Incorporating Low-Cost Sensor Measurements into High-Resolution PM2.5 Modeling at a Large Spatial Scale

Jianzhao Bi[1], Avani Wildani[1], Howard Chang[1], and Yang Liu[1]

[1]Emory University

November 24, 2022

Abstract

Low-cost air quality monitors (LCAQMs) are promising supplements to regulatory monitors for PM2.5 exposure assessment. However, the application of LCAQM in spatially extensive exposure modeling is hindered by the difficulty in performing calibration at large spatial scales and the adverse influence of LCAQM residual uncertainty after calibration. We aimed to develop an efficient spatially scalable calibration method for LCAQM and design a residual uncertainty-derived down-weighting strategy to optimize the use of LCAQM data with regulatory monitoring data in PM2.5 modeling. In California, for each monitor from PurpleAir, a global LCAQM network, we identified a station within a 500-m radius from the Air Quality System (AQS), a U.S. regulatory monitoring network. Regional calibration of PurpleAir to AQS was performed at the hourly level with Geographically Weighted Regression (GWR). The calibrated PurpleAir measurements were down-weighted according to their residual uncertainty and then incorporated into a Random Forest (RF) prediction model as a dependent variable to generate 1-km daily PM2.5 exposure estimates. The state-level PurpleAir calibration reduced the systematic bias to ~0 ug/m3 and decreased the random error by 38%. The considerably large samples also enabled quantitative analyses regarding potential factors related to the PurpleAir bias. The RF-based model with both AQS and down-weighted PurpleAir data outperformed the RF model based solely on AQS with an improved CV R2 of 0.86, an improved spatial CV R2 of 0.81, and a lower prediction error of 5.40 ug/m3. The down-weighting allowed the prediction model to show more spatial details of PM2.5 and to better detect pollution hot-spots. Our spatially scalable calibration and down-weighting strategies, for the first time, allowed an effective application of a state-level LCAQM network in high-resolution PM2.5 exposure modeling. The proposed framework can be generalized to regions worldwide for advancing the evaluation of heavy PM2.5 episodes and health-related applications.

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

Jianzhao Bi[1], Avani Wildani[2], Howard H. Chang[3], Yang Liu[*,1]

[1] Department of Environmental Health, Emory University, Rollins School of Public Health, Atlanta, GA, USA

[2] Department of Computer Science, Emory University, Atlanta, GA, USA

[3] Department of Biostatistics and Bioinformatics, Emory University, Rollins School of Public Health, Atlanta, GA, USA

* Mailing Address: 1518 Clifton Road NE, Atlanta, GA 30322, USA. E-mail: yang.liu@emory.edu.

Background

Fine particulate matter ($PM_{2.5}$) is associated with a broad range of adverse health outcomes. Ambient $PM_{2.5}$ exposure assessment has traditionally relied on sparse regulatory air quality monitoring stations. Emerging low-cost air quality sensors (<$2,500) have desirable features such as flexibility of deployment and ease of maintenance. However, 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, the side-by-side low-cost sensor calibration against reference-grade monitors has mostly been confined in a small region. **Secondly**, even though low-cost sensor data can have a relatively low systematic bias after calibration, their precision is still not comparable to reference-grade measurements. In this study, we conducted **a spatially varying calibration** and developed **a down-weighting strategy** to integrate low-cost sensor data (***PurpleAir***) with regulatory data (Air Quality System, ***AQS***) into high-resolution $PM_{2.5}$ modeling in California.

Data and Methods

Large-Scale PurpleAir Calibration

* PurpleAir sensors were paired with the nearest AQS stations within a 500-m radius (26 paired AQS/PurpleAir sites in California)
* A Geographically Weighted Regression (GWR) model with temperature, humidity, PurpleAir sensor operational time for the calibration

GWR Calibration Model

$$\text{AQS PM}_{2.5} = \beta_0 + \beta_1 \cdot \text{PurpleAir PM}_{2.5} + \beta_2 \cdot \text{T} + \beta_3 \cdot \text{RH} + \beta_4 \cdot \text{Opl.Time}$$

Weighted $PM_{2.5}$ Modeling

* PurpleAir Weights
  – A reference weight for AQS, $\mathbf{w_{AQS} = 1}$
  – Lower weights for PurpleAir, $\mathbf{w_{PA} \in (0, 1)}$
    σ^2: Errors of prediction model structure
    τ^2: Residual errors of PurpleAir
    ρ: A data-driven scale factor

PurpleAir Weights: $\quad w_{PA} = \rho \cdot \frac{\sigma^2}{\sigma^2 + \tau^2}$

* The Random Forest Prediction Model
  – **Dependent Variable:** AQS and PurpleAir $PM_{2.5}$ data with different weights
  – **Predictors:** Satellite aerosol optical depth (AOD), meteorological, and land-use data
  – 1-km, daily $PM_{2.5}$ predictions were generated

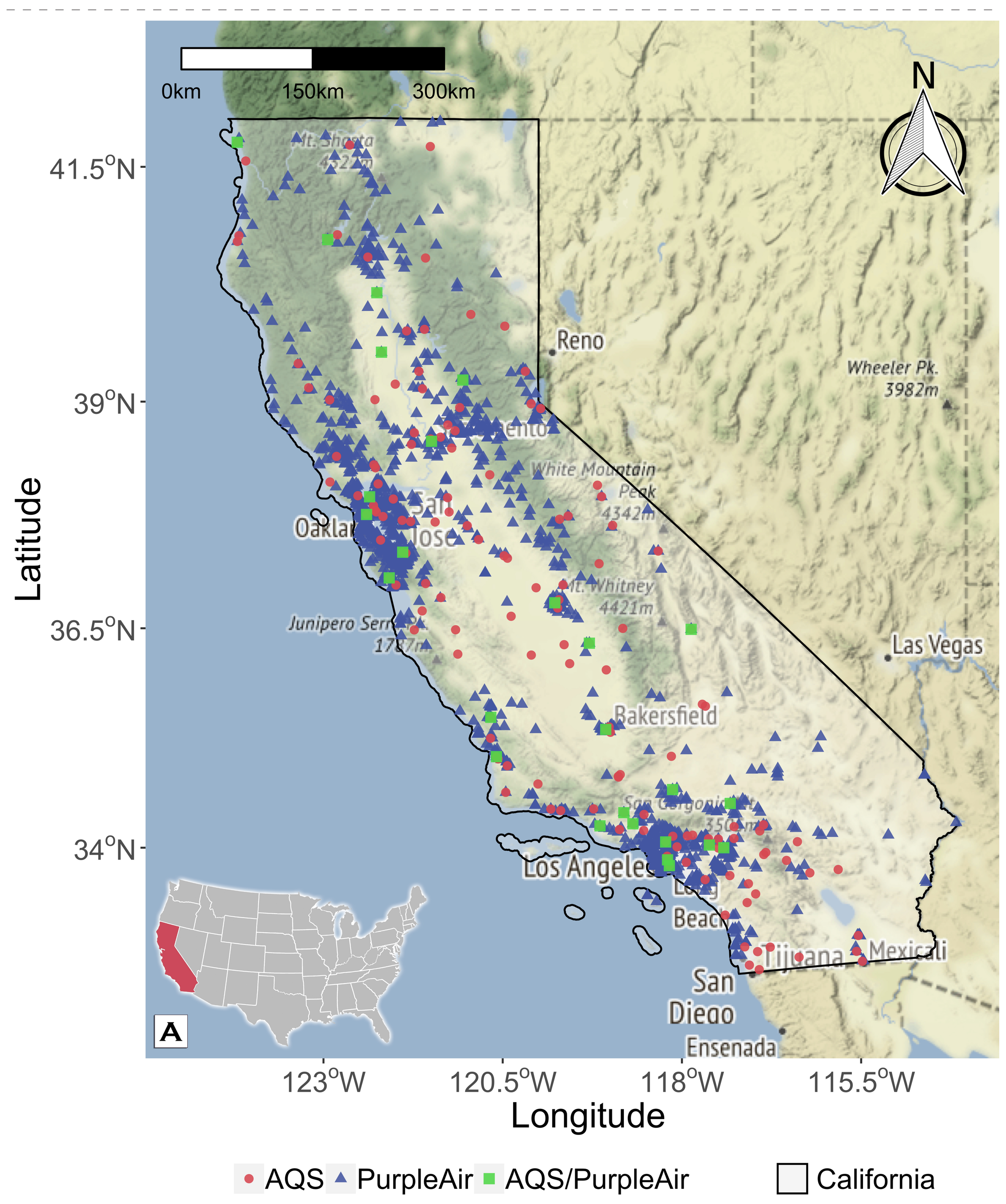

Fig 1. Study domain, California, with 157 AQS, 2,090 PurpleAir, and 26 paired AQS/PurpleAir sites.

Acknowledgements

The work of J. Bi and Y. Liu was supported by the National Aeronautics and Space Administration (NASA) Applied Sciences Program (Grant # 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 data.

Implications

* For a region with the size of California, at least ∼20 well-distributed, continuous reference-grade monitors, *i.e.*, ∼**5 stations per 100,000 km²**, are needed to effectively calibrate PurpleAir data.
* The negative impact of the large uncertainty in low-cost sensor data can be mitigated by down-weighted modeling to better take advantage of their high spatiotemporal frequency in $PM_{2.5}$ estimation.
* The two-step low-cost sensor data integration framework (calibration and down-weighting) can be **generalized to other regions** with limited regulatory monitors to advance $PM_{2.5}$ exposure assessment.
* The proposed framework can even be **transferred to other citizen science applications**, such as meteorological, geographical, and ecological citizen science programs, to combine a large volume of low-quality volunteer-generated data and few gold-standard scientific data.

Weighted $PM_{2.5}$ Modeling

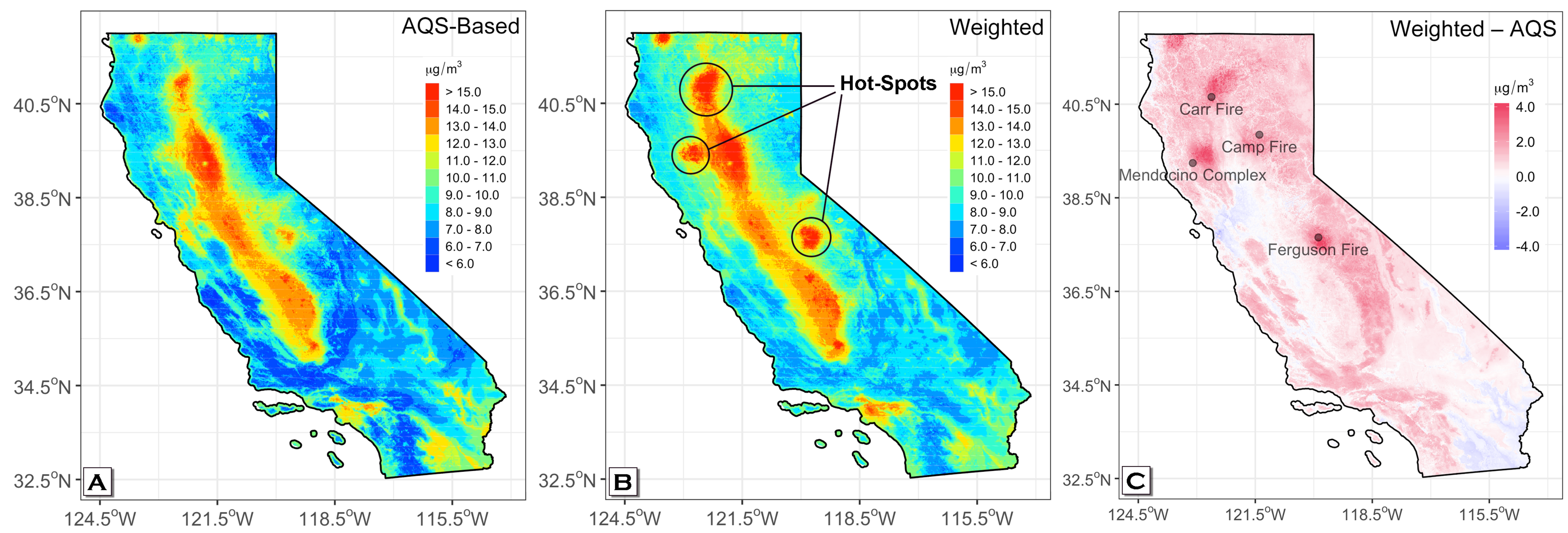

Fig 2. (A) – (B): Annual mean $PM_{2.5}$ distributions for the year of 2018 derived by (A) the AQS-based model and (B) the weighted model. (C): Annual mean $PM_{2.5}$ differences between the weighted and AQS-based models (weighted minus AQS-based) with the locations of the four most destructive wildfires in California in 2018.

Table 1. Cross-validation performance of the prediction models. CV was only performed on AQS measurements not used in calibrating PurpleAir (N = 32,981).

Model	Random CV R^2	Spatial CV R^2	Temporal CV R^2	CV RMSPE ($\mu g/m^3$)
The AQS-Based Model	0.83	0.75	0.77	6.04
The Weighted Model	0.86	0.81	0.77	5.62

* The PurpleAir weights were between 0.10 to 0.17 (against the AQS weight of 1), indicating that the contribution of PurpleAir data was no more than 20% of that of AQS data in achieving the best modeling performance.
* Dense low-cost measurements showed their potential to help the prediction model better reflect $PM_{2.5}$ hot-spots such as wildfires.

Large-Scale PurpleAir Calibration

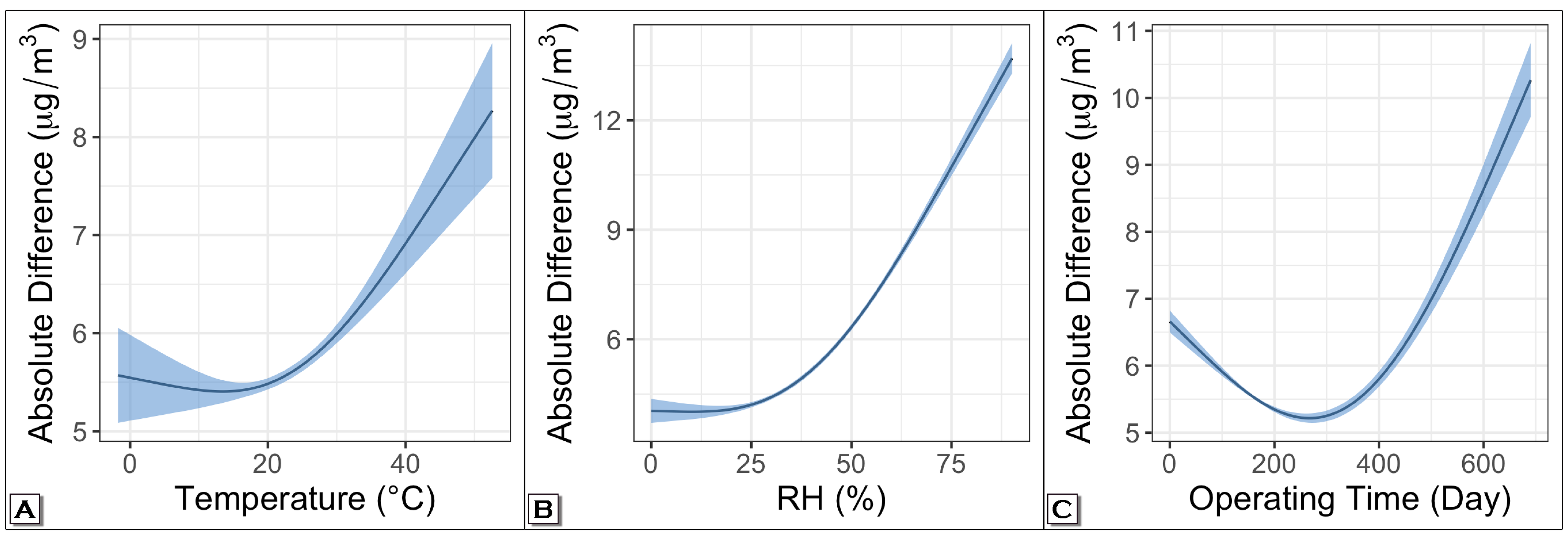

Fig 3. The nonlinear relationships with 95% confidence intervals between the absolute differences of paired AQS/PurpleAir hourly measurements and (A) temperature, (B) RH, and (C) sensor operating time.

* The calibration reduced the overall systematic bias of PurpleAir from 1.9 $\mu g/m^3$ to ∼0 $\mu g/m^3$.
* The overall residual error of PurpleAir measurements was also decreased by 36%.
* Increased temperature and humidity were related to a near-exponentially increased PurpleAir data bias.
* A sensor with an operating time of 2 years tended to have a ∼2-time higher bias than a sensor in 9 months.