British Columbia (UBC) and Canada Research Chair in Sustainability. Prior to UBC, she was a postdoctoral fellow at Carnegie Mellon University's Center for Atmospheric Particle Studies and also received a PhD in Chemical Engineering from the University of Toronto. Her research interests are in the application of experimental air pollution monitoring systems (low-cost sensor networks, mobile monitoring) to assess the air and climate impacts of technology or policy interventions, with a focus on transportation and energy systems.