Journal of Aerosol Science 158 (2021) 105833

Contents lists available at ScienceDirect

Journal of Aerosol Science

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

From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors

Michael R. Giordano [a,b,**], Carl Malings [c], Spyros N. Pandis [d,e], Albert A. Presto [f], V. F. McNeill [g], Daniel M. Westervelt [h,i], Matthias Beekmann [a,j], R. Subramanian [a,b,f,k,1,*]

[a] *Univ Paris Est Creteil, CNRS UMS 3563, Ecole Nationale des Ponts et Chaussés, Université de Paris, OSU-EFLUVE – Observatoire Sciences de L'Univers-Envelopes Fluides de La Ville à L'Exobiologie, F-94010 Créteil, France*
[b] *Univ Paris Est Creteil and Université de Paris, CNRS, LISA, F-94010 Créteil, France*
[c] *NASA Postdoctoral Program Fellow, Goddard Space Flight Center, Greenbelt, MD, USA*
[d] *Department of Chemical Engineering, University of Patras, Patras, Greece*
[e] *Institute of Chemical Engineering Sciences, FORTH, Patras, Greece*
[f] *Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA*
[g] *Departments of Chemical Engineering and Earth and Environmental Sciences, Columbia University, New York, NY, USA*
[h] *Lamont-Doherty Earth Observatory of Columbia University, New York, NY, USA*
[i] *NASA Goddard Institute for Space Studies, New York, NY, USA*
[j] *Université de Paris and Univ Paris Est Creteil, CNRS, LISA, F-75013 Paris, France*
[k] *Kigali Collaborative Research Centre, Kigali, Rwanda*

A R T I C L E I N F O

Keywords:
Low-cost sensors
Particulate matter
Air quality
Air pollution

A B S T R A C T

Low-cost sensors for particulate matter mass (PM) enable spatially dense, high temporal resolution measurements of air quality that traditional reference monitoring cannot. Low-cost PM sensors are especially beneficial in low and middle-income countries where few, if any, reference grade measurements exist and in areas where the concentration fields of air pollutants have significant spatial gradients. Unfortunately, low-cost PM sensors also come with a number of challenges that must be addressed if their data products are to be used for anything more than a qualitative characterization of air quality. The various PM sensors used in low-cost monitors are all subject to biases and calibration dependencies, corrections for which range from relatively straightforward (e.g. meteorology, age of sensor) to complex (e.g. aerosol source, composition, refractive index). The methods for correcting and calibrating these biases and dependencies that have been used in the literature likewise range from simple linear and quadratic models to complex machine learning algorithms. Here we review the needs and challenges when trying to get high-quality data from low-cost sensors. We also present a set of best practices to follow to obtain high-quality data from these low-cost sensors.

* Corresponding author. Univ Paris Est Creteil, CNRS UMS 3563, Ecole Nationale des Ponts et Chaussés, Université de Paris, OSU-EFLUVE – Observatoire Sciences de L'Univers-Envelopes Fluides de La Ville à L'Exobiologie, F-94010 Créteil, France.
** Corresponding author. Univ Paris Est Creteil, CNRS UMS 3563, Ecole Nationale des Ponts et Chaussés, Université de Paris, OSU-EFLUVE – Observatoire Sciences de L'Univers-Envelopes Fluides de La Ville à L'Exobiologie, F-94010 Créteil, France.
E-mail addresses: mike@afriqair.org (M.R. Giordano), subu@lisa.ipsl.fr, subu@cmu.edu (R. Subramanian).
[1] Currently at Qatar Environment and Energy Research Institute, Hamad Bin Khalifa University, Doha, Qatar.

https://doi.org/10.1016/j.jaerosci.2021.105833
Received 15 January 2021; Received in revised form 24 June 2021; Accepted 26 June 2021
Available online 2 July 2021

1. Introduction

Low-cost sensors have the potential to greatly alter how, where, and when air pollution monitoring is done. Until 2010or so, measurement of particulate matter mass (PM) concentrations could be routinely performed only by entities that could afford instruments on the order of USD 10,000's (generally governments or research organizations in the Global North), with overall station and personnel costs easily an order of magnitude higher. The past decade has seen radical changes in how air quality monitoring can be performed, who can perform it, and the temporal and spatial scales on which we think about air quality in general. These changes have been driven by the emergence of low-cost air quality sensors, including low-cost PM monitors. Low-cost sensors (LCS) have been utilized in (or marketed towards) projects of various scales: from citizen science projects and personal empowerment (e.g. the Airbeam monitor [https://www.habitatmap.org/airbeam], Plume Labs Flow monitor [https://plumelabs.com]), PurpleAir [https://www2.purpleair.com/]) (Ford et al., 2019; Karagulian et al., 2019; Kelly et al., 2017) to large city-scale distributed network deployments (Eilenberg et al., 2020; Jiao et al., 2016). These projects may also include more sparsely dispersed reference monitors to increase the network's accuracy. Fig. 1 shows how the incidence of "low-cost sensor" and "air quality"/"air pollution" keywords in the scientific literature has increased since 2002. With the development of low-cost sensors, the general public, researchers, and even governments in underserved regions can now attain high time resolution PM concentration measurements from a high spatial density network.

The biggest challenge to the widespread use of low-cost sensors is their out-of-the-box data quality is generally low. Paradoxically, despite the easy access to data that LCS provide, interpreting that data requires more than a cursory understanding of the sensors themselves and their outputs if one wants to compare LCS data to reference monitor data. A poor understanding of these sensors can lead to undesirable outcomes. Users from the general public could be alarmed that their measurements are much higher than that from their government's reference monitors, but this could have a simple technical explanation, such as aerosol growth at high relative humidity since government data are reported at low to intermediate RH (35% by US and 50% by EU standards). Nations with low reference-monitor coverage could conclude based on LCS data that their ambient air quality is much better or much worse than reality, leading to ineffective air quality management policies. Where the responsibility lies in training end-users to obtain high quality data from low-cost sensors is an open question.

The main benefit of low-cost PM sensors is apparent from the name; these instruments are relatively inexpensive. Most PM LCS cost on the order of US$10 to 100 for a single sensor, though the associated electronics and packaging can push costs to anywhere from $200 to $2000 – still a small fraction of the cost of traditional reference PM monitors (Table 1). Large-scale, dense networks are therefore possible, even in historically underserved nations. Some work has been done to evaluate makes and models of various low-cost sensors using formalized evaluation criteria (Fishbain et al., 2017; Williams et al., 2014) which has in general concluded that the as-reported data are often of low quality, suffering from various artifacts and biases. Attaining high-quality data from these low-cost sensors requires extra work. Here, we review and discuss these recent studies to better understand the limitations of low-cost PM sensors and methods used to address these limitations. Finally, we provide a set of suggestions supported by the literature for how to generate high-quality data from low-cost sensors.

This manuscript is also based on the experiences and lessons learned by the authors' deployments of over 200 low-cost PM sensors over much of the world for the past five years, including the USA, South Asia, and multiple countries in Europe and Africa.

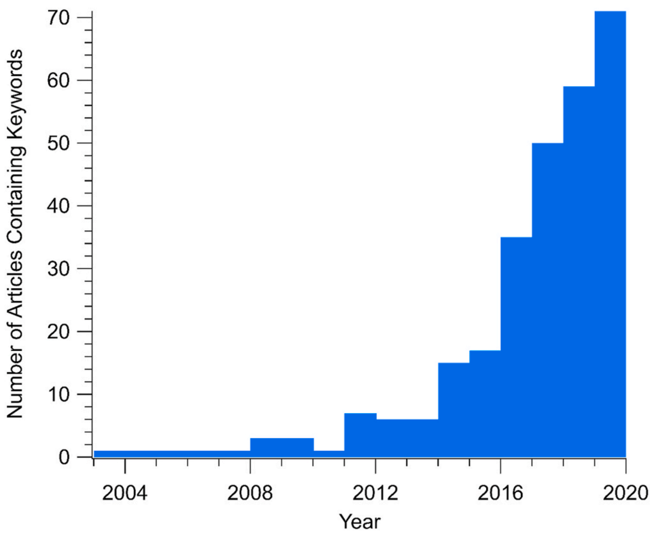

Fig. 1. Number of articles containing the keywords "low cost sensor" and either "air pollution" or "air qualitiy" on Pubmed.gov.

Table 1
General operating principles and price of some commonly used, commercially available low-cost PM sensors.

Make	Model	Light wavelength[a]	Scattering Angle (°)[a]	Aspiration	Stated Lower Particle Size Sensitivity[a]	Output	Price (USD)
[b]Shinyei	PPD42NS	Infrared	120	Convective	>1 µm	Lo Pulse Occupancy (%)	~20
[b]Sharp	GP2Y1010AU0F	Infrared	120	None	0.5 µm	Lo Pulse Occupancy (%)	~20
[b]Samyoung	DSM501	Infrared	120	None	0.5 µm	Lo Pulse Occupancy (%)	~20
Nova	SDS 011	"Red Laser"	90	Fan	0.3 µm	Particle Mass concetrations ($PM_{2.5}$, PM_{10})	~20
[d]Dylos	DC1700	"Laser"	[d]Laser particle counter	Fan (? L/min)	<0.5 µm and >2.5 µm	Particle number concentrations	~450
Plantower	PMS1003/3003/5003	650 nm	90	[c]Fan (? L/min)	0.3 µm	Particle Mass concetrations	20+
Sensirion	SPS30	660 nm	90	Fan (? L/min)	0.3 µm	Particle Mass concetrations (PM_1, $PM_{2.5}$, PM_4, PM_{10})	~40
Honeywell	HPMA115C0	650 nm	Forward	Fan (? L/min)	Unstated	Particle Mass concetrations (PM_1, $PM_{2.5}$, PM_4, PM_{10})	~40
[d]Met One	NPM2	670 nm	Forward	2L/min	0.1 µm	Particle Mass concentration	~2000
Alphasense	OPC-N2/OPC-N3	658 nm	30	5.5L/min	0.35 µm	Particle number concentrations	~500

[a] Wavelengths, scattering angle, and particle size sensitivities are taken from manufacturer specifications where available, or are not specified here. Though some manufacturers specify lower size detection limits, this may not be the case in practice, see He et al., 2020 and Tryner et al., 2020a.
[b] These sensors are essentially the same but are packaged differently by different companies.
[c] The PMS 1003/5003 flowrate is estimated to be approximately 0.1 lpm by Sayahi et al. (2019).
[d] Sensor is an integrated sensor; all other sensors are stand alone.

2. Sensor design

All commercially available low-cost PM sensors utilize light-scattering as their principle of operation. Light scattering can be implemented in a much smaller form factor than most other methods of particle counting/mass concentration measurements. In principle, a low-cost PM sensor only requires three main components: a light (generally infrared or red laser) emitting diode, a phototransistor, and a lens to focus the diode light (Wang et al., 2015a). A fan, pump, or convective heater can be used to draw air into the measurement cavity. Some commonly used commercially available low-cost sensors and their general operating principles are

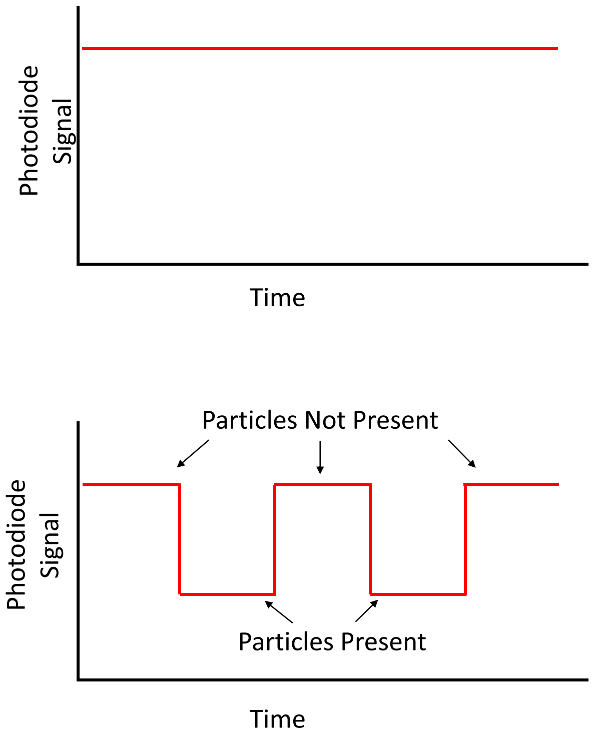

Fig. 2. Representation of the resulting output for the photodiode signal in a low-cost PM sensor using LPO with particles absent (a) and present (b) in the light path.

Table 2
Summary of findings, sensors studied, and broad categorization of sensor effects from numerous low-cost sensor studies specifically relating to confounding factors for low-cost sensor measurements.

Study	Sensor(s) Used	Topics										Major finding(s)
		Aerosol Concentration	Aerosol Size, Density, Composition	Aerosol Size Distribution	RH Effects	Optical property effects	Reference method selection	Calibration Methods	On-board Corrections	Intersensor Consistency	Averaging Periods	
Ardon-Dryer et al., 2020	Purpleair PA-II (Plantower PMS 5003)									X		Dual monitor setups like Purpeair's are often well correlated with eachother (R2 > 0.9)
Barkjohn et al., 2020	Plantower PMS5003	X			X		X					In Beijing, China, CF = atm showed higher correlation than CF = 1 with reference methods (R2 = 0.63 vs 0.52). LCS may be useful in determining indoor/outdoor filtration changes.
Castell et al. (2017)	AQMesh		X						X			Lower traffic emissions correlate to better agreement with reference monitors. Onboard calibration factors likely the cause.
Crilley et al. (2018)	Alphasense OPC-N2				X		X			X		Gravimetric and optical reference methods affect end calibration. RH > 85% significantly impacts sensors. ~20% variability in inter-unit precision for 14 sensors.
Demanega et al. (2021)	AirVisual, Awair, Clarity, Foobot, Kaiterra, uHoo	X	X	X	X							For two different humidities and multiple different sources, the tested sensors underreport PM concentrations by up to 50%
Di Antonio et al. (2018)	AlphaSense OPC-N2		X	X	X							κ-Köhler theory based RH correction significantly reduces overestimation of PM1 and PM2.5 by at least an order of magnitude
		X						X				

(*continued on next page*)

Table 2 (*continued*)

Study	Sensor(s) Used	Topics										Major finding(s)
		Aerosol Concentration	Aerosol Size, Density, Composition	Aerosol Size Distribution	RH Effects	Optical property effects	Reference method selection	Calibration Methods	On-board Corrections	Intersensor Consistency	Averaging Periods	
Gao et al. (2015)	PUWP (Shinyei PPD42NS)											High PM2.5 concentrations (above 100 μg m-3) require non-linear correction methods.
Hagan and Kroll (2020)	Theoretical, Idealized Sensors			X	X	X						Sensor technology (OPC vs nephelometer), sensor design, and calibration aerosol choices all impact the relative error of sensors when calibrating. Mie Theory can be used to reduce some of this error.
Han et al. (2017)	Dylos DC1700	X	X		X							PM2.5 and course PM (PM10-PM2.5) associations dependent on size distribution. RH > 60% causes overestimation (>70%) of mass concentrations.
He et al. (2020)	Plantower PMS5003	X		X		X						Shape of response function of LCS is closely related to light-scattering response (a complex function of composition and size distribution)
Holder et al. (2020)	Plantower PMS5003 (in RAMPS and PurpleAirs), Aeroqual micro AQ station	X	X									A specific forest-fire correction can be made to reduce the impact of high concentration BMB events and reduce MAE < 10 μg m-3
Jiao et al. (2016)	Shinyei PMS-SYS-1, Dylos DC1100, Shinyei PPD60PV, Metone Aercet 831, Shinyei PPD42				X			X				Including RH in sensor caibrations yielded higher regression coefficients (R2adj)

(*continued on next page*)

Table 2 (*continued*)

Study	Sensor(s) Used	Topics										Major finding(s)
		Aerosol Concentration	Aerosol Size, Density, Composition	Aerosol Size Distribution	RH Effects	Optical property effects	Reference method selection	Calibration Methods	On-board Corrections	Intersensor Consistency	Averaging Periods	
												for at least both the Shinyei and Airbeam sensors than including the temperature (the Dylos being the exception). Including both RH and T did not yield significantly higher correlations
Kelly et al. (2017)	Plantower PMS 1003/3003	X					X					Sensor response linear between 200 and 850 µg m-3 PM2.5 in lab tests but ambient tests can show overestimation at certain periods. High correlation (> 0.8) to light-scattering, attenuation, and gravimetric reference methods.
Kuula et al., 2020	Plantower PMS5003, Nova SDS011, Sensirion SPS30, Sharp GP2Y1010AU0F, Shinyei PPD42NS, and Omron B5W-LD0101			X								None of the tested sensors adhere to the detection ranges declared by the manufacturer and only achieve independent responses in 1 or 2 size bins
Li et al. (2020)	AirVisual, Alphasense, APT, Awair, Dylos, Foobot, PurpleAir, Wynd, and Xiaomi	X	X	X			X					Laboratory chamber tests of 8 of 9 sensors using 3 different aerosol sources generally showed high correlation (R2 > 0.9) and high linearity when compared to two different reference monitors
Liu et al. (2017)	4 Sharp, Shinyei, Samyoung and Oneair sensors		X									Linear correlations are appropriate under tested conditions but change with differing

(*continued on next page*)

Table 2 (*continued*)

Study	Sensor(s) Used	Topics										Major finding(s)
		Aerosol Concentration	Aerosol Size, Density, Composition	Aerosol Size Distribution	RH Effects	Optical property effects	Reference method selection	Calibration Methods	On-board Corrections	Intersensor Consistency	Averaging Periods	
												size distribution means and geometrical standard deviations.
Liu et al. (2019)	Nova SDS011	X			X					X		Sensors highly correlate with each other (R2 > 0.97). Between 0 and 50 μg m-3 PM2.5 sensors reasonably correlate with reference (R2 ~ 0.5). RH > 80% negatively effects sensors.
Magi et al. (2020)	Plantower PMS5003 (PurpleAir PA–II–SD)				X			X				Including both RH and T in calibration models yieded a 27–57% improvement in sensor accuracy but the improvement is more pronounced for higher RH than higher T
Malings et al. (2020)	Met-One NPM and PurpleAir PA-II				X			X		X		25 NPM and 9 PPA sensors show high inter-unit consistency for PM2.5 (r > 0.9, MAE < 2.5 μg m-3). Fully-empirical correction methods for RH/particle hygroscopicity are as good as or better than semi-theoretical approaches.
Masic et al. (2020)	Alphasense OPC-N2, Plantower PMS5003	X										Daily average aerosol concentrations at ~20 μg m-3 PM2.5 and PM10 result in worse agreement with reference monitors than daily averages of 83 μg m-3.
	Plantower PMS5003	X										MLR and RF calibration methods

(*continued on next page*)

Table 2 (*continued*)

Study	Sensor(s) Used	Topics										Major finding(s)
		Aerosol Concentration	Aerosol Size, Density, Composition	Aerosol Size Distribution	RH Effects	Optical property effects	Reference method selection	Calibration Methods	On-board Corrections	Intersensor Consistency	Averaging Periods	
McFarlane et al. (2021)												work well for collocations in extremely polluted environments such as Kampala, Uganda
Pope, Gatari, Ng'ang'a, Poynter, & Blake (2018)	AlphaSense OPC-N2									X		After collocation calibration, inter-instrument coefficient of variance of 8.8% for PM2.5.
Eilenberg et al., 2020	Plantower PMS5003										X	Instrument and sampling uncertainty percentages decrease as averaging times are increased as compared to reference instruments.
Salimifard et al. (2020)	OPC-N2, IC Sentinel, Speck, and Dylos	X	X									Linearity of sensors sensitive to particle concentration: number concentrations < 5 cm-3 are strongly non-linear. Responses to different aerosol compositions differ in strength and location of non-linearities.
Sousan et al. (2016)	Dylos DC1700, Sharp GP2Y1010AU0F and DN7C3CA006		X									Of 4 different sources of aerosols, sensors show differing bias and detection efficiencies but calibration can correct for these effects.
Subramanian et al. (2020)	MetOne NPM	X										Pittsburgh-based sensor calibration applied to Kigali, Rwanda measurements appears to significantly underestimate gravimetric PM2.5

(*continued on next page*)

Table 2 (*continued*)

Study	Sensor(s) Used	Topics										Major finding(s)
		Aerosol Concentration	Aerosol Size, Density, Composition	Aerosol Size Distribution	RH Effects	Optical property effects	Reference method selection	Calibration Methods	On-board Corrections	Intersensor Consistency	Averaging Periods	
Tryner et al. (2019b)	Custom						X					Gravimetric methods as a reference method for low-cost sensors reduces bias and relative standard deviation.
Tryner et al. (2020a)	Plantower PMS5003, Amphenol Advanced Sensors SM-UART-04 L, Sensirion SPS30	X	X		X		X					Particle count data reported by the PMS5003 not reliable for various types of measured PM2.5 aerosols. Some sensors subject to errors caused by exposure to high dust concentrations.
Tryner et al. (2020b)	Plantower PMS5003 (PurpleAir PA–II–SD)				X		X					Performing collocation calibrations only in laboratory settings does not translate to field settings. Periodic corrections with gravimetric filters reduces bias.
Wang et al., 2015	Shinyei PPD42NS, Samyoung DSM501A, and Sharp GP2Y1010AU0F	X	X		X							All 3 sensors linear (R2 > 0.89) between 0 and 1000 μg m-3 PM2.5. Differing compositions can cause up to 10× difference in measured concentrations. All sensors highly sensitive to RH but in varying directions.
Williams et al. (2014)	AirBase CanarIT, CairPol CairClip, Speck, Dylos DC1100, MetOne 831, RTI MicroPEM, Senaris ECO PM, Shinyei PMS-SYS-1	X			X							Low-cost sensors reasonably agree with equivalent reference methods for PM2.5 < 50 μg m-3. T and RH effects can be substantial.
Zheng et al. (2018)	Plantower PMS3003	X			X		X					Sensor performance increases with PM2.5

(*continued on next page*)

Table 2 (*continued*)

Study	Sensor(s) Used	Topics										Major finding(s)
		Aerosol Concentration	Aerosol Size, Density, Composition	Aerosol Size Distribution	RH Effects	Optical property effects	Reference method selection	Calibration Methods	On-board Corrections	Intersensor Consistency	Averaging Periods	
												concentrations up to ~125 μg m-3. β-attenuation-based monitors have drawbacks at low concentrations. RH effects can explain up to 30% of the variance in PM2.5 measurements.
Zikova et al. (2017)	Speck monitors (Syhitech DSM501A)		X									Separating data periods into combustion and non-combustion-derived aerosols improves correlation (R2 0.3->0.5). LOD of 10 μg m-3 PM2.5.
Zusman et al. (2020)	Shinyei PPD42NS, Plantower A003				X			X				Nonlinear corrections of RH effects yielded similar or worse model performance compared to linear RH terms

summarized in Table 1. Other techniques such as quartz crystal microbalance (Paprotny et al., 2013) are not yet commercially available at low cost for routine monitoring and so are excluded from this manuscript. Nephelometric devices can include an additional filter for later correction of the real-time data, such as the MicroPEM (Williams et al., 2014) and the AMOD sampler (Wendt et al., 2019), but the principle of operation is largely the same as for the nephelometric LCS discussed in the rest of this manuscript. The AMOD even uses a Plantower PMS5003 sensor like many of the low-cost monitors discussed here.

The measurement principles of low-cost PM sensors are simple: as particles flow through the measurement cavity, the light intensity of the infrared/red light reaching the phototransistor is modulated by the presence of particles in the light path. The intensity of the modulation, i.e. the nephelometric response, is a function of the particle (mass and number) concentration. The optics (diode, transistor, and lens) as well as the geometry of the measurement cell greatly impact the overall nephelometric response of these instruments (Williams et al., 2014). Therefore, even though the measurement principles across low-cost PM sensors are the same, different makes and models of low-cost sensors often perform quite differently.

Scattering of light by particles has been studied for decades and a full discussion is beyond the scope of this paper; readers are referred to texts such as Bohren & Huffman (1998). Instead, here it suffices to note that the light scattering of particles in low-cost PM sensors falls in the Rayleigh, Mie, or geometric scattering regimes. The majority of the number, but not necessarily the mass, of an aerosol population will fall under the Rayleigh and Mie regimes Friedlander (2000). For a given light wavelength, in the Rayleigh and geometric regimes, the scattered light intensity of a single particle is proportional to the particle size (d_p^6 and d_p^2, respectively, where d_p is the particle diameter). In the Mie regime, the transition between Rayleigh and geometric, scattered light intensity is a more complex function of particle size, shape, and refractive index. The Mie regime occurs when (for an assumed spherical particle) the size of the particle is similar to the wavelength of incident radiation. Therefore, for a PM LCS using an infrared/red diode, scattering of particles with d_p's in some range between 600 nm and 2 μm becomes a more complex, not simply proportional, function of particle size. Depending on the size distribution of the aerosol population, this can be problematic as a substantial portion of the total scattering of the aerosol can be due to particles in this size range.

One of the characteristics of low-cost PM sensors is that they generally do not measure individual particle scattering, but the scattering due to a particle population. This ensemble measurement helps mitigate a number of the drawbacks of measuring individual particles by light scattering. Many low-cost sensors use the low-pulse occupancy (LPO) method: the phototransistor outputs either a high voltage (when the light path is clear) or a low voltage (when the light path contains particles) in a modulated pulse (see Fig. 2). The LPO, or the percentage of time in a sampling period where the voltage is low, is given as a raw signal for many (e.g. the Shinyei PPD42NS, Samyoung DSM501A), but not all, sensors. The LPO is proportional to the particle mass concentration similar to nephelometers (He et al., 2020). The nephelometric response is sensitive to phenomena like coincidence of particles in the light path and particle characteristics that affect light scattering (e.g. morphology, composition; discussed below).

Other instruments (e.g. the Sharp GP2Y1010AU0F) provide an analog voltage output as opposed to the LPO. Some instruments, like the MetOne NPM, report neither the LPO nor analog voltage and instead perform simple, onboard calibrations and output mass and/or number concentrations. Regardless of provided output though, calibration of the sensor signals are generally tied to mass concentrations through collocated measurements of aerosol concentrations using standardized instruments.

While the previous discussion covers most PM LCS, a different subset of "lower-cost" sensors are the optical particle counters (OPCs) such as the Alphasense OPC-N3 or Particle Plus OPC. OPC-type LCS are differentiated from the cheaper nephelometric LCS in that OPCs detect particles individually (Hagan & Kroll 2020). As with traditional optical particle sizers that are much more expensive, OPC binning is done by measuring the intensity of scattered light from individual particles and assigning the signal to either one bin or using a probability distribution across multiple bins (Walser et al., 2017). Assignment is generally done onboard the OPC LCSs where a priori calibrations are used to relate scattered light intensity to particle diameter. Given that optical particle sizer (OPS)/OPC technology has been available and understood for several decades (Hagan & Kroll 2020 and references therein) and that OPCs cost several hundred dollars per unit putting them somewhere between true low-cost sensors and traditional OPS units, these sensors are not considered further in this manuscript.

While a fully theoretical approach to calibration would be beneficial for LCS devices deployed without a collocation calibration, such an approach is difficult due to the a priori knowledge of the aerosol population that is required for light scattering theory. Hagan & Kroll (2020) have produced a model which uses Mie Theory to calculate sensor response to various aerosol properties. The model is useful for estimating how a given sensor reacts to different conditions, but is more a supporting tool for physical calibrations. A different approach to a theoretical calibration has been presented by He et al. (2020) who showed that the Plantower PMS5003 sensor signal can be reproduced by a transfer function which has similar characteristics to the Mie scattering curve. However, the transfer function is still dependent on aerosol size and composition but its calculation is much less computationally intense than Mie theory. The fact that a transfer function is calculable suggests that exact, explicit knowledge of each sensor's optical response functions may not be necessary. Instead, as long as the sensor response over the desired size range is relatively smooth, then "simple" corrections can be applied to low-cost sensors and there is no direct need to solve the governing equations for the various scattering regimes for each sensor. Additionally, the fact that comparatively simple transfer functions are relatively accurate implies that the main drawbacks of measuring aerosol mass from nephelometric response are less important for the expected applications of low-cost PM sensors. The complex aerosol composition, shape, and morphology (Hagan & Kroll, 2020; He et al., 2020) that affect an aerosol's refractive index may all therefore be "hidden" in relatively simple calibration functions for low-cost PM sensors.

3. The complexity of sensor calibration

Calibrating low-cost sensors is a relatively simple experimental process, but a complex topic. In general, the process consists of simply measuring "actual" aerosol concentrations with an instrument which serves as a trusted reference side-by-side with a low-cost sensor. Then, one needs to find a calibration function that yields the best results, while fitting within other design constraints (e.g. computational complexity, software requirements). The complexity arises from the fact that performing a calibration leads rise to numerous questions. Over what range of particle concentrations can LCS be applied to? Which reference instrument should be used? What is an appropriate concentration measurement range for the calibration? What type (e.g. composition, morphology) of aerosols should be used in calibrations? Should calibrations be done with poly- or mono-disperse size distributions and which ones? What atmospheric conditions (RH/T) should calibrations be performed at? Where should calibrations be performed; are lab-based calibrations equivalent to field-based? How are appropriate calibration factors determined with respect to time and place? How often should calibrations at these various conditions be performed? With all of these questions answered, how does one fit the data? How long does a calibration then remain valid for a sensor in the field? Here we will discuss the various ways these questions have been answered in the past decade and end with a recommendation of best practices. Note, first, that from this section on, the term 'reference instrument' is used to represent a highly-trustworthy instrument that measures "actual" PM concentrations without explicitly specifying a method or monitor. The ultimate goal of LCS should be to provide measurements as close to a FRM measurement as possible but since access to FRM instruments is not always possible, we do not exclude the following discussion as applying to FRM, eRM, and research-grade monitors equally. The reader should be aware that ideally a FRM or eRM would be used to calibrate LCS and we phrase the following discussion to reflect that ideal but the following is not dependent on FRM/eRM access. Second, note that this manuscript deals primarily with $PM_{2.5}$. The methods described in section 3.4 apply to calibrating LCS for PM_1, PM_4, and PM_{10} equally as well as $PM_{2.5}$ but sensor performance and sensitivities at these other sizes, PM_{10} especially, can vary much more as compared to $PM_{2.5}$ (Crilley et al., 2018; Jayaratne et al., 2018; Tryner et al., 2020a). However, due to the importance of $PM_{2.5}$ on human health, discussion of measurement of the other PM sizes is left to another paper.

3.1. Sensor linearity and limits of detection

Despite the limitations and drawbacks of calculating aerosol mass from nephelometric response, some low-cost sensors have been shown to have linear responses to ambient $PM_{2.5}$ concentrations as high as 300 μg m^{-3} (e.g., Liu et al., 2019; Malings et al., 2020; Masic et al., 2020; Stavroulas et al., 2020;Wang et al., 2015; Williams et al., 2014; Zheng et al., 2018). Despite this potentially large window of $PM_{2.5}$ concentrations for which LCS can be used, the actual story of sensor linearity is fairly complicated. Much of the complication comes from the fact that sensor linearity is likely due to a combination of factors such as individual sensor design, sensor algorithms, lab versus ambient tests, and the various confounding factors that affect LCS (see section 3.3). Studies such as Wang et al. (2015a) and Hapidin et al. (2019), for example, showed that in laboratory tests using incense burning particles with concentrations up to 1000 and 500 μg m^{-3}, respectively, R^2-values exceeding 0.8 compared to research-grade monitors with no or minor corrections were attainable for the five different sensors tested (Shinyei PPD42NS, Samyoung DSM501A, Winsen ZH03A, Novafitness SDS011, and Sharp GP2Y). Contrast this result with ambient studies such as Crilley et al. (2020), Johnson, Bergin, Russell, & Hagler (2018), Zheng et al. (2018), or Han et al. (2017) which show that sensor linearity for similar sensors ends at an upper range of a few 100's of ug m^{-3} in real-world (i.e. complex aerosol sources, RH and T variations) conditions. Furthermore, sensor algorithms can play a major role in linearity where on-board corrections such as Plantower's "CF = 1" and "CF = atm" can shift the upper limit of linearity from as low as 40 μg m^{-3} to 100's of ug m^{-3} (Kelly et al., 2017; Kosmopoulos et al., 2020; Zheng et al., 2018). In the end, the only sure statement is that the upper range where aerosol concentration responses are linear for low-cost sensors has been shown to be smaller than those of both eRM and research-grade monitors (Zheng et al., 2018). This in turn may necessitate the use of non-linear corrections above certain concentration thresholds (Gao et al., 2015). This upper limit of concentration where linearity holds therefore necessitates end users to think about the anticipated concentrations at their deployment locations. Thinking about anticipated concentrations should also extend to the development of calibration models in that models developed solely in a "low" or "high" concentration range may not necessarily transfer to measurements outside of the range the calibration is built on. This is discussed more in the need to localize calibrations for LCS in the following section.

Low-cost sensors have higher limits of detection than reference monitors. Lower limits of detection (LLOD) for low-cost sensors have been shown to have a wide range of values in laboratory tests. Wang et al. (2015b) showed a LLOD for Sharp GP2Y sensors as high as 26.9 μg m^{-3}. Bulot et al. (2020) (which also provides an excellent review of a number of studies that report limits of detection in field and laboratory environments) conversely showed LLOD well below 1 μg m^{-3} for 5 different sensors. Austin et al. (2015) and Kelly et al. (2017) both showed similar results for Shinyei PPD42 sensors and Plantower PMS1003/3003 sensors with LLOD's calculated to be 1 and 1–3.22 μg m^{-3}, respectively. Outside of laboratory conditions though, sensor responses to meteorological and other conditions tends to bring the LLOD up. In the same paper, Kelly et al. (2017) showed that the units with a 1–3.22 μg m^{-3} LLOD in the laboratory rose to 10.5 μg m^{-3} in ambient conditions. Similarly, Zikova et al. (2017) showed a LLOD for Synhitech DSM501A sensors (inside Speck monitors) of 10 μg m^{-3} (at 1-min data reporting; 9 μg m^{-3} for 1-h averaging) for indoor and outdoor ambient measurements. Bulot et al. (2019) showed similar LLODs in agreement with the manufacturer recommendations for 4 different sensors in ambient conditions as well. This discrepancy between laboratory and ambient LLODs is not overly concerning given the number and impact of confounding factors for LCS (see section 3.3). Since most LCS users deploy sensors in ambient conditions, the LLOD of ∼10 μg m^{-3} should be kept in mind as sensor performance in cleaner settings (e.g. remote locations or parts of the Global North) could be impacted. Though the use of LCS in cleaner environments (average $PM_{2.5}$ < 10 μg/m^3) may seem challenging, studies have shown that with careful corrections and long-term averaging, the uncertainties can be reduced to about ±1 μg/m^3 or ∼10% at these lower levels (Malings et al., 2020; Eilenberg et al., 2020) and the data used to determine long-term source contributions of <1 μg/m^3 (Eilenberg et al., 2020), justifying their use in such environments.

3.2. The importance of reference method selection

In general, the PM mass concentration is measured to ensure regulatory compliance. In both the US and EU, PM regulations are based on reference methods which use gravimetric analysis of filters at a specific temperature and RH (Carlton & Teitz, 2002; Noble et al., 2001). However, calibrating low-cost sensors with reference methods is difficult since gravimetric analysis is a manual, off-line method that generally yields low time resolution (24-h), integrated measurements. These drawbacks mean that calibrating with filter-based gravimetric reference methods can take considerably longer to both collect enough measurements over a wide range of PM concentrations and to analyze them. Additionally, all of the time-resolution improvements that low-cost PM sensors can offer are not evaluated when using gravimetric reference methods. The next-best option is then to use equivalent reference methods (eRMs) which are used to provide higher time-resolution (typically hourly) PM mass concentrations. The most common eRMs are the beta attenuation monitor (BAM) and the tapered element oscillating microbalance (TEOM) though some light scattering instruments are also becoming commonplace (e.g. Teledyne T640, GRIMM EDM180, Fidas). The BAM, TEOM, and light scattering instruments all meet defined standards for being classified as equivalent to the reference methods (for 24 h averaging periods), but importantly yield measurements at higher time resolutions than daily. Other research-grade optical measurements (e.g. GRIMM, Dustrak) and electrical mobility measurements (SMPS) can be used in place of the eRMs when performing calibrations for low-cost PM sensors, but then the applicability of LCS data to regulatory limits becomes less clear.

The choice of which reference method, eRM or other method is used to measure the "true" PM mass concentration is often based on practical considerations, namely which instrument is physically available where and when calibrations of low-cost PM sensors can be performed. Unfortunately, the choice of method affects perceived sensor performance, with various low-cost PM sensors under- or overestimating the reference methods in various conditions. Low-cost PM sensors would ideally have high agreement with 24-h gravimetric measurements, which are the gold-standard for regulatory uses. Literature results have varied. Gao et al. (Gao et al., 2015) tested multiple Shinyei PPD42NS sensors and showed moderate correlation ($R^2 = 0.53$) between the sensors and the gravimetric method for $PM_{2.5}$ in Xi'an China. Kelly et al. (2017) and Sayahi et al. (2019), however, showed excellent correlation between $PM_{2.5}$ gravimetric analysis and Plantower 1003 and 5003 units in Salt Lake City, UT, USA. It is unclear if the difference between these studies is simply the model of sensor or some other issue. Though using gravimetric methods for calibrations has been shown to be able to generate calibrations with low biases and relative standard deviations with respect to low-cost PM sensors (Tryner et al., 2019a, 2019b), the 24-h integration time can require much longer collocations to adequately capture the entire range of meteorological conditions and PM concentrations that a PM LCS might experience. The BAM is a common choice of eRM allowing hourly measurements to mitigate this drawback of gravimetric methods. For the BAM, the overall correlations for $PM_{2.5}$ from PM LCS are similar to the gravimetric methods. Feenstra et al. (2019) showed correlations ranging from poor to great ($0.38 < R^2 < 0.95$) for 11 different sensors. Gao et al. (2015) observed higher correlations ($R^2 = 0.87$–0.92) for hourly $PM_{2.5}$ measurements from PUWP sensors in Xi'an China, as did Kosmopoulos et al. (2020) for hourly PMS 1003/3003 ($R^2 > 0.85$) in Patras, Greece. With the TEOM, the story is similar to the gravimetric and BAM methods; correlation with various low-cost sensors also ranges from poor to high ($R^2 = 0.36$–0.9) over multiple time integration periods, concentration ranges, and deployment locations (Badura, Batog, Drzeniecka-Osiadacz, & Modzel, 2018; Feinberg et al., 2019; Johnson, Bonczak, & Kontokosta, 2018; Kelly et al., 2017; Liu et al., 2019; Sayahi et al., 2019; Steinle et al., 2015). Another option to obtain even higher time-resolution reference measurements is to use optical methods such as the Teledyne T640 or Fidas 200. Zheng et al. (2018) showed that the precision of eRM instrument used in LCS calibrations is critical in evaluating sensor performance and that optical methods such as the T640 may be better than beta-attenuation-based monitors, at least at low concentrations for PMS3003 sensors. Ultimately, the differences in errors between the various reference methods and low-cost PM sensors do not necessarily arise from either the low-cost sensors or the reference methods themselves. Rather, the differences come from how each instrument handles the various confounding factors as well as the effects of factors like averaging time, which are discussed below.

3.3. Confounding factors in LCS calibration

Obtaining high-quality data from low-cost PM sensors is complicated by the fact that these sensors are sensitive to many factors. Table 2 lists a number of different categories into which low-cost PM sensor sensitivities can fall into in the Topics column. These limitations include the measurement of ensembles and not of single particles (light path coincidence and particle concentration) and the factors that affect light interactions with particles (RH, morphology, composition). Fig. 3 shows how some of these factors can affect low-cost PM sensor responses. The exact effects of morphology are highly variable and shown as multiple possible sensor responses. In general, the literature provides evidence that all of these factors are correctable and obtaining good-quality data from low-cost sensors is possible with some careful consideration of exactly what and how PM is being measured.

Despite the various sensitivities of low-cost PM sensors, literature suggests that correcting low-cost sensors is possible with a small amount of work to characterize a given sensor's reactions to various meteorological and operating conditions. Before corrections can occur though, users should determine if all their units respond similarly to changes in PM concentrations. Crilley et al. (2018), Liu et al. (2019) and Malings et al. (2020) all showed high intra-unit consistency with variability less than 20% between units. On the other hand, some work (Castell et al., 2017) has shown a notable unit-to-unit variation with different AQMesh units showing different responses to RH and temperature ranges ("high" PM bias for some units at RH < 50%, negligible for other units). SC-AQMD also showed similar notable unit-to-unit variation with Airviz Speck units (with DSM 501 sensors) (SCAQMD, AQ-SPEC). The cause of interunit variability is unclear but may be due to manufacturing tolerances with certain brands. However, users can determine if their interunit variability is within acceptable bounds for their use cases and then determine if individual or ensemble correction factors are

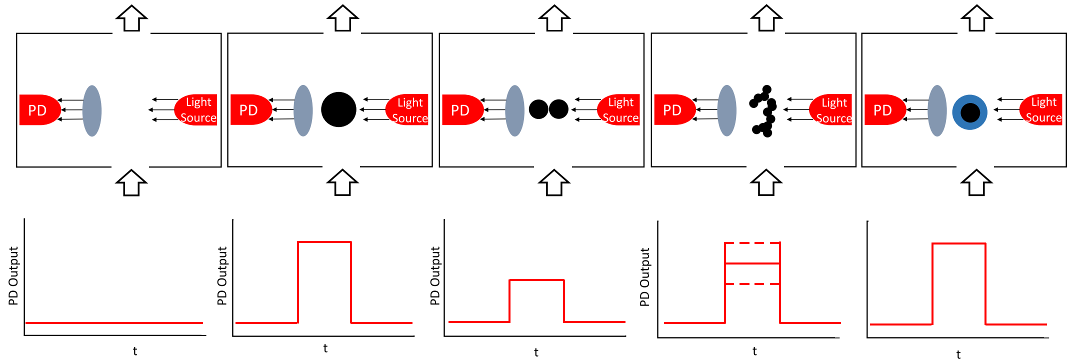

Fig. 3. Representative diagram of a low-cost sensor consisting of a photodiode (PD), light source, and a measurement cavity (top row) and the resulting output of the PD (bottom row) for 5 different cases each measuring equivalent particle volumes: (a) no particles present, (b) 1 large particle, (c) 2 coincident particles, (d) a particle with complex morphology which can yield variable sensor responses, and (e) a humidified particle with a coating of water.

more appropriate. The US EPA does have guidance for LCS performance metrics (see section 3.4) that can help users determine what variability might be appropriate for them (Duvall et al., 2021). The gains in individual versus ensemble correction factors (i.e. using a custom correction for each sensor unit versus using a common correction applied to all units of the same model), however, may not warrant the extra work in all cases with all sensors.

Some of the easier confounding factors to measure and control for are meteorological conditions, namely ambient temperature and relative humidity (RH). Because LCS generally do not include a heater or dryer at their inlets, the detected particles include water that should not be included in the measured PM mass concentration. Most work has shown that RH starts to significantly impact low-cost PM sensors when it exceeds 80–85% (Crilley et al., 2018; Liu et al., 2019; Malings et al., 2020) though Magi et al. (2020) showed that for Plantower PMS 5003 units the impact may start lower at 65–70% and Jayaratne et al. (2018) showed that even at 50% RH, Sharp GP2Y and Shinyei PPD42NS sensors show an appreciable artifact. However, it should be noted that Kosmopoulos et al. (2020) reported that the effect of RH on the response of Plantower PMS 5003 sensors for fine PM in South-East Europe was negligible so measurement location is important to consider. Generally, air quality regulations stipulate "dry" PM mass, so correcting for these hygroscopic effects is generally necessary. Some work has shown that linear RH corrections perform as well as or better than non-linear corrections (Zusman et al., 2020). On the other hand, Zheng et al. (2018) has suggested that empirical nonlinear RH corrections that have been previously applied to other nephelometric PM instruments can improve PM concentrations in low-cost sensors to within 10% of dry values. Crilley et al. (2020) have shown that applying κ-Köhler theory (using average bulk particle hygroscopicity) can improve LCS PM measurements to within 33% of TEOM measurements. Malings et al. (2020) suggested that empirical corrections are possibly a better choice than theoretical approaches because they result in lower mean absolute errors and higher correlation coefficient values between reference and low-cost sensors. One drawback of the empirical approach is that resultant models will be strictly limited to the RH range that the model is trained on. The hygroscopic growth approach should theoretically be applicable to RH ranges not explicitly measured. Ultimately the choice to use a linear, theoretical, or empirical method to correct for RH is up to the end user's preferences and needs, but there is broad agreement in the literature that some correction needs to be made in many situations.

The case for corrections for temperature, however, is a bit less convincing than the case for RH. Most of the literature has shown minimal effects of temperature on low-cost PM sensors (e.g. Liu et al., 2019; Wang et al., 2015a) with even low temperatures between -5 and 5 °C causing positive or negative errors (-5 to 5 μg m^{-3}). Oftentimes, the error terms on temperature coefficients in linear/quadratic correction formulae (see next section) are larger than the coefficients themselves and therefore temperature effects can generally be ignored, though some studies have shown this to not always be the case (Magi et al., 2020). The general approach of ignoring temperature effects in calibration models may not hold true in extreme environments (hot or cold deserts especially), but more work is needed for the application of LCS in such extreme conditions.

In addition to meteorological factors, sensor age and/or deployment history may also affect their performance. Sayahi et al. (2019) noted substantial sensor drift in one of four PMS 5003 sensors over a period of 320 days. Tryner et al. (2020a) showed some sensors report erroneously after exposure to high concentrations of Arizona road dust in the lab. On the other hand, Crilley et al. (2018) noted no substantial drift over a 4-month deployment of Alphasense OPC-N2 units under ambient conditions in Birmingham, England in September 2016. Malings et al. (2020) collocated two Met-One NPM sensors with reference BAMs (an eRM) over 16 months and found seasonal variations in mean monthly bias, with positive bias in the winter and zero or negative bias in the spring and summer, thus rather a cyclic behavior than a linear time shift. Overall, for large deployments of sensors (where singular faulty units can be identified using the ensemble) with collocation calibrations, the effects of sensor aging are likely negligible for deployments on the order of a year, but more research is needed on the topic of long-term LCS signal stability. However, the effects of deployment history or exposure to periods of high concentrations of aerosols (including dust) are factors that must be kept in mind, especially with longer deployments of sensors.

A number of studies have also shown that aerosol size, density, and composition play a potentially important role in sensors' linearity and overall sensitivity. Aerosol size and composition (i.e. refractive index) both inherently change particle scattering and absorption (see Mie Theory) and therefore are expected to significantly impact any light-based PM measurements. Similarly, particle density differences also impact light-based measurements for the same reasons and complicate the connection between particle size and mass. Several studies (Liu et al., 2017; Salimifard et al., 2020; Sousan et al., 2016) have shown that low-cost sensors exhibit different responses to aerosols of differing physical and chemical properties. Liu et al. (Liu et al., 2017) showed that sensor response slopes to changing PM concentrations, for three different low-cost sensors (Oneair, Dylos, and Samyoung), changed by up to a factor of 2–3 for different inorganic species. Salimifard et al. (2020) similarly showed up to a 20% difference in the response between organic and inorganic PM measurements for each of two sensors (Dylos and Speck) and Li et al. (2020) showed an even larger difference in response for nine different sensors measuring dust, sea salt, and incense aerosols. Demanega et al. (2021) showed an even larger discrepancy between 8 different sensor responses to multiple types of indoor aerosol sources with sensors showing up to a 50% difference as compared to reference research-grade monitor. Sensors that provide segmented PM measurements, e.g. $PM_{2.5}$ vs PM_{10}, like the Dylos 1700, are especially sensitive to particle size distributions (Han et al., 2017; Kuula et al., 2020) but even other sensors can have varying responses to different PM distributions, even to the point of size bins not being independent (Kuula et al., 2020; Liu et al., 2017). Kosmopoulos et al. (2020) showed that the PurpleAir sensors had very different response even for $PM_{2.5}$ during which their suburban site was impacted by Saharan dust and therefore the aerosol had a quite different size distribution. Aerosol composition has also been suggested to affect low-cost PM sensors such that even partly organic compositions can deviate from calibrations done with inorganic aerosols (Wang et al., 2015a). The same trend has been shown in field deployments where some sensors have better agreement with reference monitors at low-traffic than high traffic sites or time periods (both $PM_{2.5}$ and PM_{10}, (Castell et al., 2017); $PM_{2.5}$ (Subramanian et al., 2020)) and when measuring combustion and non-combustion derived aerosols (Zikova et al., 2017). Some work has been done to reconstruct PM mass from the counts reported in the different size bins of the more expensive optical sensors (i.e. the OPC-N2) thereby turning the issue of size distribution sensitivity around, but this method is generally not available to less expensive sensors with only a couple of size channels (Crilley et al., 2018; Di Antonio et al., 2018). Kuula et al. (2020), He et al. (2020) and Tryner et al. (2020a) have all shown that the particle size distribution data reported by the low-cost PMS5003 sensor are not reliable. When tested separately with particles of specific sizes over and above the submicron range (0.1 μm to 2.0 μm), PMS5003 sensors can show similar distributions in all size channels.

Another factor that needs to be considered when calibrating low-cost PM sensors (and LCS in general) is the averaging times that sensor calibrations are built on and applied to. Eilenberg et al. (2020) showed that both instrument and sampling uncertainty with respect to reference-methods drop as averaging time is increased from 1 h (40–50% individual sensor uncertainty) to 100 d (10% individual sensor uncertainty). When combined with reports of high limits of detection at low averaging times (1 min, Zikova et al., 2017) and the fact that each low-cost sensor outputs data at varying intervals, the importance of carefully selecting data averaging periods is apparent. However, though minimizing uncertainty and error and increasing limits of detection are important, completely negating the temporal resolution increases that low-cost PM sensors afford may not be helpful for a given use case. As with all of the caveats of low-cost sensors discussed here, consistency and transparency in dealing with these issues are by far the most important attributes when reporting low-cost data.

3.4. Methods to account for confounding calibration factors and their evaluation

Though there is generally a high degree of linearity between reference and the raw low-cost sensor signal for aerosol concentrations, this does not necessarily mean that linear equations are the best way to convert measured scattering to actual concentrations. Much work has been performed for various low-cost sensors determining what preprocessing of the data is necessary, what types of functions are ideal, and what terms are necessary to account for. A growing trend is the application of machine learning algorithms to low-cost sensor data, but it is unclear if the increase in computational complexity is worth the accuracy gains (if/when they exist) over simpler methods.

No matter what method is selected for correcting low-cost PM sensor data, reporting metrics that describe the fit and predictive power of the correction model is of paramount importance. Up to now we have mostly discussed the coefficient of determination, R^2, as it is the most common metric reported. R^2 by itself only quantifies the strength of association between two variables but not necessarily their agreement and is sensitive to factors such as the range of the measurements (Alexander et al., 2015). A measure of correlation (R or R^2) is therefore necessary when assessing low-cost PM sensor calibration models, but alone it is not sufficient. In addition, the bias of calibration models should also be reported. Bias should be reported both as absolute and normalized values as discussed below. The mean normalized bias (MNB) is shown in Eq. (1):

$$MNB = \frac{\sum_{i=1}^{n}\left(c_{estimated,i} - c_{true,i}\right)}{\sum_{i=1}^{n}\left(c_{true,i}\right)} \tag{1}$$

where c_{true} and $c_{estimated}$ are PM concentrations measured by the reference monitor and LCS, respectively. A mean-normalized bias helps quantify the accuracy of the measurements over the collocation period. Note that MNB here is used as a model evaluation tool, not a data correction. Once the bias has been reported, precision can be reported as bias-corrected. For precision metrics, mean absolute error (MAE; eq. (2)) and root mean squared error (RMSE; eq. (3)) are often, but not always, reported.

$$MAE = \frac{\sum_{i=1}^{n}\left|c_{estimated,i} - c_{true,i}\right|}{n} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(c_{estimated,i} - c_{true,i}\right)^2}{n}} \tag{3}$$

Both MAE and RMSE give the average model prediction error and can range from zero to values as high as the measured concentrations themselves. RMSE weights large errors much more than MAE, which may or may not be desirable to a specific use case. Since calculating both metrics is fairly straightforward, both should be reported. However, because calibration models are often built on different scales of PM concentrations, normalizing the accuracy metric is important so that models can be compared. Normalizing performance metrics allows models to be appropriately compared between environments where concentration ranges are different. For example, an uncertainty of 10 μg/m^3 on a measurement of 10 μg/m^3 is very different than the same error on a measurement of 100 μg/m^3. Both the MAE and RMSE can be normalized by the mean, standard deviation, difference between the maximum and minimum, the interquartile range, or any similar factor. The mean is likely the most physically intuitive statistic for PM concentration data, so low-cost PM sensor users should be reporting the bias-corrected mean normalized MAE and RMSE (CvMAE, nRMSE, eqs. (4) and (5)) for their models.

$$CvMAE = \frac{\sum_{i=1}^{n}\left|c_{estimated,i} - n_{bias} - c_{true,i}\right|}{\sum_{i=1}^{n}\left(c_{true,i}\right)} \tag{4}$$

$$nRMSE = \frac{\sqrt{\sum_{i=1}^{n}\left(c_{estimated,i} - n_{bias} - c_{true,i}\right)^2}}{\sum_{i=1}^{n}\left(c_{true,i}\right)} \tag{5}$$

where

$$n_{bias} = \frac{1}{n}\sum_{i=1}^{n}\left(c_{estimated,i} - c_{true,i}\right) \tag{6}$$

It should be noted that only reporting normalized metrics may also be misleading; in the above example, reporting uncertainty as 100% based on the first set of data may not be applicable when conditions change to the second set of data. Instrument manufacturers themselves generally recognize this and provide both percentages (more relevant at high concentrations) and absolute uncertainty (relevant at low concentrations) (e.g. see Plantower Data Sheet). Therefore, reporting both the normalized and absolute metrics is necessary. These accuracy metrics should relate to the shortest time intervals that are used (e.g. hourly) when the models are constructed. Though many LCS can give data on the order of minutes, if the reference data used to build the calibration model is not also on the same order of timescale, evaluating the performance of calibrations at the shorter timescale is not explicitly knowable. The best that can be done is to report the (e.g.) hourly metrics and assume (while clearly stating so) that the minute data will have similar performance. The question of whether even using data on timescales that cannot be validated is desirable is left to the end-user and specific applications of LCS. If 1-min sensor data is really the end-goal of using LCS, then 1-min FEM (generally optical methods) should be used to build the required calibration models. Regardless of this caveat, only with these metrics, at a minimum, can calibration models or correction methods be properly compared against one another. Many previous sensor papers have simply not reported enough metrics to confidently evaluate the sensor or calibration model performance (Williams et al., 2019). As it is straightforward to calculate the bias, MAE, RMSE, CvMAE, and nRMSE, at a minimum all of these metrics should be reported.

Linear and quadratic/higher-order polynomial fits are computationally the least expensive ways to increase the agreement between low-cost PM sensors and reference methods. Here the issue of how differences in various sensor outputs and how the choice of output (e.g. raw signal vs built-in processing) can affect sensor calibration is set aside. This section deals with the post-processing done on whatever sensor output is used. Because of the high linearity of low-cost sensors with various reference methods and aerosol concentrations, nearly all of the literature on the subject includes linear methods for converting sensor output to calibrated PM values. As calibration functions should ideally correct for (or at least approximate, in the case of non-linear terms, e.g. hygroscopic growth) biases, the terms that show up in calibration equations are consistent with many of the sensor biases discussed in the previous section. Hagler et al. (2018) define most of these variables as defensible (relative humidity, temperature, elapsed time, etc.) when creating correction formulae, as opposed to questionable (wind speed/direction, temporal factors other than elapsed time, atmospheric mixing height, etc.) variables which move the calibration more into the realm of a statistical predictive model. The vast majority of studies have shown that including temperature and relative humidity in linear fits can improve calibration quality to some degree (e.g. Crilley et al., 2020; Jiao et al., 2016; Zheng et al., 2018). Similarly, many have also included higher-order polynomial terms to capture non-linear behaviors (e.g. (Malings et al., 2020; Zheng et al., 2018)). However, given inherent uncertainty with low-cost sensors, complex equations of multiple variables may not be needed; Malings et al. (2020) examined a wide (130+) combination of linear and quadratic regression fits but most did not reduce uncertainties much more than linear regressions that use reported PM mass and relative humidity. In general, if an explicit correction term for RH is included in a linear or quadratic calibration equation, then particle hygroscopic effects are implicitly assumed to be captured in that term even though actual hygroscopic effects are likely more complex in nature (see efflorescence/deliquescence curves, Köhler theory, etc). Aerosol hygroscopic growth and the increase or decrease of aerosol size as a function of relative humidity is the subject of much research but, for the purpose of sensor correction, linear and quadratic RH fits (i.e. c*RH or c*RH^2) appear to be adequate though they are masking complex behaviors. Other aerosol properties that are known to influence nephelometric responses (e.g. particle shape, refractive index, etc.) are generally explicitly ignored and are implicitly assumed to be wrapped up in the linear/polynomial coefficients. As a consequence, calibrations are then specific for a given aerosol source/composition/etc. (see below). Linear calibrations generally perform extremely well for most field deployments of low-cost sensors, commonly reaching R^2 values above 0.9 with relatively low errors (though errors are not reported nearly often enough in the literature). The ability of linear and polynomial fitting, depending on the exact choice of terms, to be performed either on a circuit board or with basic computer skills allows for a very inclusive set of users to understand how their data are processed.

The next step on the computational complexity ladder for calibration methods for low-cost PM sensors are the statistical methods. Often times, these methods are used to determine what terms are most important in linear/polynomial fitting. Bayesian information criteria has been a common method over the past five years to accomplish this (Austin et al., 2015; Castell et al., 2017; Gao et al., 2015; Johnson, Bonczak et al., 2018). Applications of statistical methods require much higher statistical, mathematical, and computational literacy than linear/polynomial fitting, but are much more transparent than machine learning algorithms.

The last step on the computational complexity ladder is machine learning. A number of studies have applied various machine learning algorithms to various makes and models of LCS from neural networks (Chen et al., 2018; Li et al., 2018) to random forest models (Liu et al., 2019). A number of LCS manufacturers also offer calibrations of data using unspecified and/or proprietary machine learning algorithms e.g. (Morawska et al., 2018). Overall, the results of machine learning applications are positive, with R^2 values $>$ 0.5 generally reported (Chen et al., 2018; Johnson, Bonczak et al., 2018; Liu et al., 2019). However, machine learning algorithms are not necessarily better than simpler and completely transparent calibration methods; as discussed above, linear calibrations can yield strong correlations ($R^2 > 0.8$) and near-zero bias. Given the computational requirements, the lack of transparency, and most importantly the lack of greatly improved results, machine learning algorithms do not (at present) appear to be preferable to simple linear fits to obtain good quality data from low-cost PM sensors.

4. Best practices for sensor calibration

Overall, the performance of low-cost PM sensors is subject to numerous factors and both the accuracy and precision of the corresponding measurements can be highly variable depending on how these factors are addressed. The ideal approach when deploying low-cost PM sensors is, therefore, to perform a full battery of characterization and calibration tests, in a similar micro-environment the sensors will be deployed in, over all of the settings and conditions expected for the duration of the deployment. Here, a full battery of tests refers to characterizing sensor linearity, RH and T response, as well as examining sensor response to different aerosol sources. For collocation calibrations, these tests are either inherent to the collocation itself or simply require parsing the collocation data into relevant periods (e.g., relatively clean versus relatively polluted periods, high RH and T versus low RH and T periods, examining different times of day or year to draw general conclusions about source contributions where applicable, etc.). Deploying sensors in the same environment where they will collect data helps minimize the impacts that particle composition and size distribution differences have on LCS, effectively by smoothing over a large ensemble of conditions. However, for practical reasons this approach is often not possible. Given the literature already published on low-cost PM sensors, we can recommend a set of best practices that should generally yield high-quality data from these sensors. Note that these recommendations are aimed primarily at ambient measurements, the LCS use case that the authors are most familiar with. Measurements in specific micro-environments dominated by a specific size and/or composition of PM (e.g. Zamora et al., 2019 may require significantly shorter LCS characterization periods. In general, if any of these recommendations preclude or hinder attaining high-quality data in a different use case, then the primary recommendation of being transparent and explicit about the steps taken to obtain high-quality data is all that must be followed.

The physical setup of a collocation calibration should generally consist of a reference (FRM, eRM) or research-grade monitor and all of the low-cost PM sensors of interest deployed at the same location (ideally in the same environment that the units will ultimately be deployed in) for some period of time. A simultaneous collocated calibration with all low-cost units is the easiest way to identify faulty units and the best way to account for intra-unit variability if making a general (as opposed to unit-specific) calibration model. While collocations of very large networks (50+ LCS) are not always feasible due to space constraints, efforts should be taken to collocate as many units as possible. Outstanding units can then be handled in one of two ways. One way is to keep a subset of units collocated with the reference monitor for the entirety of the available collocation time and rotate out the remaining units for a few days/weeks until every unit is collocated. The other way would be to examine the intra-sensor consistency over the course of the post-collocation deployment during periods when pollution across the network is expected to be relatively similar (e.g. due to minimal influence of local sources) but this requires additional information and may not always be feasible. The choice of which approach to take with large network deployments is up to the end user but either approach should yield the same benefits. The variables of interest for the collocation then revolve around three factors: the conditions (atmospheric and PM concentration, composition, and size distribution) over which measurements are made, the choice of reference monitor, and the length of time of collocated deployment.

The most important factor in a collocation is ensuring that the collocation period covers as wide a range of concentrations, compositions, and atmospheric conditions as could possibly be experienced by the sensors during post-collocation deployment. Note that in-situ measurements of factors like aerosol size distribution and composition are not necessary for LCS calibration but it is important that collocation periods and locations be selected with some a priori knowledge of the deployment location. Questions such as “will this collocation cover the rainy season/biomass burning season/winter”, “does the collocation site have access to air transported from all the potentially major source regions for the eventual deployment location?”, or “is the collocation site dominated by a single nearby source?” are questions that do not necessarily require size distribution or chemical speciation to answer but rather require knowledge of the major sources that affect the site. One potential way to ensure that major sources are covered (and therefore the ranges of aerosol concentrations and compositions that LCS would be expected to measure) is to do a multi-site collocation where LCS are moved between, e.g., an urban background site, a roadside site, and at a near-source site (e.g. industrial complex), similar to the approach of Malings et al. (2020). This approach to collocation is more work than a single-site study but can help alleviate pressures to find a perfect single site. It is important to note that the overall length of time required for a given collocation will depend on the reference monitor used. If the reference monitor of choice is a gravimetric reference method with a 24-h time resolution, then the collocation may have to take place over much longer periods than if the reference monitoring was at higher (~hourly) time resolution, as the integrated samples will average out variations in meteorological conditions (e.g. RH and T) and source impacts. On the other hand, a number of studies have noted that when comparing LCS and reference instruments with longer time integrations, R^2 values increase and errors decrease (Holstius et al., 2014; Kelly et al., 2017; Malings et al., 2020; Zikova et al., 2017). In practice, measuring over the entire meteorological and concentration range can be difficult without a full year of collocation (Zimmerman et al., 2018). One way to ensure this requirement is met is to periodically collocate (preferably seasonally) sensors with the reference monitors in environments representative of sites across the network, e.g. urban background, source-oriented, or roadside, for a minimum of 4 weeks per year, at different times of the year. A 4-week collocation was shown to decrease mean absolute error from low-cost gas sensors by 50% (Zimmerman et al., 2018). If there is a high correlation between all the sensors in a network, only a subset of LCS devices may need to be collocated seasonally with the reference monitors, reducing operational effort and costs. This result has not been explicitly verified for low-cost PM sensors and should be investigated to see if this holds as true as it does for low-cost gas sensors. Permanently collocating some LCS devices with the reference site is another way to capture seasonal effects on LCS performance without bringing all the nodes back to the central calibration site(s). There may also be utility in including 2 or more sensors in one housing (e.g. as PurpleAir does) to add redundancy and confidence to measurements, especially as these setups often correlate well with each other (Ardon-Dryer et al., 2020). As low-cost sensors are generally cheap, this approach should be adopted where possible. Overall, as long as a collocation covers the range of temperature, humidity, PM concentration, and major sources which are likely to be encountered in a field deployment, the choice of length of time and method of collocation is simply a matter of logistics. If a single or multi-site continuous collocation approach cannot be adopted, an alternative approach could be to perform both a pre-deployment and a post-deployment calibration and utilize an average or a moving calibration model.

Once the collocation is completed, the question then arises of how to process the data obtained to retrieve a useful correction equation. The first part of this question is what data output of a sensor is most applicable. This question is most relevant to Plantower PMS sensors, which provide mass concentrations with two correction factors, CF = 1 "for laboratory evaluations" and CF = atm "for field evaluations", as per the manufacturer's datasheet (Digial Universal Particle Concentration Sensor PMS5003 Series Data Manual, 2016). The details of these calibration factors are proprietary and unavailable from the manufacturer. According to Kaiterra (Bates, personal communication, 2020; https://www.kaiterra.com/en/index/), a company that uses Plantower sensors to make air quality monitors and works closely with Plantower, the CF = 1 output is based on laboratory testing of Plantower sensors collocated with a BAM while sampling a smoke/soot mixture. The CF = atm output is based on ambient collocations in Beijing, China. Barkjohn et al. (2020a) showed a slightly higher correlation compared to a TEOM when using CF = atm versus CF = 1 measuring in various locations in Beijing, China with $PM_{2.5}$ concentrations between 0 and 60 μg m^{-3} (R^2 = 0.63 vs 0.52, respectively). Stavroulas et al. (2020) found that the CF = 1 was closer and more linear with respect to reference methods at high concentrations (from 110 μg/m^3 to over 300 μg/m^3) – when their sites were most impacted by residential wood burning. On the other hand, below 20 μg/m^3, the two outputs were the same. We have observed similar results at similar low concentrations in Welgegund, South Africa and Pittsburgh, Pennsylvania, USA. Tryner et al. (2020a), on the other hand, showed a large nonlinearity for laboratory measurements at 40 μg m^{-3} when using CF = 1. Wallace et al. (2021) try to use the binned size counts to develop their own correction factors, but the wider applicability of such approaches needs to be studied further, given that Tryner et al. (2020a) and He et al. (2020) show the binned counts are not related to actual aerosol size distributions. Thus, choice of which output to use as the "raw" value may be guided by the source(s) impacting the location, though depending on the ambient levels observed in practice, there may not be much difference. Met One and Alphasense both also provide on-board corrections that are user-definable. Regardless of what on-board correction end users decide to use, however, consistency and transparency should be stressed. As long as details about both the calibration models and the data streams used are provided and consistency between calibration and application periods is observed, data integrity is maintainable. Additionally, in data selection, users must be explicit about the QA/QC procedures applied to their data. The elimination of potentially compromised data should be performed but the criteria for these QA/QC checks must be provided. The fraction of data thus eliminated should also be reported.

With the data output of the sensor selected, a calibration model can be built between the low-cost PM sensor data and reference monitor (or eRM). Here there are three questions to be answered by each end user: "How should the models be built?", "What model (and model inputs) should be used?" and "How many models are needed?" As with all the aspects of sensor calibration discussed here, the goals of the LCS deployment dictate some of the answers but there are a number of best practices that should be followed regardless of the end goals, especially with regards to how to build models. First and foremost, transparent models (i.e. models that can be reproduced elsewhere using the same data) should be prioritized over black-box models. As such, open-source code should generally be selected over proprietary code. Second, model building should include some process for model validation. This is commonly referred to as creating test/train splits within a data set, i.e. reserving some amount of data from the model training to use as a test set. "Leave one out" has historically been used over many fields of science but k-fold cross validation has become a general recommendation in machine learning. K-fold cross validation is simply a resampling procedure used to evaluate models on a limited data set. This technique is generally purported to be a less biased estimate of model skill than a simple train/test split within the machine learning community (Alpaydin, 2020). Models that do not perform similarly across all k-folds of the data set should be disregarded or tuned differently. In reality, the number of folds to choose depends on the learning curve (a diagnostic tool visualizing model learning performance over experience or time) (Hastie et al., 2001) but a k-value between 5 and 10 appears to be adequate for LCS applications (Tryner et al., 2019a; Zimmerman et al., 2018). Some work has shown that K-fold cross validation is acceptable for even highly autocorrelated time-series as long as the errors are uncorrelated (which seems a reasonable assumption for LCS; Bergmeir et al., 2018).

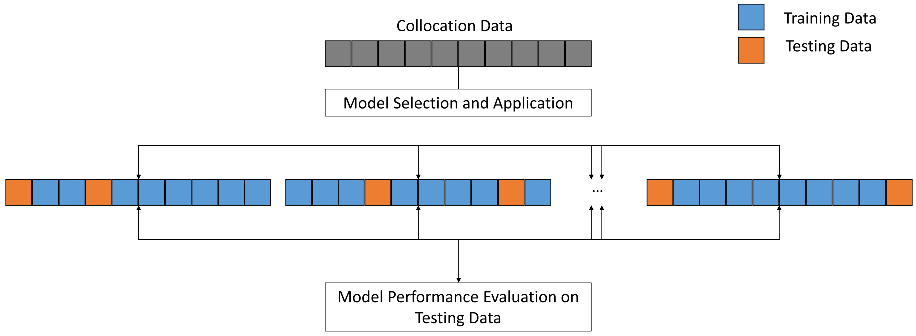

Fig. 4. Visual representation of a random 5-fold cross validation on collocation data.

Because of this, overly constraining the folds by forcing certain time periods of sampling is likely unnecessary. Fig. 4 shows a graphical representation of what a k-folds cross validation actually entails.

The second question regarding what model and inputs to choose is generally more open-ended and requires some experimentation by end users. As discussed in section 3.4, simpler models generally perform better than or as well as complex machine learning algorithms for PM LCS. We therefore recommend that simple linear or quadratic models should be employed most often. Complex models run a risk of over-fitting model training datasets and therefore require additional justification for their use. As for model inputs, users should examine model performance using all of their potential inputs. Users should avoid questionable input parameters not directly related to measurement physics (e.g. time of day, wind speed; see Hagler et al., 2018), but parameters such as meteorological conditions (T and RH) and auxiliary aerosol data such as composition, hygroscopicity, or black carbon concentrations are all good potential input parameters to test. In machine learning this is called feature selection and simply entails creating models with combinations of inputs and choosing only the inputs that increase model performance while also being aware of estimated errors on each model term. As this subject of model creation is naturally open-ended, this discussion is difficult but can be illustrated with an imaginary, yet concrete, example where we limit model inputs to LCS PM output, RH, and T. If a 3-coefficient linear model (PM, RH, intercept) outperforms a 4-coefficient model (PM, RH, T, intercept), then the 3-coefficient model is the better choice as the sensors in question are apparently not temperature sensitive. However, situations can arise where models with more variables outperform models with less variables but the error bounds on a coefficient are greater than the value of the coefficient (e.g. 0.1 ± 0.2). Should the better performing model be used or the model with more certainty in its coefficients? In the end, the authors would err towards the latter but either is likely acceptable as long as the model is transparent.

The third question regarding how many models are needed deals with the idea of whether each sensor should be treated as an individual or whether an ensemble treatment is good enough. For a single collocation of multiple sensors, a general calibration model could be made for the whole batch of sensors to be deployed, or a model for each individual sensor could be made. The former may be more appropriate for dense networks with dozens of sensors (as in Malings et al., 2020), while users with one or a few sensors could generate sensor-specific calibration models (as in Subramanian et al., 2018). A general model is less resource-intensive and for low-cost gas sensors has been suggested to result in a minimal overall performance loss (Malings et al., 2019). It may be assumed (pending a dedicated investigation) that a similar conclusion will hold true for PM LCS, as was shown in Malings et al. (2020). A further abstraction is to create a universal model based off many different collocations as done by Barkjohn et al. (2020c) for Purpleair sensors deployed in the United States. Users should be cautious in this approach as the resulting model may not “travel” well to sensors that are far outside the mean of the sensor network used to build the universal calibration. Sensors impacted by sources that are not well-enough represented in the universal calibration (e.g. dust from deserts, ocean/sea-spray influence, combustion sources unique to a certain area such as dung or trash burning) may not calibrate well but more work is needed to explore this. Ultimately, the answer to this question is more based around the goals and limitations of a LCS deployment. As was previously discussed though, whatever the answers to these questions that end-users decide are, details of the model performance including correlation (R or R^2), bias (absolute and mean normalized), and accuracy (bias-corrected absolute and mean normalized MAE or RMSE) should also be reported so an idea of model performance can be conveyed.

In summary, low-cost PM sensors need to be collocated against each other (to ensure network precision) and against a reference method/eRM (or other instrument of choice) to ensure measurement accuracy. Calibration periods should ideally cover the entire range of meteorological conditions and PM concentration levels over which the sensors are expected to operate. Though collocation calibrations are only discussed here as a monolith calibration, it should be noted that there do exist different practical strategies to apply collocation-derived calibrations. More nuanced approaches can be applied such as event-specific corrections for different types of pollution episodes in a large network of sensors (Kelly et al., 2021). Centralized collocation measurements can even be “distributed” to futher sensors through nearest-neighbor calibrations to study the spatial variations in LCS measurements (Chu et al., 2020).

5. Moving forward with low-cost PM sensors

Low-cost PM sensors are already altering the landscape of who, how, where, and when PM measurements are made, and who has access to the data. Though obtaining high-quality data from these sensors presents a unique set of challenges, these challenges are surmountable. Ultimately, low-cost PM sensor users must be as open and transparent as possible with which sensors they use, how the calibrations were performed, and under what conditions the calibrations should apply. As long as sensor users perform collocation calibrations that span the entire range of expected operating conditions (RH, T, PM concentrations), report the specific correction factors (equations) obtained from these collocation studies, and appropriate descriptive metrics for their correction factors (correlation, accuracy, and bias on separate testing data) then trust can be established that low-cost PM sensors are reporting high-quality data. Low-cost PM sensors have already yielded valuable insights into source contributions of aerosols, even for small changes. For example, Eilenberg et al. (2020) showed that a large metallurgical coke plant near Pittsburgh, PA was only responsible for 0.3 ± 0.2 μg m^{-3} $PM_{2.5}$ over the long-term. Tanzer-Gruener et al. (2020) showed COVID-19 related lockdowns reduced $PM_{2.5}$ by up to 1.5 μg m^{-3} during morning rush hours, also in Pittsburgh, PA. Similarly, Chadwick et al. (2021) showed statistically significant reductions in PM due to COVID-19 in Salt Lake City County, Utah. Mallia et al. (2020) used LCS to validate wildfire smoke plume modeling. Stavroulas et al. (2020) used the measurements of network of PM sensors to quantify the impact of a nearby wildfire in different areas of Athens during the summer of 2019. Similarly, Stampfer et al. (2020) paired LCS with an aetholometer to understand how different sources of $PM_{2.5}$ and black carbon impact the Yakama Reservation airshed. Holder et al. (2020) showed that low-cost sensors can be used to fill in spatial gaps in air quality monitoring networks near wildfires with MAE less than 10 μg m^{-3} in hourly $PM_{2.5}$ concentrations. Subramanian et al. (2020) used low-cost sensors corrected with the methods used in this review to conduct the first long-term (over 1 year) measurements of air pollution in Kigali, Rwanda. The study found that annual average $PM_{2.5}$ was 52 μg/m^3 (well above the WHO first interim target of 35 μg/m^3), with local sources contributing half the ambient $PM_{2.5}$ in wet seasons but regional transported pollution likely dominant during the dry seasons. LCS calibrated with methods as described in this review were used to make the first observations of $PM_{2.5}$ in Kinshasa, DRC, and Brazzaville, ROC, which have a combined population of around 17 million. Findings suggest that annual mean $PM_{2.5}$ is about 4 times the WHO guidelines of 10 μg/m^3 (Air quality guidelines, 2017). Barkjohn et al. (2020a, 2020b) and others have shown how LCS can be used to quantify the effects of air filtration on indoor $PM_{2.5}$ exposure. Thus, low-cost PM sensors provide unique and useful insights into spatio-temporal impacts and sources of PM even if they require additional efforts to generate high-quality data.

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.

Acknowledgements

“This work benefited from State assistance managed by the National Research Agency under the “Programme d’Investissements d’Avenir” under the reference “ANR-18-MPGA-0011” (“Make Our Planet Great Again” initiative). CM was supported by an appointment to the NASA Postdoctoral Program at the Goddard Space Flight Center, administered by USRA through a contract with NASA. VFM acknowledges the Earth Institute of Columbia University. DMW acknowledges the National Science Foundation Accelnet Program (Award OISE 2020677) and a Columbia University Climate and Life Fellowship for support. SNP acknowledges the support by the project “Smart Air Quality Monitoring (SmartAQM)", funded by Western Greece’s Smart Specialisation Strategy (RIS3) 2014–2020, co-financed by Greece and the European Union. AAP acknowledges funding under Assistance Agreements no. RD83587301 and 83628601 awarded by the U.S. Environmental Protection Agency (EPA)."

References

Air quality guidelines. Global update. (2005). Particulate matter, ozone, nitrogen dioxide and sulfur dioxide, 2017 https://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/pre2009/air-quality-guidelines.-global-update-2005.-particulate-matter,-ozone,-nitrogen-dioxide-and-sulfur-dioxide.

Alexander, D. L. J., Tropsha, A., & Winkler, D. A. (2015). Beware of R2: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling, 55*(7), 1316–1322. https://doi.org/10.1021/acs.jcim.5b00206

Alpaydin, E. (2020). *Introduction to machine learning* (4th E). MIT press.

Ardon-Dryer, K., Dryer, Y., Williams, J. N., & Moghimi, N. (2020). Measurements of PM2.5 with PurpleAir under atmospheric conditions. *Atmospheric Measurement Techniques, 13*(10), 5441–5458. https://doi.org/10.5194/amt-13-5441-2020

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. https://doi.org/10.1371/journal.pone.0137789

Badura, M., Batog, P., Drzeniecka-Osiadacz, A., & Modzel, P. (2018). Evaluation of low-cost sensors for ambient PM2.5 monitoring. *Journal of Sensors*. , Article 5096540. https://doi.org/10.1155/2018/5096540, 2018.

Barkjohn, K. K., Bergin, M. H., Norris, C., Schauer, J. J., Zhang, Y., Black, M., Hu, M., & Zhang, J. (2020a). Using low-cost sensors to quantify the effects of air filtration on indoor and personal exposure relevant PM2.5 concentrations in beijing, China. *Aerosol and Air Quality Research, 20*(2), 297–313. https://doi.org/10.4209/aaqr.2018.11.0394

Barkjohn, K. K., Gantt, B., & Clements, A. (2020c). Development and Application of a United States wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. *Atmospheric Measurement Techniques Discussions*, 1–34. https://doi.org/10.5194/amt-2020-413

Barkjohn, K. K., Norris, C., Cui, X., Fang, L., Zheng, T., Schauer, J. J., Li, Z., Zhang, Y., Black, M., Zhang, J., Jim), & Bergin, M. H. (2020b). Real-time measurements of PM 2.5 and ozone to assess the effectiveness of residential indoor air filtration in Shanghai homes. *Indoor Air, 31*(1), 74–87. https://doi.org/10.1111/ina.12716

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis, 120*, 70–83. https://doi.org/10.1016/j.csda.2017.11.003

Bohren, Craig F., & Huffman, Donald R. (1998). *Absorption and Scattering of Light by Small Particles*. Wiley. In press.

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). https://doi.org/10.1038/s41598-019-43716-3

Bulot, F. M. J., Russell, H. S., Rezaei, M., Johnson, M. S., Ossont, S. J. J., Morris, A. K. R., & Cox, S. J. (2020). Laboratory comparison of low-cost particulate matter sensors to measure transient events of pollution. *Sensors, 20*(8). https://doi.org/10.3390/s20082219

Carlton, A. G., & Teitz, A. (2002). Design of a cost-effective weighing facility for PM2.5 quality assurance. *Journal of the Air and Waste Management Association, 52*(5), 506–510. https://doi.org/10.1080/10473289.2002.10470802

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., & Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International, 99*, 293–302. https://doi.org/10.1016/j.envint.2016.12.007

Chadwick, E., Le, K., Pei, Z., Sayahi, T., Rapp, C., Butterfield, A. E., & Kelly, K. E. (2021). Technical note: Understanding the effect of COVID-19 on particle pollution using a low-cost sensor network. *Journal of Aerosol Science, 155*, 105766. https://doi.org/10.1016/j.jaerosci.2021.105766

Chen, C. C., Kuo, C. T., Chen, S. Y., Lin, C. H., Chue, J. J., Hsieh, Y. J., … Huang, C. M. (2018). Calibration of low-cost particle sensors by using machine-learning method. 2018 IEEE Asia pacific conference on circuits and systems. *Proceedings of the 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 111–114. https://doi.org/10.1109/APCCAS.2018.8605619

Chu, H. J., Ali, M. Z., & He, Y. C. (2020). Spatial calibration and PM2.5 mapping of low-cost air quality sensors. *Scientific Reports, 10*(1), 22079. https://doi.org/10.1038/s41598-020-79064-w

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. https://doi.org/10.5194/amt-11-709-2018

Crilley, L. R., Singh, A., Kramer, L. J., Shaw, M. D., Alam, M. S., Apte, J. S., Bloss, W. J., Hildebrandt Ruiz, L., Fu, P., Fu, W., Gani, S., Gatari, M., Ilyinskaya, E., Lewis, A. C., Ng'ang'a, D., Sun, Y., Whitty, R. C. W., … Pope, F. D. (2020). Effect of aerosol composition on the performance of low-cost optical particle counter correction factors. *Atmospheric Measurement Techniques, 13*(3), 1181–1193. https://doi.org/10.5194/amt-13-1181-2020

Demanega, I., Mujan, I., Singer, B. C., Anđelković, A. S., Babich, F., & Licina, D. (2021). Performance assessment of low-cost environmental monitors and single sensors under variable indoor air quality and thermal conditions. *Building and Environment, 187*, 107415. https://doi.org/10.1016/j.buildenv.2020.107415

Di Antonio, A., Popoola, O., Ouyang, B., Saffell, J., & Jones, R. (2018). Developing a relative humidity correction for low-cost sensors measuring ambient particulate matter. *Sensors, 18*(9), 2790. https://doi.org/10.3390/s18092790

Digital universal particle concentration sensor PMS5003 series data manual.(2016).

Duvall, R. M., Clements, A. L., Hagler, G., Kamal, A., Kilaru, V., Goodman, L., Frederick, S., Barkjohn, K. K., VonWald, I., & Greene, D. (2021). Performance testing protocols, metrics, and target values for fine particulate matter air SensorsUse in ambient, outdoor, fixed site, non-regulatory supplemental and informational monitoring applications. EPA/600/R-20/280. Available at: https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=542106&Lab=CEMM.

Eilenberg, S. R., Subramanian, R., Malings, C., Hauryliuk, A., Presto, A. A., & Robinson, A. L. (2020). Using a network of lower-cost monitors to identify the influence of modifiable factors driving spatial patterns in fine particulate matter concentrations in an urban environment. *Journal of Exposure Science and Environmental Epidemiology, 30*(6), 949–961. https://doi.org/10.1038/s41370-020-0255-x

Feenstra, B., Papapostolou, V., Hasheminassab, S., Zhang, H., Boghossian, B. Der, Cocker, D., & Polidori, A. (2019). Performance evaluation of twelve low-cost PM2.5 sensors at an ambient air monitoring site. *Atmospheric Environment, 216*, 116946. https://doi.org/10.1016/j.atmosenv.2019.116946

Feinberg, S. N., Williams, R., Hagler, G., Low, J., Smith, L., Brown, R., Garver, D., Davis, M., Morton, M., Schaefer, J., & Campbell, J. (2019). Examining spatiotemporal variability of urban particulate matter and application of high-time resolution data from a network of low-cost air pollution sensors. *Atmospheric Environment, 213*, 579–584. https://doi.org/10.1016/j.atmosenv.2019.06.026

Fishbain, B., Lerner, U., Castell, N., Cole-Hunter, T., Popoola, O., Broday, D. M., Martinez Iñiguez, T., Nieuwenhuijsen, M., Jovasevic-Stojanovic, M., Topalovic, D., Jones, R. L., Galea, K. S., Golumbic, Y. N., Golumbic, Y. N., Baram-Tsabari, A., Yacobi, T., Drahler, D., Robinson, J. A., Kocman, D., & Bartonova, A. (2017). An evaluation tool kit of air quality micro-sensing units. *The Science of the Total Environment, 575*, 639–648. https://doi.org/10.1016/j.scitotenv.2016.09.061

Ford, B., Pierce, J. R., Wendt, E., Long, M., Jathar, S., Mehaffy, J., Tryner, J., Quinn, C., van Zyl, L., L'Orange, C., Miller-Lionberg, D., & Volckens, J. (2019). A low-cost monitor for measurement of fine particulate matter and aerosol optical depth – Part 2: Citizen-science pilot campaign in northern Colorado. *Atmospheric Measurement Techniques, 12*(12), 6385–6399. https://doi.org/10.5194/amt-12-6385-2019

Friedlander, S. K. (2000). *Smoke, Dust, and Haze* (2nd). Oxford University Press.

Gao, M., Cao, J., & Seto, E. (2015). A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *Environmental Pollution, 199*, 56–65. https://doi.org/10.1016/j.envpol.2015.01.013

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. https://doi.org/10.5194/amt-13-6343-2020

Hagler, G. S. W., Williams, R., Papapostolou, V., & Polidori, A. (2018). Air quality sensors and data adjustment algorithms: When is it No longer a measurement? *Environmental Science and Technology, 52*(10), 5530–5531. https://doi.org/10.1021/acs.est.8b01826. American Chemical Society.

Han, I., Symanski, E., & Stock, T. H. (2017). Feasibility of using low-cost portable particle monitors for measurement of fine and coarse particulate matter in urban ambient air. *Journal of the Air & Waste Management Association, 67*(3), 330–340. https://doi.org/10.1080/10962247.2016.1241195

Hapidin, D. A., Saputra, C., Maulana, D. S., Munir, M. M., & Khairurrijal, K. (2019). Aerosol chamber characterization for commercial particulate matter (PM) sensor evaluation. *Aerosol and Air Quality Research, 19*(1), 181–194. https://doi.org/10.4209/aaqr.2017.12.0611

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *Elements of statistical learning: Data mining, Inference, and prediction* (2nd ed.). Springer.

He, M., Kuerbanjiang, N., & Dhaniyala, S. (2020). Performance characteristics of the low-cost Plantower PMS optical sensor. *Aerosol Science and Technology, 54*(2), 232–241. https://doi.org/10.1080/02786826.2019.1696015

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. *Sensors, 20*(17), 4796. https://doi.org/10.3390/s20174796

Holstius, D. M., Pillarisetti, A., Smith, K. R., & Seto, E. (2014). Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. *Atmospheric Measurement Techniques, 7*(4), 1121–1131. https://doi.org/10.5194/amt-7-1121-2014

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. https://doi.org/10.5194/amt-11-4883-2018

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M., Rickard, J., Davis, M., Weinstock, L., Zimmer-Dauphinee, S., & Buckley, K. (2016). Community air sensor network (CAIRSENSE) project: Evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmospheric Measurement Techniques, 9*(11), 5281–5292. https://doi.org/10.5194/amt-9-5281-2016

Johnson, K. K., Bergin, M. H., Russell, A. G., & Hagler, G. S. W. (2018). Field test of several low-cost particulate matter sensors in high and low concentration urban environments. *Aerosol and Air Quality Research, 18*(3), 565–578. https://doi.org/10.4209/aaqr.2017.10.0418

Johnson, N. E., Bonczak, B., & Kontokosta, C. E. (2018). Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment. *Atmospheric Environment, 184*, 9–16. https://doi.org/10.1016/j.atmosenv.2018.04.019

Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., Redon, N., Crunaire, S., & Borowiak, A. (2019). Review of the performance of low-cost sensors for air quality monitoring. *Atmosphere, 10*(9), 506. https://doi.org/10.3390/atmos10090506

Kelly, K. E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., & Butterfield, A. (2017). Ambient and laboratory evaluation of a low-cost particulate matter sensor. *Environmental Pollution, 221*, 491–500. https://doi.org/10.1016/j.envpol.2016.12.039

Kelly, K. E., Xing, W. W., Sayahi, T., Mitchell, L., Becnel, T., Gaillardon, P. E., Meyer, M., & Whitaker, R. T. (2021). Community-based measurements reveal unseen differences during air pollution episodes. *Environmental Science and Technology, 55*(1), 120–128. https://doi.org/10.1021/acs.est.0c02341

Kosmopoulos, G., Salamalikis, V., Pandis, S. N., Yannopoulos, P., Bloutsos, A. A., & Kazantzidis, A. (2020). Low-cost sensors for measuring airborne particulate matter: Field evaluation and calibration at a South-Eastern European site. *The Science of the Total Environment, 748*, 141396. https://doi.org/10.1016/j.scitotenv.2020.141396

Kuula, J., Mäkelä, T., Aurela, M., Teinilä, K., Varjonen, S., González, Ó., & Timonen, H. (2020). Laboratory evaluation of particle-size selectivity of optical low-cost particulate matter sensors. *Atmospheric Measurement Techniques, 13*(5), 2413–2423. https://doi.org/10.5194/amt-13-2413-2020

Li, J., Li, H., Ma, Y., Wang, Y., Abokifa, A. A., Lu, C., & Biswas, P. (2018). Spatiotemporal distribution of indoor particulate matter concentration with a low-cost sensor network. *Building and Environment, 127*, 138–147. https://doi.org/10.1016/j.buildenv.2017.11.001

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. https://doi.org/10.4209/aaqr.2018.12.0485

Liu, H. Y., Schneider, P., Haugen, R., & Vogt, M. (2019). Performance assessment of a low-cost PM 2.5 sensor for a near four-month period in Oslo, Norway. *Atmosphere, 10*(2). https://doi.org/10.3390/atmos10020041

Liu, D., Zhang, Q., Jiang, J., & Chen, D. R. (2017). Performance calibration of low-cost and portable particular matter (PM) sensors. *Journal of Aerosol Science, 112*, 1–10. https://doi.org/10.1016/j.jaerosci.2017.05.011

Magi, B. I., Cupini, C., Francis, J., Green, M., & Hauser, C. (2020). Evaluation of PM2.5 measured in an urban setting using a low-cost optical particle counter and a Federal Equivalent Method Beta Attenuation Monitor. *Aerosol Science and Technology, 54*(2), 147–159. https://doi.org/10.1080/02786826.2019.1619915

Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S. P. N., Zimmerman, N., Kara, L. B., … Subramanian, R. (2019). Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. *Atmos. Meas. Tech., 12*, 903–920. https://doi.org/10.5194/amt-12-903-2019

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. https://doi.org/10.1080/02786826.2019.1623863

Mallia, D. V., Kochanski, A. K., Kelly, K. E., Whitaker, R., Xing, W., Mitchell, L. E., Jacques, A., Farguell, A., Mandel, J., Gaillardon, P., Becnel, T., & Krueger, S. K. (2020). Evaluating wildfire smoke transport within a coupled fire-atmosphere model using a high-density observation network for an episodic smoke event along Utah's wasatch front. *Journal of Geophysical Research: Atmosphere, 125*(20). https://doi.org/10.1029/2020JD032712. e2020JD032712.

Masic, A., Bibic, D., Pikula, B., Blazevic, A., Huremovic, J., & Zero, S. (2020). Evaluation of optical particulate matter sensors under realistic conditions of strong and mild urban pollution. *Atmospheric Measurement Techniques Discussions*, 1–22. https://doi.org/10.5194/amt-2020-237

McFarlane, C, Isevulambire, P. K., Lumbuenamo, R. S., Ndinga, A. M. E., Dhammapala, R., Jin, X., … Westervelt, D. M. (2021). First Measurements of Ambient $PM_{2.5}$ in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo Using Field-calibrated Low-cost Sensors. *Aerosol Air Qual. Res., 21*(7). https://doi.org/10.4209/aaqr.200619

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. W., Jayaratne, R., Kumar, P., Lau, A. K. H., Louie, P. K. 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? *Environment International, 116*, 286–299. https://doi.org/10.1016/j.envint.2018.04.018. Elsevier Ltd.

Noble, C. A., Vanderpool, R. W., Peters, T. M., McElroy, F. F., Gemmill, D. B., & Wiener, R. W. (2001). Federal reference and equivalent methods for measuring fine particulate matter. *Aerosol Science and Technology, 34*(5), 457–464. https://doi.org/10.1080/02786820121582

Paprotny, I., Doering, F., Solomon, P. A., White, R. M., & Gundel, L. A. (2013). Microfabricated air-microfluidic sensor for personal monitoring of airborne particulate matter: Design, fabrication, and experimental results. *Sensors and Actuators, A: Physical, 201*, 506–516. https://doi.org/10.1016/j.sna.2012.12.026

Pope, F. D., Gatari, M., Ng'ang'a, D., Poynter, A., & Blake, Rhiannon (2018). Airborne particulate matter monitoring in Kenya using calibrated low-cost sensors. *Atmos. Chem. Phys., 18*, 15403–15418. https://doi.org/10.5194/acp-18-15403-2018

Salimifard, P., Rim, D., & Freihaut, J. D. (2020). Evaluation of low-cost optical particle counters for monitoring individual indoor aerosol sources. *Aerosol Science and Technology, 54*(2), 217–231. https://doi.org/10.1080/02786826.2019.1697423

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. https://doi.org/10.1016/j.envpol.2018.11.065

Sousan, S., Koehler, K., Thomas, G., Park, J. H., Hillman, M., Halterman, A., & Peters, T. M. (2016). Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols. *Aerosol Science and Technology, 50*(5), 462–473. https://doi.org/10.1080/02786826.2016.1162901

Stampfer, O., Austin, E., Ganuelas, T., Fiander, T., Seto, E., & Karr, C. J. (2020). Use of low-cost PM monitors and a multi-wavelength aethalometer to characterize PM2.5 in the Yakama Nation reservation. *Atmospheric Environment, 224*, 117292. https://doi.org/10.1016/j.atmosenv.2020.117292

Stavroulas, I., Grivas, G., Michalopoulos, P., Liakakou, E., Bougiatioti, A., Kalkavouras, P., Fameli, K., Hatzianastassiou, N., Mihalopoulos, N., & Gerasopoulos, E. (2020). Field evaluation of low-cost PM sensors (purple air PA-II) under variable urban air quality conditions, in Greece. *Atmosphere, 11*(9), 926. https://doi.org/10.3390/atmos11090926

Steinle, S., Reis, S., Sabel, C. E., Semple, S., Twigg, M. M., Braban, C. F., Leeson, S. R., Heal, M. R., Harrison, D., Lin, C., & Wu, H. (2015). Personal exposure monitoring of PM2.5 in indoor and outdoor microenvironments. *The Science of the Total Environment, 508*, 383–394. https://doi.org/10.1016/j.scitotenv.2014.12.003

Subramanian, R., Ellis, A., Torres-Delgado, E., Tanzer, R., Malings, C., Rivera, F., … Mayol-Bracero, O. L. (2018). Air Quality in Puerto Rico in the Aftermath of Hurricane Maria: A Case Study on the Use of Lower Cost Air Quality Monitors. *ACS Earth Space Chem., 2*(11), 1179–1186. https://doi.org/10.1021/acsearthspacechem.8b00079

Subramanian, R., Kagabo, A. S., Baharane, V., Guhirwa, S., Sindayigaya, C., Malings, C., Williams, N. J., Kalisa, E., Li, H., Adams, P., Robinson, A. L., DeWitt, H. L., Gasore, J., & Jaramillo, P. (2020). Air pollution in Kigali, Rwanda: Spatial and temporal variability, source contributions, and the impact of car-free sundays. *Clean Air Journal, 30*(2). https://doi.org/10.17159/caj/2020/30/2.8023

Tanzer-Gruener, R., Li, J., Eilenberg, S. R., Robinson, A. L., & Presto, A. A. (2020). *Impacts of Modifiable factors on ambient air pollution: A case study of COVID-19 Shutdowns*. Environmental Science & Technology Letters. https://doi.org/10.1021/acs.estlett.0c00365

Tryner, J., Good, N., Wilson, A., Clark, M. L., Peel, J. L., & Volckens, J. (2019a). Variation in gravimetric correction factors for nephelometer-derived estimates of personal exposure to PM2.5. *Environmental Pollution, 250*, 251–261. https://doi.org/10.1016/j.envpol.2019.03.121

Tryner, J., L'Orange, C., Mehaffy, J., Miller-Lionberg, D., Hofstetter, J. C., Wilson, A., & Volckens, J. (2020b). Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers. *Atmospheric Environment, 220*, 117067. https://doi.org/10.1016/j.atmosenv.2019.117067

Tryner, J., Mehaffy, J., Miller-Lionberg, D., & Volckens, J. (2020a). Effects of aerosol type and simulated aging on performance of low-cost PM sensors. *Journal of Aerosol Science, 703*, 105654. https://doi.org/10.1016/j.jaerosci.2020.105654

Tryner, J., Quinn, C., Windom, B. C., & Volckens, J. (2019b). Design and evaluation of a portable PM2.5 monitor featuring a low-cost sensor in line with an active filter sampler. *Environmental Sciences: Processes and Impacts, 21*(8), 1403–1415. https://doi.org/10.1039/c9em00234k

Wallace, L., Bi, J., Ott, W. R., Sarnat, J., & Liu, Y. (2021). Calibration of low-cost PurpleAir outdoor monitors using an improved method of calculating $PM_{2.5}$. *Atmospheric Environment, 256*(7), 118432. https://doi.org/10.1016/j.atmosenv.2021.118432

Walser, A., Sauer, D., Spanu, A., Gasteiger, J., & Weinzierl, B. (2017). On the parametrization of optical particle counter response including instrument-induced broadening of size spectra and a self-consistent evaluation of calibration measurements. *Atmospheric Measurement Techniques, 10*(11), 4341–4361. https://doi.org/10.5194/amt-10-4341-2017

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

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

Wendt, E. A., Quinn, C. W., Miller-Lionberg, D. D., Tryner, J., L'Orange, C., Ford, B., Yalin, A. P., Jeffrey, P., Jathar, S., & Volckens, J. (2019). A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth - Part 1: Specifications and testing. *Atmospheric Measurement Techniques, 12*(10), 5431–5441. https://doi.org/10.5194/amt-12-5431-2019

Williams, R., Duvall, R., Kilaru, V., Hagler, G., Hassinger, L., Benedict, K., Rice, J., Kaufman, A., Judge, R., Pierce, G., Allen, G., Bergin, M., Cohen, R. C., Fransioli, P., Gerboles, M., Habre, R., Hannigan, M., Jack, D., Louie, P., … Ning, Z. (2019). Deliberating performance targets workshop: Potential paths for emerging PM2.5 and O3 air sensor progress. *Atmospheric Environment X, 2*, 100031. https://doi.org/10.1016/j.aeaoa.2019.100031, 2019.

Williams, R., Kaufman, A., Hanley, T., Rice, J., & Garvey, S. (2014). In *Evaluation of field-deployed low cost PM sensors*. U. S. Environmental Protection Agency (EPA). Vol. EPA/600/R- (Issue Washington DC), EPA/600/R-14/464 (NTIS PB 2015-102104).
Zamora, M. L., Xiong, F., Gentner, D., Kerkez, B., Kohrman-Glaser, J., & Koehler, K. (2019). Field and laboratory evaluations of the low-cost plantower particulate matter sensor. *Environmental Science and Technology, 53*(2), 838–849. https://doi.org/10.1021/acs.est.8b05174
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. https://doi.org/10.5194/amt-11-4823-2018
Zikova, N., Hopke, P. K., & Ferro, A. R. (2017). Evaluation of new low-cost particle monitors for PM2.5 concentrations measurements. *Journal of Aerosol Science, 105*, 24–34. https://doi.org/10.1016/j.jaerosci.2016.11.010
Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., & Robinson, A. L. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques, 11*(1), 291–313. https://doi.org/10.5194/amt-11-291-2018
Zusman, M., Schumacher, C. S., Gassett, A. J., Spalt, E. W., Austin, E., Larson, T. V., Carvlin, G., Seto, E., Kaufman, J. D., & Sheppard, L. (2020). Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environment International, 134*, 105329. https://doi.org/10.1016/j.envint.2019.105329