Environmental Science & Technology

pubs.acs.org/est

Article

Incorporating Low-Cost Sensor Measurements into High-Resolution $PM_{2.5}$ Modeling at a Large Spatial Scale

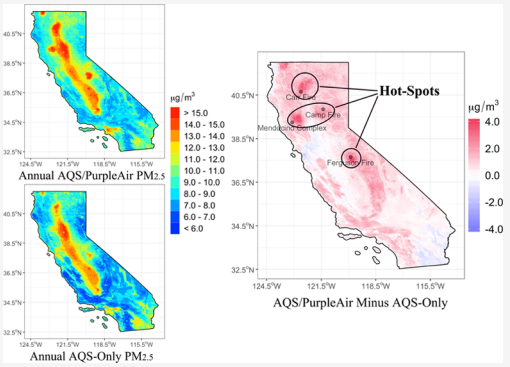
Jianzhao Bi, Avani Wildani, Howard H. Chang, and Yang Liu*

Cite This: *Environ. Sci. Technol.* 2020, 54, 2152−2162

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Low-cost air quality sensors are promising supplements to regulatory monitors for $PM_{2.5}$ exposure assessment. However, little has been done to incorporate the low-cost sensor measurements in large-scale $PM_{2.5}$ exposure modeling. We conducted spatially varying calibration and developed a downweighting strategy to optimize the use of low-cost sensor data in $PM_{2.5}$ estimation. In California, PurpleAir low-cost sensors were paired with air quality system (AQS) regulatory stations, and calibration of the sensors was performed by geographically weighted regression. The calibrated PurpleAir measurements were then given lower weights according to their residual errors and fused with AQS measurements into a random forest model to generate 1 km daily $PM_{2.5}$ estimates. The calibration reduced PurpleAir's systematic bias to ~0 $\mu g/m^3$ and residual errors by 36%. Increased sensor bias was found to be associated with higher temperature and humidity, as well as longer operating time. The weighted prediction model outperformed the AQS-based prediction model with an improved random cross-validation (CV) R^2 of 0.86, an improved spatial CV R^2 of 0.81, and a lower prediction error. The temporal CV R^2 did not improve due to the temporal discontinuity of PurpleAir. The inclusion of PurpleAir data allowed the predictions to better reflect $PM_{2.5}$ spatial details and hotspots.



1. INTRODUCTION

Particulate matter with aerodynamic diameter ≤ 2.5 μm ($PM_{2.5}$) is associated with a broad range of adverse health outcomes[1,2] and is a major contributor to the global burden of disease.[3] Precise and detailed ambient $PM_{2.5}$ exposure assessment is fundamental to reliably describing $PM_{2.5}$−disease relationships[4−6] and developing $PM_{2.5}$ pollution control policies.[7,8] Ambient $PM_{2.5}$ exposure assessment has traditionally relied on regulatory air quality monitoring stations such as the U.S. Environmental Protection Agency (EPA) air quality system (AQS) stations. Due to high instrumentation and maintenance cost, regulatory monitoring is only performed at limited locations for examining the compliance of air quality standards. Given the spatial variability of $PM_{2.5}$ at the kilometer scale,[9] sparse and uneven regulatory monitoring has a limited ability to reflect $PM_{2.5}$ pollution details,[10] especially at remote communities or when impacted by episodic events such as wildfires.[11,12] This paradigm is shifting with the development of citizen science where many individuals voluntarily collect large amounts of air quality data through low-cost air quality sensors. These low-cost sensors typically cost < $2500 and have desirable features such as flexibility of deployment and ease of maintenance. Due to the lower costs, they can be deployed more densely than government-operated regulatory stations. Low-cost sensor data have the potential to provide meaningful air quality information in a spatiotemporally more frequent manner.

Since the majority of low-cost $PM_{2.5}$ sensors are based on the light-scattering principle,[13] they tend to have a higher uncertainty than reference-grade monitors. The uncertainty may be caused by the measurement principle itself such as the uncertainty in measured particle counts and the conversion from particle counts to mass concentrations.[14,15] The manufacturing calibration, which uses manufactured aerosols with different compositions and properties than those in the ambient environment, is another source of uncertainty.[13,16] The sensors may also experience quality degradation over time[17] and other logistical issues during deployment and maintenance. Therefore, the data quality of low-cost sensors varies with sampling locations and conditions.[13,14,18] Previous studies suggested that the pretest and calibration of low-cost sensors should be conducted where the sensors are intended to be deployed.[13,14,18] Current laboratory and field calibrations of low-cost sensors mainly focus on reducing their systematic bias.[19−21] Humidity and temperature were found to be two important factors affecting the systematic bias,[16,22] especially when humidity is high.[15,19,23] Multivariate regression models with these factors as covariates have been widely used to

Received: October 7, 2019
Revised: December 11, 2019
Accepted: January 6, 2020
Published: January 13, 2020

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

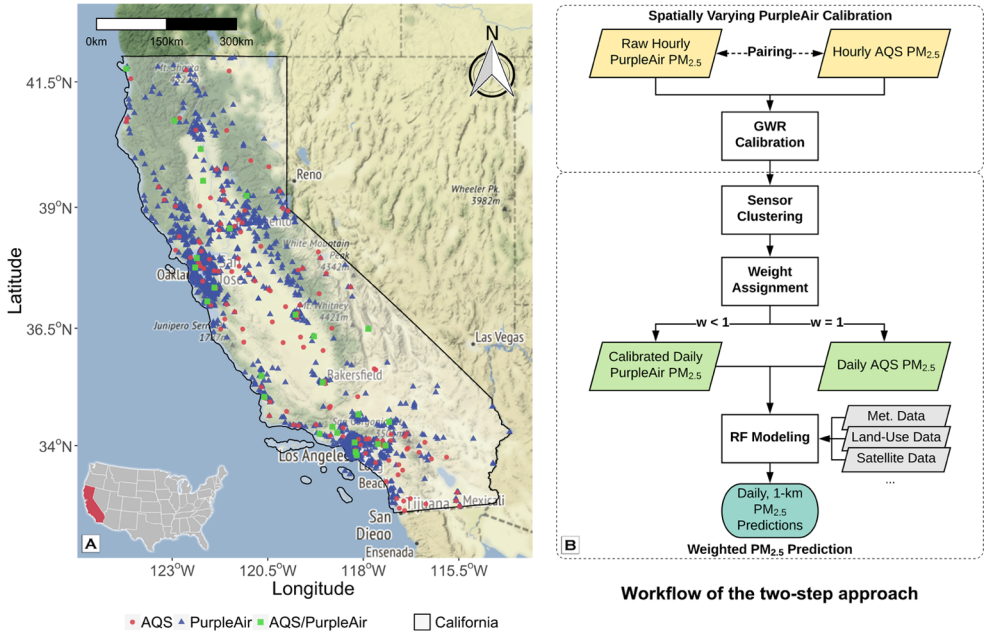


Figure 1. (A) Study domain, California, with the locations of AQS (red), PurpleAir (blue), and paired AQS/PurpleAir monitors (green). The latitude/longitude ranges of California are [32°30′N, 42°N] and [114°8′W, 124°24′W]. (B) Workflow of the two-step approach including spatially varying PurpleAir calibration and weighted $PM_{2.5}$ prediction.

calibrate the sensors against collocated reference-grade monitors, which are able to significantly improve the accuracy of the data but have a limited ability to reduce their residual errors.[19,24,25]

Currently, there are two primary uses for low-cost $PM_{2.5}$ sensors. First, they improve monitoring coverage in areas where there is insufficient regulatory monitoring. For example, Pope et al.[26] identified spatial features and diurnal behavior of PM pollution based on low-cost sensor data deployed at three locations in Nairobi, Kenya, a city without long-term reference-grade PM measurements. The applications of low-cost sensors in regions with limited access to regulatory monitoring have advanced local communities' awareness and understanding of air pollution.[11,13,27] The second major use of low-cost sensors is to assist with revealing the fine-scale variability of $PM_{2.5}$, especially in developed countries. In this case, low-cost sensor data are used as a supplement to regulatory measurements in physical or statistical models to fill in the spatiotemporal gaps of $PM_{2.5}$ concentrations. For example, Masiol et al.[28] incorporated continuous PM concentrations from commercially available low-cost sensors in land-use regression models to derive hourly resolved PM predictions in Monroe County of New York State. With the addition of calibrated low-cost sensor data to the few regulatory measurements from existing air quality stations in Imperial County of California, Bi et al.[29] found that low-cost sensor data could improve the accuracy of $PM_{2.5}$ predictions with more reasonable spatial details.

Though it holds promise, there are two major limitations with regard to using a low-cost sensor network to improve $PM_{2.5}$ pollution mapping and exposure assessment. First, due to the significant cost of extensive field testing by trained scientists,[26] the side-by-side low-cost sensor calibration against reference-grade monitors has mostly been confined in a small region, e.g., at a city or county level. In other words, even though low-cost sensors are individually cheap, the high cost of field calibration makes their use at large spatial scales expensive. Field calibration is more difficult for the low-cost sensor networks established by third parties for other purposes. Second, even though low-cost sensor data attain a relatively low systematic bias after calibration, their precision is still not comparable to that of reference-grade measurements. The residual measurement errors of sensor data are difficult to be reduced by current multivariate calibration models.[29] When the calibrated sensor data are treated as ground truth, their residual errors may still significantly influence the reliability of their downstream applications such as hotspot detection, source identification, and epidemiologic analysis.[28] These limitations also apply to other citizen science programs with large amounts of low-quality volunteer-generated data, such as the personal weather data collected by citizens across the U.S. for the Citizen Weather Observer Program.[10]

In this study, we proposed a two-step approach to address the aforementioned limitations and optimize the use of low-cost sensor measurements in a spatially extensive, high-resolution $PM_{2.5}$ exposure assessment. Using a commercial low-cost sensor network as an example, we first conducted a large-scale spatially varying calibration for low-cost $PM_{2.5}$ data against existing reference-grade measurements. A downweighting process was then conducted in the prediction stage to reduce the negative impacts of the residual errors of the calibrated sensor data. Our framework is designed to integrate low-cost sensor data with regulatory monitoring data and other sources of information such as satellite, meteorological, and land-use data to improve the high-resolution $PM_{2.5}$ exposure assessment. This framework could also be informative to other citizen science programs to improve the accuracy of volunteer-generated data.

2. DATA AND METHODS

2.1. Study Domain and Modeling Strategy. California is the most populous U.S. state with over 39 million residents and the one with the most severe PM pollution, especially in metropolitan areas and the Central Valley.[30] California has a relatively dense regulatory air quality monitoring network and

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

dense low-cost sensors for tracking local air quality. By the end of 2018, there were 157 AQS stations providing $PM_{2.5}$ measurements and 2090 outdoor sensors from PurpleAir, a commercial low-cost sensor network, providing subhourly $PM_{2.5}$ measurements within the state. Figure 1A shows our study domain with the locations of AQS and PurpleAir monitors. To take advantage of the dense ground monitors and high-resolution satellite aerosol data, we defined a grid at a 1 km resolution for $PM_{2.5}$ modeling. The entire study domain consists of 493 561 grid cells. A brief workflow of our two-step modeling approach is shown in Figure 1B.

2.2. Data. *2.2.1. $PM_{2.5}$ Measurements.* PurpleAir is a citizen-based, real-time low-cost PM sensor network started in 2015 (https://www.purpleair.com/). By the end of 2018, PurpleAir had almost 7000 sensors worldwide with a growing rate of ∼30 sensors per day. PurpleAir provides minute-level indoor/outdoor measurement for $PM_{2.5}$ and other environmental parameters (humidity, barometric pressure, and temperature). We obtained hourly $PM_{2.5}$ measurements from 2090 outdoor PurpleAir sensors in 2018 in California (N = 5 842 404). Quality control was conducted for these measurements to minimize the outliers (Section 1, Supporting Information). The raw PurpleAir $PM_{2.5}$ measurements appeared to bias substantially high against reference-grade measurements (Section 2, Supporting Information).

Reference-grade PM measurements were obtained from the EPA AQS regulatory monitoring network (https://www.epa.gov/aqs). In 2018, 157 AQS stations provided 50 870 daily $PM_{2.5}$ measurements in California, and 109 of them provided 499 940 hourly $PM_{2.5}$ measurements. The hourly $PM_{2.5}$ measurements from the AQS stations near the PurpleAir monitors were used for PurpleAir evaluation and calibration. Daily PM_{10} measurements were also obtained, which were utilized to generate an ancillary predictor, the $PM_{2.5}/PM_{10}$ ratio. This predictor is a continuous surface interpolated from the $PM_{2.5}/PM_{10}$ ratio scatters at the locations of AQS stations, representing the distribution of the percentages of $PM_{2.5}$ in PM_{10} in the study domain. The interpolation was performed by ordinary kriging with month-specific variograms fitted in a spherical model. The $PM_{2.5}/PM_{10}$ ratio was shown to be an important predictor of ground-level $PM_{2.5}$ in California due to relatively high coarse-particle loadings.[29]

2.2.2. Ancillary Data. Aerosol optical depth (AOD) is the integral of aerosol extinction of the solar beam along the entire vertical atmospheric column, which is an important predictor of ground-level $PM_{2.5}$.[9,31] We adopted the satellite AOD retrievals from the moderate resolution imaging spectroradiometer (MODIS) multiangle implementation of atmospheric correction (MAIAC) product (MCD19A2, https://lpdaac.usgs.gov/products/mcd19a2v006/).[32] MODIS aerosol retrievals have 40−50% missing values on average in California due to cloud cover.[33] Therefore, AOD gap-filling was performed by following Bi et al.,[9] in which daily-level AOD prediction models were built with satellite-observed cloud fractions and AOD-related meteorological parameters (humidity, visibility, downward shortwave radiation, and wind speed and direction) to derive complete daily AOD surfaces.

Meteorological parameters were obtained from the North American Regional Reanalysis (http://www.emc.ncep.noaa.gov/) at a 32 kilometer (∼0.3°) resolution[34] and the North American Land Data Assimilation System (https://ldas.gsfc.nasa.gov/) at a 0.125° resolution.[35] The meteorological parameters used in $PM_{2.5}$ modeling include visibility, 2 meter air temperature and specific humidity, planetary boundary layer height, 10 meter zonal and meridional wind speeds, shortwave/longwave radiation flux downwards, aerodynamic conductance, convective available potential energy, convective precipitation, and total precipitation. These reanalysis data were aggregated from subdaily to daily to match the $PM_{2.5}$ data.

The land-use and demographic parameters were obtained from (1) the 2011 National Land Cover Database at a 30 meter resolution (https://www.mrlc.gov/), (2) the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation at a 1 arc-second resolution (https://asterweb.jpl.nasa.gov/), (3) the LandScan ambient population in 2017 at a 900 meter resolution (https://landscan.ornl.gov/), (4) the Normalized Difference Vegetation Index (NDVI) from MODIS vegetation indices at a 500 meter resolution (https://modis.gsfc.nasa.gov/), (5) the distances to the nearest primary and secondary roads computed from the U.S. Census TIGER/Line Shapefiles (https://www.census.gov/), and (6) active fire distributions computed from satellite-retrieved active fire spots (https://firms.modaps.eosdis.nasa.gov).

2.3. PurpleAir $PM_{2.5}$ Calibration and Weighted $PM_{2.5}$ Modeling. *2.3.1. Spatially Varying PurpleAir $PM_{2.5}$ Calibration.* The PurpleAir measurements were calibrated against the "gold-standard" AQS measurements. Since both AQS and PurpleAir were existing networks, there were very few strictly collocated AQS/PurpleAir sites in California during the time this analysis was conducted. Instead, we matched a PurpleAir sensor to its nearest AQS station within a 500 m radius so that each AQS/PurpleAir pair was within a 1 km modeling grid cell. The calibration was conducted at the level of single PurpleAir sensors, i.e., the measurements from multiple PurpleAir sensors around the same AQS station were treated separately rather than aggregated together in calibration. A sensitivity analysis indicated that the selected AQS/PurpleAir pairs were robust within a range between 100 and 1000 m without a significant change in the number of pairs. During the study period, 54 PurpleAir sensors were matched to 26 AQS stations, providing 128 777 paired hourly $PM_{2.5}$ measurements.

Given the spatially varying agreement between paired AQS and PurpleAir measurements, geographically weighted regression (GWR) was conducted for the PurpleAir calibration. GWR allows smoothed local relationships between AQS and PurpleAir measurements. Temperature and relative humidity (RH) were used as covariates of the GWR calibration model because these parameters are associated with the data quality of low-cost sensors.[14,19] In addition, low-cost sensors may experience quality degradation over time;[17,36,37] thereby, the total operating time of a sensor (the duration between the measurement time and the installation time) was used to adjust the effect of sensor aging. Finally, the sensor uptime (the time during which a sensor is in consecutive operation from the last boot time) was used to adjust the potential impact of sensor's operational stability on data quality. A linear specification was used to describe the relationship between the bias of PurpleAir measurements and four covariates (temperature, RH, operating time, and uptime) (see Section 3 of the Supporting Information for the nonlinearity analysis of PurpleAir bias). The GWR model can be expressed as

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

$$\begin{aligned}\text{AQS PM}_{2.5i} &= \beta_0(u_i, v_i) + \beta_1(u_i, v_i)\cdot\text{PurpleAir PM}_{2.5i} \\ &+ \beta_2(u_i, v_i)\cdot T_i + \beta_3(u_i, v_i)\cdot\text{RH}_i \\ &+ \beta_4(u_i, v_i)\cdot\text{optime}_i + \beta_5(u_i, v_i)\cdot\text{uptime}_i + \epsilon_i \\ \epsilon_i &\sim N(0, \tau^2) \end{aligned} \quad (1)$$

where $\beta(u_i v_i)$ indicates the vector of the location-specific parameter estimates, and $(u_i v_i)$ represents the geographic coordinates of location i. AQS $PM_{2.5i}$ and PurpleAir $PM_{2.5i}$ are the paired hourly $PM_{2.5}$ measurements at location i. T_i, RH_i, $optime_i$, and $uptime_i$ represent temperature, relative humidity, operating time, and uptime of the PurpleAir sensor at location i, respectively. The error term ϵ_i is normally distributed with a mean of zero and an overall error variance τ^2. The optimal hyperparameters of GWR, i.e., the kernel and bandwidth, were chosen based on the corrected Akaike information criterion. In this analysis, the optimal kernel was a Gaussian kernel, and the optimal bandwidth was 5401 nearest-neighboring points. All covariates were statistically significant at an α level of 0.05 in the GWR calibration model. The GWR was fitted using the R package "GWmodel" version 2.0-7.[38] To examine the impact of the number of collocated AQS stations for PurpleAir calibration, a sensitivity analysis was conducted with subsets of randomly selected collocated stations (Table S1).

Besides the calibration, another generalized additive model (GAM) was built to quantify the impacts of temperature, RH, operating time, and uptime on the bias of PurpleAir measurements (eq 2). This model describes the relationships of the absolute bias of PurpleAir measurements (against paired AQS measurements) and the four covariates. The model can be expressed as

$$\begin{aligned}&|\text{AQS PM}_{2.5i} - \text{PurpleAir PM}_{2.5i}| \\ &= \alpha_0 + s(T_i) + s(\text{RH}_i) + s(\text{optime}_i) + \text{s}(\text{uptime}_i) + \epsilon_i \end{aligned} \quad (2)$$

where i represents a specific paired record, and s() indicates the smooth function with degrees of freedom of 2 to minimize the random fluctuation in the estimated relationships.

2.3.2. Weighted $PM_{2.5}$ Modeling with AQS and PurpleAir Data. After calibration, AQS and PurpleAir measurements were aggregated to daily, 1 km averages for $PM_{2.5}$ modeling. For the 1 km grid cells containing both AQS and PurpleAir measurements, only the AQS measurements were selected to better represent the pollution levels. A weighted random forest (RF) model was adopted to generate daily, 1 km $PM_{2.5}$ predictions based on the aggregated daily measurements. Random forests are an ensemble learning method combining the predictions from a multitude of decision trees.[39] RF provides variable importance measures to explain the relative importance and contribution of each predictor. The RF algorithm has been increasingly applied to predicting ground $PM_{2.5}$ levels.[9,40] An advantage of using RF in this analysis is that it can assign an individual weight to each dependent observation so that the high-quality AQS measurements could have a higher weight than the PurpleAir measurements.[41] An observation with a higher weight will be selected with a higher probability in the samples for building decision trees, therefore having a greater influence on the predictions.

We followed Hu et al.[40] and Bi et al.[9] to perform variable selection and model evaluation based on RF variable importance and random cross-validation (CV). The independent variables used in the prediction models are shown in Table 1. Two major RF hyperparameters, the number of decision trees (n_{tree}) and the number of predictors randomly tried at each split (m_{try}), were tuned based on CV performance. In this analysis, the optimal values of n_{tree} and m_{try} were 500 and 5, respectively. Apart from the RF model with individual weights (refer to as "the weighted model" hereinafter), two reference models were built: one based solely on the AQS measurements (a.k.a. the AQS-based model) and another based on the AQS and PurpleAir measurements without weighting (a.k.a. the nonweighted model). We used 10-fold random, spatial, and temporal CV to evaluate these models. The 10-fold spatial CV procedure creates validation sets according to the locations of the measurements (i.e., dropping 10% of all locations), and the temporal CV creates validation sets according to Julian days. R^2 and root-mean-square prediction error (RMSPE) were the major gauging metrics of CV. It is worth noting that CV was only performed on AQS measurements not used in calibrating PurpleAir to ensure the CV only evaluates out-of-sample model prediction performance. This avoids the issue that calibrated PurpleAir measurements will likely share similar features of matched AQS monitors.

Table 1. Independent Variables Used in the $PM_{2.5}$ Prediction Models (s: Spatially Varying; t: Temporally Varying)

prediction variables	
MAIAC AOD	**ancillary variables**
gap-filled Terra $AOD_{(s,t)}$	$PM_{2.5}/PM_{10}$ $ratio_{(s,t)}$
gap-filled Aqua $AOD_{(s,t)}$	$day\ of\ year_{(t)}$
	$month_{(t)}$
land-use variables	**meteorological variables**
$elevation_{(s)}$	$visibility_{(s,t)}$
$population_{(s)}$	2 meter air $temperature_{(s,t)}$
$NDVI_{(s,t)}$	2 meter specific $humidity_{(s,t)}$
nearest distance to $roads_{(s)}$	planetary boundary layer $height_{(s,t)}$
percentage of $shrublands_{(s)}$	longwave radiation flux $downwards_{(s,t)}$
percentage of $herbaceous\ areas_{(s)}$	shortwave radiation flux $downwards_{(s,t)}$
percentage of $developed\ areas_{(s)}$	10 meter zonal wind $speed_{(s,t)}$
percentage of $cultivated\ areas_{(s)}$	10 meter meridional wind $speed_{(s,t)}$
percentage of $forests_{(s)}$	$aerodynamic\ conductance_{(s,t)}$
percentage of water $bodies_{(s)}$	$convective\ available\ potential\ energy_{(s,t)}$
percentage of $wetlands_{(s)}$	$convective\ precipitation_{(s,t)}$
percentage of barren $lands_{(s)}$	total $precipitation_{(s,t)}$
active fire $distribution_{(s,t)}$	

Although the systematic bias of PurpleAir data could be reduced by calibration, these measurements still had substantial residual errors, which might adversely impact the accuracy of $PM_{2.5}$ predictions. We assigned lower weights to PurpleAir measurements in the prediction process according to their estimated residual errors to mitigate such influences. Similar to the bias, we assumed that the residual errors in calibrated PurpleAir measurements would vary under different environmental conditions. Accordingly, the study domain was partitioned into several subdomains based on selected variables using hierarchical agglomerative clustering (HAC).[42] The domain partitioning aimed to obtain distinct $PM_{2.5}$ pollution conditions under which the PurpleAir residual errors would vary. The selected variables were the top 10 predictors with the highest importance values in the AQS-based prediction model (Table S2). HAC performs "bottom-up" clustering, i.e., each

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

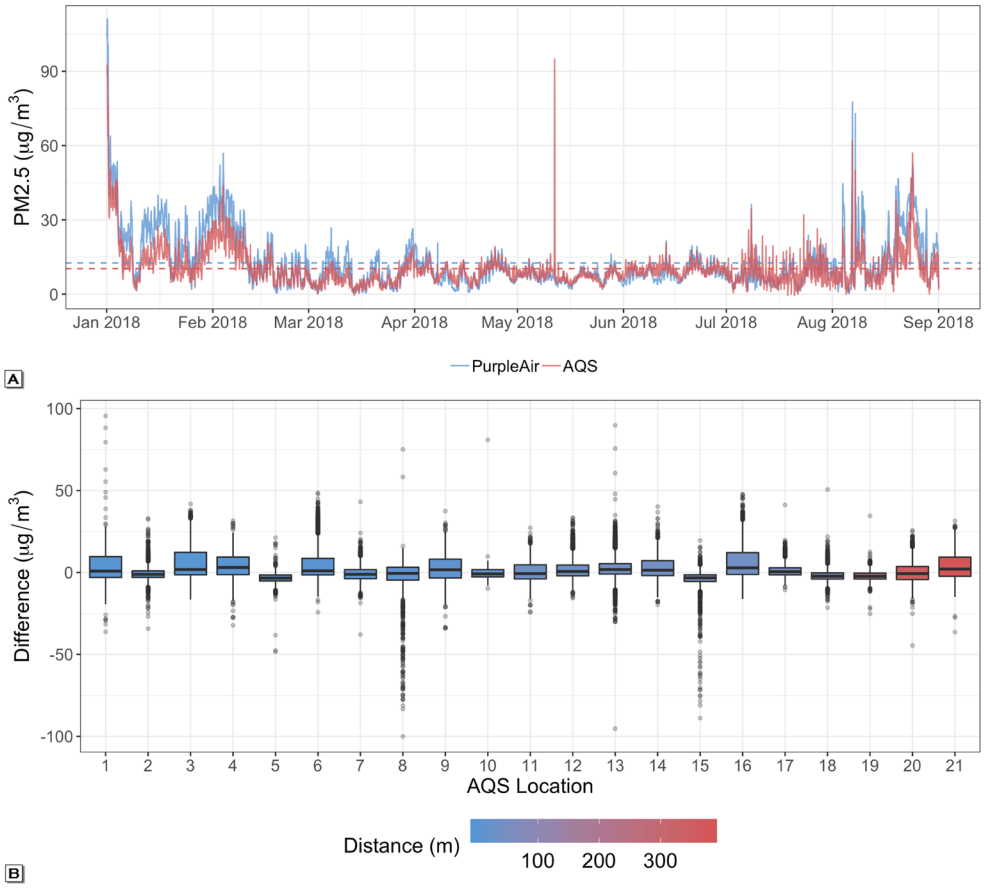


Figure 2. (A) Time series of paired AQS (red) and PurpleAir (blue) hourly measurements with their mean values (dashed lines). The paired measurements were only available from January to August of 2018. (B) Box plots of the differences between paired AQS and PurpleAir hourly measurements (PurpleAir minus AQS) at the locations of 21 AQS stations with more than 70 paired hourly $PM_{2.5}$ measurements (in the ascending order of mean distance).

unclassified item starts in its own cluster and the two most similar items are merged into a new cluster. This step is iterated until all items are aggregated into a single cluster to form a hierarchical structure. Hierarchical clustering has no hidden assumptions about the distribution of underlying data, which was suitable in our case as we had little a priori understanding of the 10-variable feature space. The number of clusters (K) was determined with the R package "NbClust".[43] NbClust provides 30 indices for determining K, and the optimal K can be decided through the majority vote of these indices. The optimal K of our feature space was determined to be 3. Month-specific clustering was conducted as a sensitivity test, which showed that our three-cluster partitioning was robust over time.

Weights assigned to calibrated PurpleAir measurements reflected their relative importance to AQS measurements in the $PM_{2.5}$ prediction process. All calibrated PurpleAir measurements in each subdomain were given the same weight determined using a three-parameter formula (eq 3). First, the mean PurpleAir residual variance in each subdomain (τ_j^2), measured by the variance of the differences between paired PurpleAir and AQS measurements, represents the overall PurpleAir residual error in the subdomain. Secondly, the error associated with the prediction model structure (σ^2) was estimated as the CV mean-squared prediction error (MSPE) of the AQS-based model. σ^2 was the same across different subdomains. The proportion of the model structure error (σ^2) in the total possible variance ($\sigma^2 + \tau_j^2$) served as the upper bound of the weight. Finally, in addition to the uncertainty related to the low-cost sensing technology, other factors such as the lack of a consistent siting plan for spatial representativeness can also influence the quality of PurpleAir measurements. To summarize the impact of these unquantifiable circumstances, we included a data-driven scale factor (ρ) with a range (0, 1) in the weighting formula. Its value was tuned based on CV RMSPE and was determined to be 0.23 in this analysis (Section 4, Supporting Information). Intuitively, as the overall residual error and unquantifiable uncertainty in calibrated PurpleAir measurements decrease, the weight of the PurpleAir measurements increases within the range (0, 1) in the prediction model. As a reference, the weight of the gold-standard AQS measurements was fixed to 1.

$$w_j = \rho \cdot \frac{\sigma^2}{\sigma^2 + \tau_j^2} \qquad 0 < w_j < 1 \quad 0 < \rho < 1 \tag{3}$$

3. RESULTS

3.1. PurpleAir $PM_{2.5}$ Calibration. *3.1.1. Evaluation of Uncalibrated PurpleAir Measurements.* A linear regression of uncalibrated PurpleAir measurements against AQS had an R^2 of 0.74 and a slope of 0.61. This R^2 was slightly lower than the R^2 values reported by previous studies.[44,45] Our relaxed pairing strategy between AQS and PurpleAir might lead to this lower agreement. Figure 2A shows that uncalibrated PurpleAir $PM_{2.5}$

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

measurements tracked well with AQS in time but detected more spikes and biased high against AQS by 1.9 $\mu g/m^3$. These spikes might be caused by high-level local pollution in the microenvironments near the PurpleAir sensors such as cigarette smoke, barbecues, fireplaces, and idling trucks[25] as most PurpleAir sensors have been installed in residential areas by citizens.

When evaluating PurpleAir sensors at the locations of 21 AQS stations with more than 70 paired $PM_{2.5}$ measurements, a clear variation of the AQS/PurpleAir agreement was observed. Figure 2B shows the box plots of the differences between PurpleAir and AQS data at each AQS site. The site-specific R^2 between PurpleAir and AQS ranged from 0.03 to 0.93. The site-specific slope also had a large variation from 0.06 to 1.23. These substantial variations emphasize the necessity of calibrating and assigning lower weights to PurpleAir measurements in a spatially varying manner. Figure 2B also shows that the variation of the AQS/PurpleAir agreement was not correlated with their actual distance (correlation coefficient < 0.001), suggesting that 500 m was a reasonable distance for pairing AQS and PurpleAir. As shown in Figure 2B, some paired AQS/PurpleAir measurements had large differences (>50 $\mu g/m^3$). No temporal patterns among these large differences were found. High-level local pollution in the microenvironments of PurpleAir sensors is believed to be a potential reason for these occasional inconsistencies.

3.1.2. Spatially Varying PurpleAir $PM_{2.5}$ Calibration. The GWR slopes of PurpleAir (β_1 in eq 1) averaged 0.64 with an interquartile range of 0.02. The largest slope was 0.67 near the U.S.−Mexico border in Southern California, and the smallest was 0.62 near the coast of Northern California (Figure S1). Although the individual slopes of the AQS/PurpleAir pairs varied significantly across the domain, the calibration slopes had a narrower range because the GWR model fitted the paired measurements in a wider area at each location and other covariates also worked to remove much of the variation. This is a conservative strategy for mitigating the influence of few paired measurements with extreme coefficients on the calibration model. After calibration, the overall systematic bias of PurpleAir decreased from 1.9 to ~0 $\mu g/m^3$. The overall PurpleAir residual error was also reduced to some degree, reflected in a decreased standard deviation of the AQS/PurpleAir differences from 8.18 to 5.20 $\mu g/m^3$ (i.e., a 36% decrease). The calibration model had a 10-fold CV R^2 of 0.78, which is higher than the R^2 of 0.74 between AQS and uncalibrated PurpleAir data, again indicating the improvement of the overall precision of PurpleAir data. Table S1 shows the results of the sensitivity analysis based on randomly selected subsets of collocated AQS stations. When keeping 90% of the collocated AQS stations (23 stations), the calibrated PurpleAir data only had negligible changes. However, when keeping ~80% of the collocated stations (20 stations), although the hourly level mean absolute difference between the fully calibrated data and the calibrated data based on the subset of collocated stations was still minor (0.35 $\mu g/m^3$), the maximum absolute difference started becoming significant (>10 $\mu g/m^3$).

3.1.3. PurpleAir Sensor Bias and Influential Factors. Figure 3 shows the GAM-fitted relationships of the AQS/PurpleAir absolute differences and temperature, RH, operating time, and uptime. The 95% confidence intervals (CIs) of the relationships are shown as the shaded area. In the paired data, temperatures ranged from −1.8 to 52.6 °C with an average of

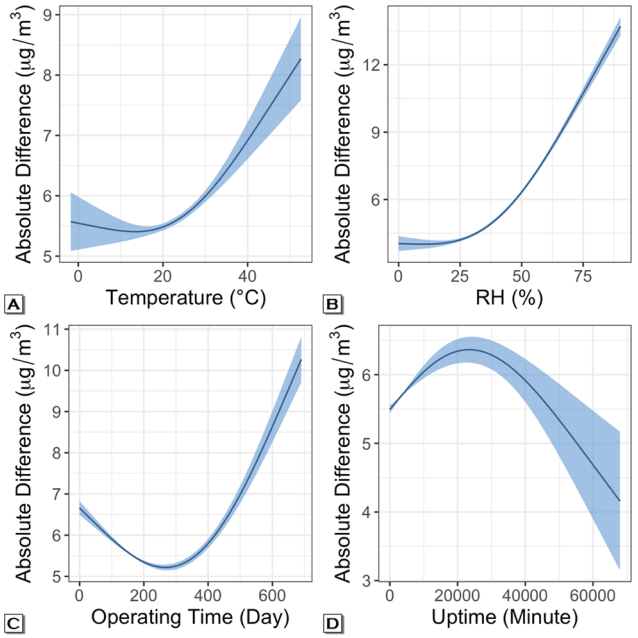


Figure 3. GAM-fitted relationships with 95% confidence intervals between the absolute differences of paired AQS/PurpleAir hourly measurements and (A) temperature, (B) RH, (C) sensor operating time, and (D) sensor uptime after controlling for other three factors.

22.4°C. Temperature was associated with the smallest absolute bias at ~20°C after adjusting for other covariates (Figure 3A). The bias significantly increased when the temperature became higher. At 50°C, the absolute bias was ~1.5 times (~2.5 $\mu g/m^3$) higher than at 20°C. In contrast, a lower temperature was only associated with a minor increase of bias. RH measures ranged from 0 to 90.1% with an average of 38.8%. RH was positively associated with the absolute bias after adjusting for other covariates (Figure 3B). Specifically, the absolute bias was relatively stable at RH < 25% but increased exponentially at RH > 25%. At 90%, the absolute bias was ~3 times (~9 $\mu g/m^3$) higher than at 25%. Operating times ranged from 1 to 690 days with an average of 198 days. After controlling for temperature, RH, and uptime, the absolute bias showed a U-curve with a minimum value at ~280 days (~9 months) (Figure 3C). As the operating time became shorter or longer, the bias increased, and the rate of increase was faster for a longer operating time. A sensor with an operating time of 700 days (~23 months) had a ~2 times (~5 $\mu g/m^3$) higher absolute bias than a sensor with an operating time of 280 days. Sensor uptimes ranged from 1 to 67 964 min (~47 days) with an average of 4881 min (~3.5 days). After adjusting for temperature, RH, and operating time, the absolute bias peaked at ~23 000 min (~16 days) and became smaller when the uptime was shorter or longer (Figure 3D). However, the derived relationship about the uptime had a large uncertainty.

3.2. Weighted $PM_{2.5}$ Modeling. *3.2.1. Residual Errors and Weights.* The clustered subdomains correspond well with the topographic, meteorological, and land-cover features in California (Figure S2): (1) the first subdomain consisting of agricultural, humid, and developed areas where most of the population resides, (2) the second subdomain consisting of mountainous areas such as the Sierra Nevada, and (3) the third subdomain mainly consisting of the arid areas in the state. The estimated mean residual variances (τ_j^2) and the corresponding weights of PurpleAir (w_j) in the subdomains are summarized

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

in Table 2. The residual variances were distinct in different subdomains, varying from 11.2 to 50.0. The variance was smallest in arid areas and largest in mountainous areas and was modest in agricultural/developed areas. The domain-specific weights of PurpleAir ranged from 0.10 to 0.17.

Table 2. Numbers of Paired AQS/PurpleAir Hourly Measurements, Mean PurpleAir Residual Variances, and PurpleAir Weights in Three Clustered Subdomains[a]

subdomain (j)	N	variance (τ_j^2)	weight (w_j)
agricultural	118 912	27.22	0.13
mountainous	3531	50.00	0.10
arid	6334	11.21	0.17

[a]The weights were calculated based on an AQS-based CV MSPE (σ^2) of 33.4 and a scale factor (ρ) of 0.23.

3.2.2. Modeling Performance and $PM_{2.5}$ Predictions. Table 3 shows the CV performance of the prediction models. Figure S3 shows the CV scatter plots of the models, indicating that the predictions from all models were slightly underestimated against AQS measurements with slopes of ~1.1. The underestimation is mainly because the RF algorithm is conservative for extreme values, and in this analysis, extremely high $PM_{2.5}$ pollution levels tended to be predicted as lower values. The spatial and temporal CV of the AQS-based model had baseline R^2 values of 0.75 and 0.77, respectively, which were lower than its random CV R^2 of 0.83. The lower spatial/temporal R^2 values reflect slightly decreased abilities to extrapolate the $PM_{2.5}$ estimates from the spatial/temporal ranges of the training data to the entire domain/time span. The contribution of PurpleAir data is shown by the higher random CV R^2 values of both nonweighted and weighted models than that of the AQS-based model. The spatial CV R^2 values of both nonweighted and weighted models also increased from 0.75 to 0.81. This change indicates that PurpleAir measurements captured $PM_{2.5}$ pollution in more microenvironments despite the network's lack of a coordinated siting strategy. The temporal CV R^2 of the nonweighted model decreased from the baseline value of 0.77 to 0.75, possibly due to the lack of sampling continuity of PurpleAir. Unlike AQS, most of the PurpleAir sensors were newly installed during the study period and maintained by untrained citizens; so, the operations were often intermittent. This lack of sampling continuity could render the measurements less representative in time. The weighted model had a spatial CV R^2 higher than the baseline value and a temporal CV R^2 comparable to the baseline value. The model also had the best random CV R^2 of 0.86 and the lowest RMSPE of 5.62 $\mu g/m^3$. These results indicate that the weighting strategy could not only result in higher spatial predictability provided by dense PurpleAir sensors but also maintain high temporal predictability provided by continuous AQS monitors.

Table 3. Cross-Validation Performance of Three Prediction Models[a]

model	random CV R^2	spatial CV R^2	temporal CV R^2	CV RMSPE ($\mu g/m^3$)
AQS-based	0.83	0.75	0.77	6.04
nonweighted	0.85	0.81	0.75	5.95
weighted	0.86	0.81	0.77	5.62

[a]CV was only performed on AQS measurements not used in calibrating PurpleAir (N = 32 981).

The $PM_{2.5}$ prediction surfaces illustrate the contribution of PurpleAir and the weighting strategy from a different angle. Figure 4 shows the annual mean $PM_{2.5}$ distributions generated from the AQS-based and weighted models, as well as their differences. The AQS-based model had an averaged $PM_{2.5}$ prediction of 9.4 $\mu g/m^3$, and the weighted model had an average of 10.0 $\mu g/m^3$. The weighted predictions were higher than the AQS-based predictions in almost all areas except for the San Francisco Bay Area, Imperial Valley, and the desert mountain ranges in Southeastern California. The higher predictions were mainly caused by higher calibrated PurpleAir measurements. During the study period, the daily calibrated PurpleAir $PM_{2.5}$ measurements had an average of 12.1 $\mu g/m^3$, higher than the daily AQS measurements averaged 11.5 $\mu g/m^3$. Figure 4C shows some hotspots where the weighted predictions were considerably higher than the AQS-based predictions. These hotspots appear to spatiotemporally coincide with the California wildfires. The black points in Figure 4C label the locations of the four-most destructive California wildfires in 2018 (i.e., Carr Fire, Camp Fire, Mendocino Complex Fire, and Ferguson Fire). The extreme weather conditions during the wildfire events, especially high air temperatures, might influence the quality of PurpleAir sensors. However, when checking the temperature measurements from the PurpleAir sensors near the wildfires, we found

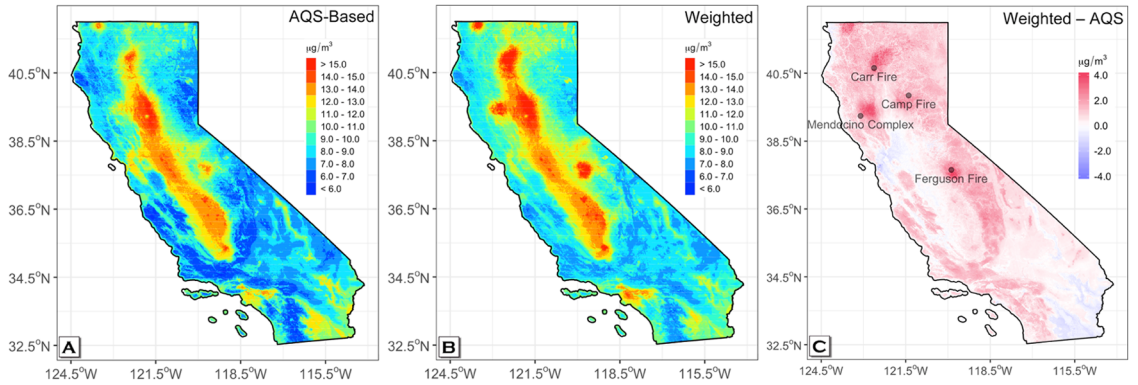


Figure 4. Annual mean $PM_{2.5}$ distributions for the year 2018 derived by (A) the AQS-based model and (B) the weighted model. (C) Annual mean $PM_{2.5}$ differences between the weighted and AQS-based models (weighted minus AQS-based) with the locations of the four-most destructive wildfires in California in 2018.

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

that the maximum measurement, 50.3°C, was still within the temperature range of the calibration model (Figure 3). This finding indicates that the PurpleAir measurements near these wildfires were well calibrated and the observed $PM_{2.5}$ hotspots are not likely caused by highly biased PurpleAir measurements due to high temperatures. As shown in Figure S4, we infer that the density of PurpleAir sensors in the study domain, allowing them to measure such episodic and high-level pollution events, could be one of the reasons for their ability to better reflect the hotspots. The utility of low-cost sensors under more extreme conditions still warrants further research. Figure S5 shows that the nonweighted predictions were higher in most of the study domain than the weighted predictions. However, the larger impact of PurpleAir residual errors on the nonweighted model reduced the credibility of its predictions. As PurpleAir measurements were aggregated to daily-level data in this analysis due to the difficulty of current $PM_{2.5}$ models generating predictions at a finer temporal scale, the contribution of high temporal frequency of PurpleAir data on $PM_{2.5}$ predictions warrants further research with improved prediction models.

4. DISCUSSION

In this study, we conducted a spatially varying calibration and developed a downweighting strategy to integrate low-cost sensor data into high-resolution $PM_{2.5}$ modeling in California. To the best of our knowledge, this is the first time such a framework has been proposed to enable $PM_{2.5}$ prediction models to take advantage of a large volume of low-cost sensor measurements while minimizing the adverse influence of their uncertainties.

Strict side-by-side collocation against reference-grade monitors has been reported in many field calibration studies of low-cost sensors.[15,19,20,25] Although it ensures the robustness of calibration, a number of limitations prevent larger-scale implementation of this method. Because side-by-side collocation in a field measurement campaign is costly and time consuming, it is usually restricted to a relatively small area such as a city[20,21] or a county.[19] Calibration coefficients fitted in a small area are difficult to apply in other regions as the low-cost sensor bias may vary under different environmental conditions.[13,14,18] More importantly, strict collocation is difficult when both low-cost sensor and regulatory networks are already established. Instead, we tested a less stringent collocation strategy by matching a PurpleAir sensor to its nearest AQS station within a radius of 500 m so that each AQS/PurpleAir pair was within a 1 km modeling grid cell. The reasonability of our collocation strategy was bolstered by the fact that an agreement between AQS and PurpleAir data was not related to the actual distance between monitors within 500 m. Furthermore, it allowed for sufficient collocated samples to conduct the calibration. As the AQS/PurpleAir agreement was heterogeneous across the domain where the bias of PurpleAir data was lower near the coastal area of Northern California and higher near the U.S.−Mexico border, the PurpleAir calibration was performed with a GWR model in a spatially varying manner. The calibration reduced the overall systematic bias of PurpleAir data from 1.9 to ∼0 $\mu g/m^3$. The overall residual error of the measurements was also reduced by 36%. Results from the sensitivity analysis examining a reduced set of collocated AQS stations (Table S1) suggest that for a region the size of California, at least ∼20 well-distributed, continuous reference-grade monitors (capable of providing hourly $PM_{2.5}$ measurements) are needed to effectively calibrate hourly level PurpleAir measurements. Thus, our recommended reference-grade monitor density is ∼5 stations per 100 000 km^2. By the end of 2018, 37 states in the contiguous United States (CONUS) other than California had a density of continuous AQS stations greater than 5 per 100 000 km^2 (Table S3). Accordingly, the proposed PurpleAir calibration framework may potentially be generalized to the majority of CONUS states without deploying new reference-grade air quality stations. In fact, for the states with a lower network density, an effective calibration could still be conducted by grouping with surrounding states.

Due to the lack of detailed operational conditions of the PurpleAir sensors, assigning site-specific weights was unrealistic. Instead, we clustered these sensors into groups and assigned each group a single weight. We demonstrated a downweighting approach to minimize the influences of residual errors and other factors affecting low-cost sensor measurements. The population-level downweighting formula consists of two parts: an explicit error variance ratio and a data-driven scale factor (ρ). The error variance ratio accounts for the proportion of the mean PurpleAir residual variance in the total possible error variance of the model. As the residual error becomes smaller, this ratio becomes larger and so does the weight. The scale factor ρ was used as a proxy of the negative impacts of implicit factors of low-cost sensors on modeling performance, such as sampling discontinuity and less representative sensor siting.[46] Since the impacts of these implicit factors were unquantifiable, the optimal ρ was determined by our model-fitting data within a range from 0 to 1. It is worth noting that for a set of measurements with an overall quality close to the gold-standard measurements, ρ may approach 1 (Section 4, Supporting Information). In this analysis, the population-level weights were between 0.10 and 0.17, indicating that even though the bias of PurpleAir measurements could be eliminated by a statistical calibration, the contribution of PurpleAir measurements was still no more than 20% of that of AQS measurements in achieving the best model-predicting performance of daily $PM_{2.5}$ concentrations. Compared to the nonweighted prediction model, the weighted prediction model had a higher random CV R^2 of 0.86, a higher temporal CV R^2 of 0.81, and a lower RMSPE, indicating that the weighting strategy was able to compensate the loss of predictability caused by the PurpleAir residual errors. Through this study and its pilot study conducted in a region with insufficient AQS stations,[29] we found that the improvement of statistical metrics, such as the increase of CV R^2 values, can only partly reflect the contribution of low-cost sensors to the quality of $PM_{2.5}$ prediction, especially when CV is based solely on reference-grade measurements as in this analysis. In the absence of high-quality and high-coverage measurements, we believe that examining the spatial distribution of $PM_{2.5}$ predictions (Figures 4 and S5) is an important complement to CV metrics to evaluate low-cost sensor data. As shown in this analysis, despite the small improvement in CV metrics, small-scale pollution features are able to be well captured by low-cost sensors. Another possible way to mitigate the influence of low-cost sensor measurement errors on $PM_{2.5}$ modeling is interpolating these discrete measurements into continuous surfaces and treating the surfaces as an independent variable in the model. We think our weighting strategy is advantageous because incorporating low-cost sensor measurements into the dependent variable will lead to a

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

significantly larger training sample capable of providing considerably more detailed spatiotemporal information about the pollutant. In this analysis, PurpleAir provided ∼5 times more training samples than AQS, and these samples could help improve the model predictability and better measure pollution hotspots.

The dense and spatially extensive low-cost sensor measurements allowed us to analyze the potential factors related to the bias and residual error of low-cost sensor measurements. Increased temperature and RH were associated with a near-exponentially increased PurpleAir data bias. The observed influences of high temperature/humidity on low-cost sensor bias may be related to the issues in electronic circuits and the hygroscopic growth of fine particulates.[23,47,48] The sensor operating time was an influential factor of the bias as well, where a PurpleAir sensor with an operating time of 2 years tended to have a ∼2 times higher bias than a sensor with an operating time of 9 months. The increased bias over time may reflect the aging effect of sensors.[17,36,37] A shorter operating time than 9 months was also associated with a slightly increased sensor bias, suggesting a "break-in" or "warm-up" period of the sensor. The mechanism of the break-in warrants further investigation. A longer sensor uptime was, in general, associated with a lower sensor bias, indicating that stable operation would generally result in better data quality. However, this relationship had a large degree of uncertainty, probably because the sensor's operational stability is associated with many factors other than the sensor itself, such as the reliability of power supply. In terms of the residual error, the mountainous areas had the highest estimated PurpleAir residual error, while the arid areas had the lowest. This difference indicates that (1) humidity may still play a role in the residual errors of low-cost sensor data even after controlling it in the calibration stage and (2) the change of $PM_{2.5}$ composition in different land-use types may differentially affect the accuracy of the formula the manufacturer of PurpleAir used to convert light scatter to mass concentration. Given the limited information about the sensors we were able to collect, the factors other than temperature, humidity, sensor operating time, and uptime could not be analyzed. More in-depth analyses on the influential factors of sensor data quality are needed but they are beyond the scope of this study.

Overall, a two-step approach, i.e., spatially varying calibration and downweighting modeling, was developed to combine low-cost sensor data with regulatory measurements to improve the quality of high-resolution spatiotemporal $PM_{2.5}$ modeling. The proposed approach was able to mitigate the negative impact of the high noises in low-cost sensor measurements on $PM_{2.5}$ prediction accuracy. Dense low-cost sensor measurements in the study domain also showed their potential to help the prediction model better reflect $PM_{2.5}$ hotspots such as wildfires. This study demonstrated that the integration of low-cost sensors with regulatory monitoring and other sources of information such as satellite remote sensing can provide new insights into $PM_{2.5}$ pollution. PurpleAir is a global monitoring network with a rapid growth rate. All other supporting data in this analysis, including satellite, meteorological, land-use, and demographic data, are not limited to our study domain. Therefore, our two-step approach can be generalized to other regions to derive high-resolution $PM_{2.5}$ exposure estimates. The proposed approach is also informative to other meteorological, geographical, and ecological citizen science applications to calibrate large volumes of low-quality volunteer-generated data.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.9b06046.

Quality control for PurpleAir $PM_{2.5}$ measurements (Section 1); evaluation of PurpleAir $PM_{2.5}$ measurements (Section 2); nonlinearity of PurpleAir systematic bias (Section 3); validation of scale factor ρ (Section 4); PurpleAir calibration based on subsets of AQS stations (Table S1); HAC clustering feature space (Table S2); numbers and densities of continuous AQS stations (Table S3); spatial distribution of GWR slopes (Figure S1); clustered subdomains (Figure S2); ten-fold CV scatter plots (Figure S3); $PM_{2.5}$ predictions at Ferguson Fire (Figure S4); $PM_{2.5}$ predictions of the nonweighted model (Figure S5); PurpleAir dual-channel hourly measurements (Figure S6); AQS and PurpleAir hourly measurements (Figure S7); determination of scale factor ρ (Figure S8) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yang Liu − *Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322, United States;* orcid.org/0000-0001-5477-2186; Email: yang.liu@emory.edu

Authors

Jianzhao Bi − *Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322, United States*

Avani Wildani − *Department of Computer Science, Emory University, Atlanta, Georgia 30307, United States*

Howard H. Chang − *Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.9b06046

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work of J.B. and Y.L. was supported by the National Aeronautics and Space Administration (NASA) Applied Sciences Program (Grant Nos. NNX16AQ28Q and 80NSSC19K0191). The content is solely the responsibility of the authors and does not necessarily represent the official views of NASA. The study is not promoted or endorsed by any organization in terms of the use of PurpleAir sensors or data.

■ REFERENCES

(1) Brook, R. D.; Rajagopalan, S.; Pope, C. A., 3rd; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y.; Luepker, R. V.; Mittleman, M. A.; Peters, A.; Siscovick, D.; Smith, S. C., Jr.; Whitsel, L.; Kaufman, J. D.; American Heart Association Council on Epidemiology and Prevention, Council on the Kidney in Cardiovascular Disease, and Council on Nutrition, Physical Activity and Metabolism. Metabolism, Particulate matter air pollution and

cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation* **2010**, *121*, 2331−2378.
(2) Pope, C. A., 3rd; Dockery, D. W. Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Manage. Assoc.* **2006**, *56*, 709−742.
(3) Cohen, A. J.; Brauer, M.; Burnett, R.; Anderson, H. R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; Feigin, V.; Freedman, G.; Hubbell, B.; Jobling, A.; Kan, H.; Knibbs, L.; Liu, Y.; Martin, R.; Morawska, L.; Pope, C. A., 3rd; Shin, H.; Straif, K.; Shaddick, G.; Thomas, M.; van Dingenen, R.; van Donkelaar, A.; Vos, T.; Murray, C. J. L.; Forouzanfar, M. H. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **2017**, *389*, 1907−1918.
(4) Burnett, R. T.; Pope, C. A., 3rd; Ezzati, M.; Olives, C.; Lim, S. S.; Mehta, S.; Shin, H. H.; Singh, G.; Hubbell, B.; Brauer, M.; Anderson, H. R.; Smith, K. R.; Balmes, J. R.; Bruce, N. G.; Kan, H.; Laden, F.; Pruss-Ustun, A.; Turner, M. C.; Gapstur, S. M.; Diver, W. R.; Cohen, A. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect* **2014**, *122*, 397−403.
(5) Sarnat, S. E.; Winquist, A.; Schauer, J. J.; Turner, J. R.; Sarnat, J. A. Fine particulate matter components and emergency department visits for cardiovascular and respiratory diseases in the St. Louis, Missouri-Illinois, metropolitan area. *Environ. Health Perspect.* **2015**, *123*, 437−444.
(6) O'Lenick, C. R.; Winquist, A.; Mulholland, J. A.; Friberg, M. D.; Chang, H. H.; Kramer, M. R.; Darrow, L. A.; Sarnat, S. E. Assessment of neighbourhood-level socioeconomic status as a modifier of air pollution-asthma associations among children in Atlanta. *J. Epidemiol. Community Health* **2017**, *71*, 129−136.
(7) Zhao, Y.; Wang, S.; Duan, L.; Lei, Y.; Cao, P.; Hao, J. Primary air pollutant emissions of coal-fired power plants in China: Current status and future prediction. *Atmos. Environ.* **2008**, *42*, 8442−8452.
(8) Ma, Z.; Hu, X.; Sayer, A. M.; Levy, R.; Zhang, Q.; Xue, Y.; Tong, S.; Bi, J.; Huang, L.; Liu, Y. Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004-2013. *Environ. Health Perspect.* **2016**, *124*, 184−192.
(9) Bi, J.; Belle, J. H.; Wang, Y.; Lyapustin, A. I.; Wildani, A.; Liu, Y. Impacts of snow and cloud covers on satellite-derived PM2.5 levels. *Remote Sens. Environ.* **2019**, *221*, 665−674.
(10) Snyder, E. G.; Watkins, T. H.; Solomon, P. A.; Thoma, E. D.; Williams, R. W.; Hagler, G. S.; Shelow, D.; Hindin, D. A.; Kilaru, V. J.; Preuss, P. W. The changing paradigm of air pollution monitoring. *Environ. Sci. Technol.* **2013**, *47*, 11369−11377.
(11) Gupta, P.; Doraiswamy, P.; Levy, R.; Pikelnaya, O.; Maibach, J.; Feenstra, B.; Polidori, A.; Kiros, F.; Mills, K. C. Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations. *Geohealth* **2018**, *2*, 172−181.
(12) Al-Saadi, J.; Szykman, J.; Pierce, R. B.; Kittaka, C.; Neil, D.; Chu, D. A.; Remer, L.; Gumley, L.; Prins, E.; Weinstock, L.; MacDonald, C.; Wayland, R.; Dimmick, F.; Fishman, J. Improving National Air Quality Forecasts with Satellite Aerosol Observations. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 1249−1262.
(13) Morawska, L.; Thai, P. K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environ. Int.* **2018**, *116*, 286−299.
(14) Castell, N.; Dauge, F. R.; Schneider, P.; Vogt, M.; Lerner, U.; Fishbain, B.; Broday, D.; Bartonova, A. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environ. Int.* **2017**, *99*, 293−302.
(15) Jiao, W.; Hagler, G.; Williams, R.; Sharpe, R.; Brown, R.; Garver, D.; Judge, R.; Caudill, M.; Rickard, J.; Davis, M. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmos. Meas. Tech.* **2016**, *9*, 5281−5292.
(16) Clements, A. L.; Griswold, W. G.; Rs, A.; Johnston, J. E.; Herting, M. M.; Thorson, J.; Collier-Oxandale, A.; Hannigan, M. Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). *Sensors* **2017**, *17*, No. 2478.
(17) Masson, N.; Piedrahita, R.; Hannigan, M. Quantification Method for Electrolytic Sensors in Long-Term Monitoring of Ambient Air Quality. *Sensors* **2015**, *15*, 27283−272302.
(18) Austin, E.; Novosselov, I.; Seto, E.; Yost, M. G. Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor. *PLoS One* **2015**, *10*, No. e0137789.
(19) Carvlin, G. N.; Lugo, H.; Olmedo, L.; Bejarano, E.; Wilkie, A.; Meltzer, D.; Wong, M.; King, G.; Northcross, A.; Jerrett, M.; English, P. B.; Hammond, D.; Seto, E. Development and field validation of a community-engaged particulate matter air quality monitoring network in Imperial, California, USA. *J. Air Waste Manage. Assoc.* **2017**, *67*, 1342−1352.
(20) Liu, H.-Y.; Schneider, P.; Haugen, R.; Vogt, M. Performance Assessment of a Low-Cost PM2.5 Sensor for a near Four-Month Period in Oslo, Norway. *Atmosphere* **2019**, *10*, No. 41.
(21) Gao, M.; Cao, J.; Seto, E. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *Environ. Pollut.* **2015**, *199*, 56−65.
(22) Ripoll, A.; Viana, M.; Padrosa, M.; Querol, X.; Minutolo, A.; Hou, K. M.; Barcelo-Ordinas, J. M.; Garcia-Vidal, J. Testing the performance of sensors for ozone pollution monitoring in a citizen science approach. *Sci. Total Environ.* **2019**, *651*, 1166−1179.
(23) Crilley, L. R.; Shaw, M.; Pound, R.; Kramer, L. J.; Price, R.; Young, S.; Lewis, A. C.; Pope, F. D. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmos. Meas. Tech.* **2018**, *11*, 709−720.
(24) Holstius, D. M.; Pillarisetti, A.; Smith, K.; Seto, E. Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. *Atmos. Meas. Tech.* **2014**, *7*, 1121−1131.
(25) Zheng, T.; Bergin, M. H.; Johnson, K. K.; Tripathi, S. N.; Shirodkar, S.; Landis, M. S.; Sutaria, R.; Carlson, D. E. Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments. *Atmos. Meas. Tech.* **2018**, *11*, 4823−4846.
(26) Pope, F. D.; Gatari, M.; Ng'ang'a, D.; Poynter, A.; Blake, R. Airborne particulate matter monitoring in Kenya using calibrated low-cost sensors. *Atmos. Chem. Phys.* **2018**, *18*, 15403−15418.
(27) English, P. B.; Olmedo, L.; Bejarano, E.; Lugo, H.; Murillo, E.; Seto, E.; Wong, M.; King, G.; Wilkie, A.; Meltzer, D.; Carvlin, G.; Jerrett, M.; Northcross, A. The Imperial County Community Air Monitoring Network: A Model for Community-based Environmental Monitoring for Public Health Action. *Environ. Health Perspect* **2017**, *125*, No. 074501.
(28) Masiol, M.; Zikova, N.; Chalupa, D. C.; Rich, D. Q.; Ferro, A. R.; Hopke, P. K. Hourly land-use regression models based on low-cost PM monitor data. *Environ. Res.* **2018**, *167*, 7−14.
(29) Bi, J.; Stowell, J.; Seto, E. Y. W.; English, P. B.; Al-Hamdan, M. Z.; Kinney, P. L.; Freedman, F. R.; Liu, Y. Contribution of low-cost sensor measurements to the prediction of PM2.5 levels: A case study in Imperial County, California, USA. *Environ. Res.* **2020**, *180*, No. 108810.
(30) Mahmud, A.; Hixson, M.; Hu, J.; Zhao, Z.; Chen, S. H.; Kleeman, M. J. Climate impact on airborne particulate matter concentrations in California using seven year analysis periods. *Atmos. Chem. Phys.* **2010**, *10*, 11097−11114.
(31) Koelemeijer, R. B. A.; Homan, C. D.; Matthijsen, J. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmos. Environ.* **2006**, *40*, 5304−5315.
(32) Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. MODIS collection 6 MAIAC algorithm. *Atmos. Meas. Tech. Discuss* **2018**, *11*, 1−50.
(33) Belle, J.; Liu, Y. Evaluation of Aqua MODIS Collection 6 AOD Parameters for Air Quality Research over the Continental United States. *Remote Sens.* **2016**, *8*, No. 815.

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162

(34) Mesinger, F.; DiMego, G.; Kalnay, E.; Mitchell, K.; Shafran, P. C.; Ebisuzaki, W.; Jovi, D.; Woollen, J.; Rogers, E.; Berbery, E. H. North American regional reanalysis. *Bull. Am. Meteorol. Soc.* **2006**, *87*, 343−360.
(35) Mitchell, K. E.; Lohmann, D.; Houser, P. R.; Wood, E. F.; Schaake, J. C.; Robock, A.; Cosgrove, B. A.; Sheffield, J.; Duan, Q.; Luo, L. The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.: Atmos.* **2004**, *109*, No. 7449.
(36) Feinberg, S.; Williams, R.; Hagler, G. S.; Rickard, J.; Brown, R.; Garver, D.; Harshfield, G.; Stauffer, P.; Mattson, E.; Judge, R. Long-term evaluation of air sensor technology under ambient conditions in Denver, Colorado. *Atmos. Meas. Tech.* **2018**, *11*, 4605−4615.
(37) Rai, A. C.; Kumar, P.; Pilla, F.; Skouloudis, A. N.; Di Sabatino, S.; Ratti, C.; Yasar, A.; Rickerby, D. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Sci. Total Environ.* **2017**, *607−608*, 691−705.
(38) Gollini, I.; Lu, B. B.; Charlton, M.; Brunsdon, C.; Harris, P. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *J. Stat. Software* **2015**, *63*, 1−50.
(39) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(40) Hu, X.; Belle, J. H.; Meng, X.; Wildani, A.; Waller, L. A.; Strickland, M. J.; Liu, Y. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. *Environ. Sci. Technol.* **2017**, *51*, 6936−6944.
(41) Wright, M. N.; Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Software* **2017**, *77*, 1−17.
(42) Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236−244.
(43) Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A.; Charrad, M. M. Package 'nbclust'. *J. Stat. Software* **2014**, *61*, 1−36.
(44) Kelly, K. E.; Whitaker, J.; Petty, A.; Widmer, C.; Dybwad, A.; Sleeth, D.; Martin, R.; Butterfield, A. Ambient and laboratory evaluation of a low-cost particulate matter sensor. *Environ. Pollut.* **2017**, *221*, 491−500.
(45) Levy Zamora, M.; Xiong, F.; Gentner, D.; Kerkez, B.; Kohrman-Glaser, J.; Koehler, K. Field and Laboratory Evaluations of the Low-Cost Plantower Particulate Matter Sensor. *Environ. Sci. Technol.* **2019**, *53*, 838−849.
(46) Geng, G.; Murray, N. L.; Chang, H. H.; Liu, Y. The sensitivity of satellite-based PM2.5 estimates to its inputs: Implications to model development in data-poor regions. *Environ. Int.* **2018**, *121*, 550−560.
(47) Allen, G.; Sioutas, C.; Koutrakis, P.; Reiss, R.; Lurmann, F. W.; Roberts, P. T. Evaluation of the TEOM method for measurement of ambient particulate mass in urban areas. *J. Air Waste Manage. Assoc.* **1997**, *47*, 682−689.
(48) Day, D. E.; Malm, W. C.; Kreidenweis, S. M. Aerosol light scattering measurements as a function of relative humidity. *J. Air Waste Manage. Assoc.* **2000**, *50*, 710−716.

https://dx.doi.org/10.1021/acs.est.9b06046
Environ. Sci. Technol. 2020, 54, 2152−2162