Science of the Total Environment 1040 (2026) 181915

Contents lists available at ScienceDirect

Science of the Total Environment

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

Winter inversions and summer smoke: A season-dependent approach to $PM_{2.5}$ modeling with low-cost sensors in complex terrain

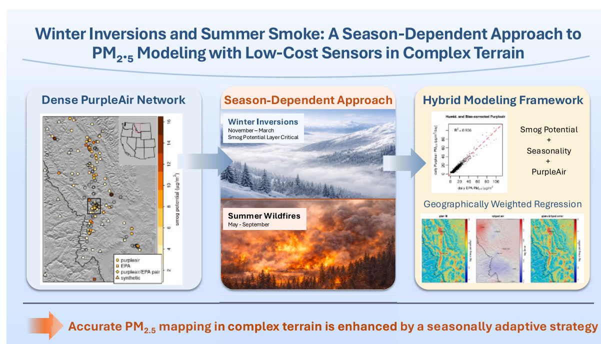
Alan Swanson [a,1], Ava Orr [a,*,1], Zachary A. Holden [b], Kyle Bocinsky [c,d], Erin L. Landguth [a]

[a] Center for Population Health Research, University of Montana, 32 Campus Drive, Missoula, MT, USA
[b] Rocky Mountain Research Station, USDA Forest Service, Missoula, MT, USA
[c] Montana Climate Office, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT, USA
[d] Department of Society and Conservation, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT, USA

HIGHLIGHTS

- PurpleAir sensors substantially improve $PM_{2.5}$ accuracy in complex terrain.
- Smog potential maps capture winter inversion-driven pollution in mountain valleys.
- Daily 600 m $PM_{2.5}$ surfaces generated using geographically weighted regression.
- $PM_{2.5}$ modeling strategy differs between winter inversions and summer wildfires.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Wildfires
$PM_{2.5}$
air pollution
PurpleAir
Air Quality Modeling
Air Sensor Data

ABSTRACT

Accurate $PM_{2.5}$ monitoring in complex terrain is challenged by sparse regulatory networks and distinct seasonal pollution sources, like winter temperature inversions and summer wildfires. This study develops and evaluates a hybrid modeling framework that integrates bias-corrected low-cost sensor data to improve spatial and temporal $PM_{2.5}$ mapping in Montana, USA. We first applied a local, observation-based bias correction to dense PurpleAir sensor network, substantially improving agreement with reference monitors (R^2 increased from 0.750 for a generic correction to 0.838). This corrected data was used to create high-resolution smog potential (smogP) surface, a static layer representing terrain-driven pollution accumulation, which explained 59.5% of the spatial variance of mean $PM_{2.5}$ during wintertime inversion events. Daily 600 m resolution $PM_{2.5}$ grids were then generated using geographically weighted regression framework. Cross-validation revealed a strong seasonal dichotomy in model performance: during the winter inversion season (November–March), models incorporating the smogP layer were critical and consistently outperformed other approaches. Conversely, during the summer wildfire season (May–September), the high density of sensor data alone was often sufficient, with simpler spatial interpolation models proving as effective at capturing the diffuse nature of large smoke plumes than models constrained by static covariates like smogP or satellite aerosol optical depth (AOD). We conclude that the optimal

* Corresponding author.
E-mail address: ava.orr@mso.umt.edu (A. Orr).
[1] Co-first author.

https://doi.org/10.1016/j.scitotenv.2026.181915
Received 14 January 2026; Received in revised form 6 May 2026; Accepted 25 May 2026
0048-9697/© 2026 Published by Elsevier B.V.

strategy for $PM_{2.5}$ mapping in complex terrain is season-dependent, and this work provides a framework for leveraging citizen-science data to enhance air quality forecasting in these challenging environments.

Synopsis: Effective $PM_{2.5}$ monitoring/modeling in complex terrain requires a seasonally adaptive strategy, using a terrain-based smog potential model for winter inversions and flexible spatial interpolation of dense sensor data for summer wildfire smoke.

1. Introduction

Fine particulate matter ($PM_{2.5}$), airborne particles with a diameter of less than 2.5 μm, is consistently linked to adverse health (Horne et al., 2018; Dockery and Pope, 1994; Jeong et al., 2019; Jainonthee et al., 2022; Orr et al., 2025a; Chen et al., 2021; DeFlorio-Barker et al., 2019; Ebisu et al., 2019; Magzamen et al., 2021; Wu et al., 2022). The Intermountain West of the United States faces a dual air quality challenge: valleys choked with stagnant, polluted air during winter inversions, and landscapes blanketed by wildfire smoke in the summer. As residential wood burning (Hadley et al., 2021; Ward and Lange, 2010), prescribed fires (Liu et al., 2017; Maji et al., 2024), and large-scale wildfires continue to shape the region's air quality, the need for accurate, high-resolution exposure estimates has never been more critical. Rural mountainous communities are often underrepresented in traditional monitoring networks despite experiencing substantial wildfire smoke exposure and wintertime pollution episodes.

Current exposure models, often based on satellite products or chemical transport models, offer broad spatial coverage but frequently fail to capture the hyperlocal pollution patterns mediated by complex terrain. In the mountainous valleys and basins of the West, topographical features and localized meteorological phenomena, such as temperature inversions and valley winds, create unique conditions where pollutants become trapped and concentrated. This results in exposure hotspots that broad-scale models miss (Orr et al., 2025b), leading to a significant underestimation of public health risk (Hadley et al., 2021). These terrain-driven processes also produce strong spatial nonstationarity in $PM_{2.5}$ concentrations, where pollutant behavior and predictor relationships vary substantially across valleys, basins, urban corridors, and high-elevation terrain.

To improve $PM_{2.5}$ exposure estimation, a wide range of statistical, machine learning, and hybrid modeling frameworks have been developed that integrate multiple data sources, including satellite aerosol products, land-use variables, meteorological data, chemical transport model outputs, and ground monitoring observations. Common approaches include land-use regression, geographically weighted regression (GWR), random forest and gradient boosting models, Bayesian hierarchical methods, and ensemble fusion frameworks (Di et al., 2021; Li et al., 2020; van Donkelaar et al., 2019; Chen et al., 2018; Hu et al., 2017; Safarov et al., 2025). These methods have substantially improved the spatial and temporal resolution of $PM_{2.5}$ estimates and have become increasingly important for epidemiological and environmental justice applications. However, many existing models are developed for broad regional or national scales and may inadequately represent localized pollution dynamics in mountainous terrain, where inversion frequency, valley confinement, elevation gradients, and wildfire smoke transport can create highly heterogeneous exposure patterns over short distances and timescales (O'Neill et al., 2021). Accurately characterizing these fine-scale processes remains particularly challenging in rural western regions with sparse regulatory monitoring coverage.

The recent proliferation of low-cost sensor networks, such as the citizen-science-driven PurpleAir platform ("PurpleAir Low-Cost PM2.5 Sensor Network Data", 2024), presents a promising opportunity to fill these observational gaps, enhancing the spatial density of existing regulatory instruments at a fraction of the cost. However, their raw data are subject to biases and require robust correction (Barkjohn et al., 2022; Jaffe et al., 2023). While a growing body of literature has focused on sensor calibration (Chu et al., 2020; Kang and Choi, 2024; Wang et al., 2019; Nana et al., 2025), a critical gap remains in understanding how to best integrate the dense data into operational air quality models to improve daily air quality mapping, especially in topographically complex settings. In addition to calibration challenges, integrating dense low-cost sensor observations into spatial prediction frameworks introduces methodological questions regarding spatial representativeness, temporal stability, sensor clustering, and spatial autocorrelation.

Despite major advances in $PM_{2.5}$ exposure modeling, relatively little work has evaluated how dense low-cost sensor networks can be integrated into daily spatial prediction frameworks in rural mountainous environments characterized by strong terrain-driven meteorological heterogeneity. In particular, it remains unclear whether dense citizen-science monitoring networks can substantially improve prediction of localized winter inversion events and wildfire smoke exposure beyond what is achievable using sparse regulatory monitoring networks alone.

To address this challenge, this study develops and evaluates a seasonally adaptive modeling framework for daily $PM_{2.5}$ mapping in Montana, USA. Unlike many $PM_{2.5}$ mapping approaches that apply a single modeling framework year-round, our approach explicitly evaluates seasonal differences in pollution dynamics by incorporating a terrain-based "smog potential" surface designed to capture winter stagnation behavior while also assessing the role of satellite-derived aerosol optical depth (AOD) during wildfire seasons. We investigate whether a dense, bias-corrected sensor network can improve predictions, particularly assessing the role of a terrain-based "smog potential" map for capturing winter inversion episodes. Ultimately, we test if this hybrid approach can provide a new pathway for enhancing air quality forecasting and risk communication in historically underserved rural and mountainous communities.

2. Data and methods

2.1. Study domain

Our study focused on Western Montana (bounding box: −114.9, −111.4, 44.7, 49.9) (Fig. 1), a USA state characterized by steep topographic gradients, frequent winter temperature inversions, spring and fall prescribed burning, and episodic wildfire smoke events. These features and unique sources create a particularly challenging environment for air quality monitoring and modeling.

2.2. Datasets

Our modeling framework integrated ground-based $PM_{2.5}$ measurements with ancillary meteorological and satellite data for the 2023 calendar year.

2.2.1. $PM_{2.5}$ monitoring data

We acquired hourly $PM_{2.5}$ observations from two primary sources. The first was the PurpleAir network, a citizen-science platform from which we compiled 10.8 million observations from 186 monitors (Fig. 1). The second was the U.S. Environmental Protection Agency's (EPA) Air Quality System (AQS), which serves as the regulatory standard. We obtained hourly concentrations from six EPA stations within the study domain (Fig. 1) from the EPA AirNow (https://www.epa.gov/outdoor-air-quality-data). Negative values were set to zero and 24-h averages were calculated in local time.

2.2.2. Ancillary datasets

To account for atmospheric conditions and land use patterns, we incorporated several predictor datasets. First, a daily smog intensity (smogI) metric was used to quantify large-scale meteorological conditions conducive to pollutant buildup. This metric, described by (Swanson et al., 2022), is a linear combination of geopotential height, longwave radiation and wind. Second, we obtained 1-km resolution MODIS Aerosol Optical Depth (AOD) data (MCD19A2), a widely used satellite-based proxy for columnar particulate matter. AOD was unavailable across the entire study area for much of the year due cloud and snow cover, so we applied standard quality masks and a nearest-neighbor interpolation approach to fill data gaps. Although crude, we found that our GWR method zeroes out the AOD coefficient under such conditions. Additional geophysical and land use predictors, detailed in Table 1 and Appendix, were used to model the spatial distribution of smog potential. We modeled $PM_{2.5}$ at daily resolution because air pollution processes in the Intermountain West are highly episodic and strongly influenced by day-specific meteorological conditions, including wildfire smoke transport, winter inversions, and boundary layer dynamics. Daily modeling allowed the framework to capture transient atmospheric behavior and acute pollution episodes while retaining stable geophysical predictors describing terrain and land-use influences. Daily estimates were also necessary for downstream exposure assessment and epidemiologic applications focused on short-term $PM_{2.5}$ variability.

Table 1
Predictor variables used in the Generalized Boosted Model (GBM) to create the smog potential (smogP) surface.

Predictor Category	Variable Description	Number of Predictors
Land Surface Temp	Monthly daytime mean MODIS LST	12
	Monthly nighttime mean MODIS LST	12
Topography	Local minima function (valley depth) at various radii	11
	Topographic position index (Jenness 2006) at various radii	5
Emission Proxies	4 km housing density (EPA surrogate)	1
	4 km rural housing density (EPA surrogate)	1
	4 km wood heat (EPA surrogate)	1
	Kernel-smoothed population density	1

2.3. PurpleAir data preprocessing

To transform the raw citizen-science data into research-grade product, we performed a three-step preprocessing workflow to (1) screen, (2)

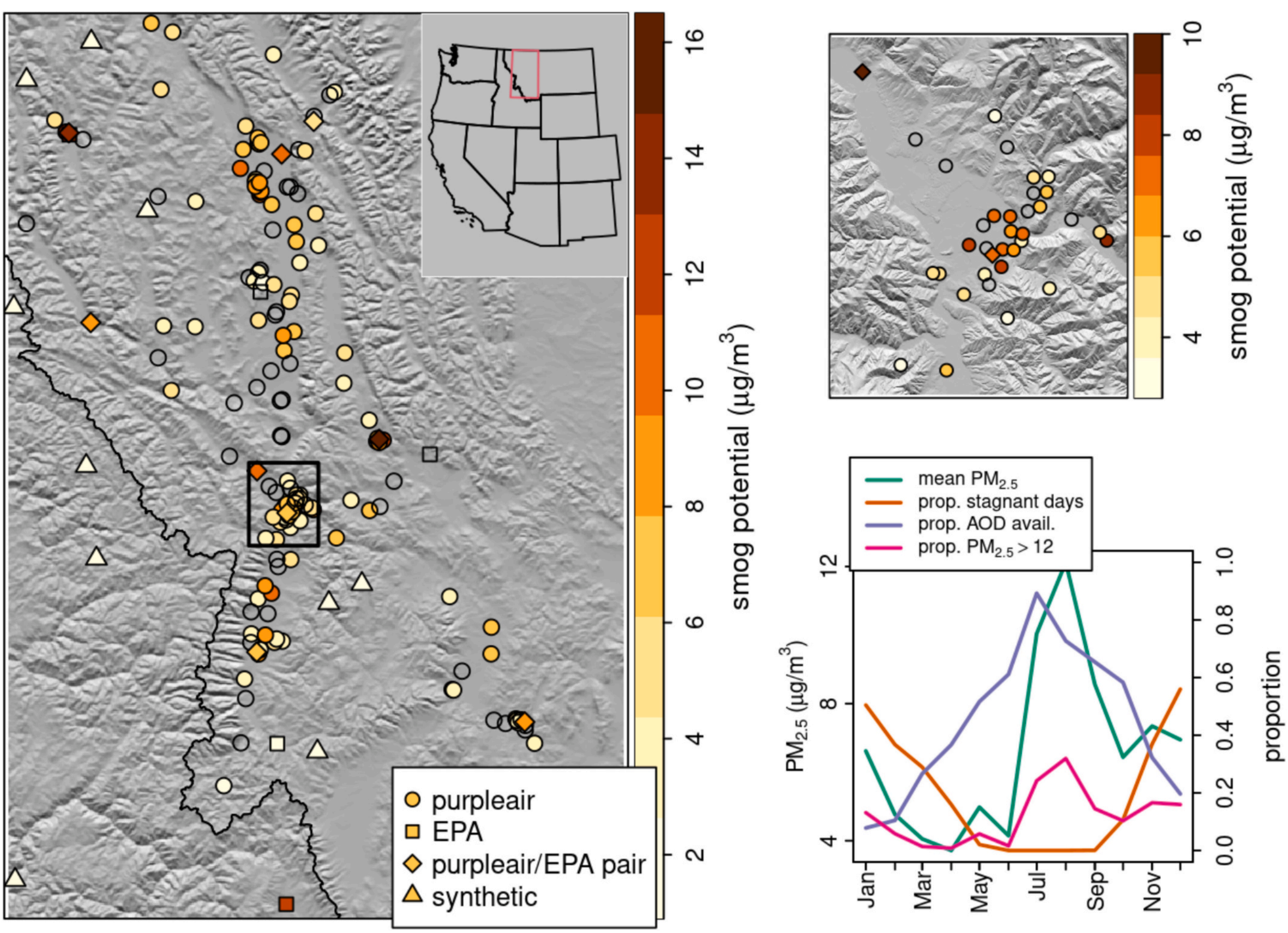


Fig. 1. Study Area and Monitoring Network in Western Montana. The study domain spans longitudes −114.9 to −111.4 and latitudes 44.7 to 49.9. Locations of the 186 PurpleAir low-cost sensors (circles) and six EPA AQS regulatory monitors (squares and diamonds) used in the analysis are shown. Point colors indicate smog potential or mean $PM_{2.5}$ under standardized "stagnant" conditions, open points did not have enough inversion days to provide a reliable estimate. (Top right) Inset of the Missoula valley illustrates how particulate matter pollution concentrates in populated valleys under stagnant conditions. (Bottom right) Mean monthly conditions averaged across the study area; $PM_{2.5}$ shows a bimodal pattern with peaks during fire season (July–September) and when stagnant conditions are dominant (November–March). Satellite AOD data, which requires cloud-free and snow-free conditions, becomes nearly unavailable in winter.

correct, and (3) harmonize the observations.

First, we removed anomalous observations exceeding 250 μg/m^3, a conservative threshold based on the highest valid hourly concentrations recorded in the regional EPA AQS dataset during the study period. The screened hourly data were then aggregated to daily means, requiring a minimum of 23 valid hourly observations (96% completeness) to ensure a representative daily value. This step reduced the influence of short-term fluctuations and aligns the data with the 24-h average exposure metrics common in health studies.

Previous work has identified systematic biases in PurpleAir and other low-cost sensors (Barkjohn et al., 2022; Jaffe et al., 2023). We correct systematic errors in our data using co-located EPA-PurpleAir pairs and fitting a machine learning model with humidity and temperature as predictors.

Third, we reconciled closely spaced observations by calculating a weighted average of the daily values of monitors located within 600-m of each other, assigning PurpleAir monitors a 33% weight and EPA monitors a 67% weight. Finally, because $PM_{2.5}$ concentrations are skewed toward large values, we applied a log-transformation ($y = \ln(PM_{2.5} + 1)$) to all data to stabilize the variance and normalize residuals for subsequent modeling.

2.4. Modeling approach

We used a two-stage approach: first, we developed a static surface representing the long-term potential for winter smog accumulation (smogP), and second, we used this surface as a predictor in a dynamic model to generate daily $PM_{2.5}$ estimates.

2.4.1. Smog potential (smogP) surface model

To represent the static, topographically driven propensity for pollution buildup during stagnant winter conditions, we developed a 600-m resolution smog potential (smogP) surface (Swanson et al., 2022). The smogP metric was intended to characterize the long-term tendency of a location to accumulate $PM_{2.5}$ during winter stagnation events, independent of day-to-day meteorological variability.

We first identified the top 20% most stagnant winter days using the smogI index. For each monitoring location, we then calculated the mean $PM_{2.5}$ concentration across these stagnant days. This monitor specific stagnant day mean $PM_{2.5}$ value served as the response variable for subsequent spatial modeling. Sensors with fewer than 50 valid stagnant day observations were excluded from analysis.

We modeled these smogP values across the study area using a hybrid machine learning and geostatistical framework. First, a Gradient Boosted Model (GBM) (Jerome and Friedman, 2001) was trained using predictor variables representing topographic confinement, land use, and emissions related characteristics (Table 1). GBM was selected because of its ability to capture complex nonlinear relationships and interactions among predictors.

Because GBM does not explicitly account for spatial autocorrelation, residuals from the GBM predictions were subsequently modeled using ordinary kriging. An exponential variogram model was fitted to the sample semi variogram using least squares estimation, with nugget effect treated as measurement error. The kriged residual surface was then added back to the GBM predictions to generate the final smogP surface. This hybrid approach combines the nonlinear predictive capability of machine learning with the ability of geostatistical interpolation to represent fine-scale spatial structure not captured by the predictor variables alone.

Model selection was performed using Forward Feature Selection (FFS) implemented in the R package CAST (Meyer et al., 2018), using root mean square error as the optimization metric. Five-fold spatial cross-validation was used during model tuning to reduce overfitting and improve spatial generalizability. Final model performance was using leave-one-out cross-validation (LOOCV) where all observations associated with a monitoring location were withheld in turn and predicted using models trained on the remaining stations. Spatial autocorrelation in model residuals was evaluated using Moran's I test with inverse distance weighting (Moran, 1950).

Finally, preliminary models were found to overpredict smog potential in high elevation mountainous terrain far from populated valleys, where monitoring data were sparse, but air quality is generally known to remain clean during winter stagnation events. To constrain extrapolation in these regions, we added 10 ‘synthetic’ training points at randomly selected high-elevation, low-population locations and assigned them a smog potential of 1 μg/m^3. These synthetic points were included during model training but excluded from validation analysis.

2.4.2. Daily $PM_{2.5}$ surface generation and evaluation

To generate daily 600-m resolution $PM_{2.5}$ surfaces, we used Geographically Weighted Regression (GWR) applied daily. A 600-m grid resolution was selected as a balance between the fine-scale spatial variability imposed by complex valley terrain and the spatial density of the PurpleAir monitoring network, while also remaining computationally tractable for daily surface generation across the study domain. This resolution was sufficiently fine to resolve major topographic gradients and urban-scale pollution variability without implying unsupported sub-grid precision. Hourly $PM_{2.5}$ observations were aggregated into daily averages using local standard time at each monitoring location. This method was chosen because the relationships between $PM_{2.5}$ and its spatial predictors (e.g., AOD, smogP) are non-stationary in complex terrain; that is, the strength of their effects are assumed to vary across the study area and from day to day. GWR was chosen because the relationships between $PM_{2.5}$ and its spatial predictors (e.g., AOD, smogP) are non-stationary in complex terrain; that is, the strength of their effects are assumed to vary across the study area and from day to day. For each day, we fit multiple GWR models with different combinations of predictors (AOD, smogP, smogP$_{+}$AOD) and compared them to a null model that used spatial interpolation alone.

Model performance was assessed using a daily LOOCV. This robust method provides an unbiased estimate of how well the model predicts $PM_{2.5}$ at unmonitored locations by iteratively withholding each station, refitting the daily GWR model, and comparing the prediction to the observed value. We used gaussian weights and an adaptive bandwidth, with the bandwidth for each model chosen to maximize R^2 over the full set of LOOCV predictions. To rigorously assess prediction of accuracy at unsampled locations, we used LOOCV estimates of smog potential, such that each point was withheld from the entire modeling process (both the static surface generation and the daily GWR).

To compare with the results of Swanson et al. (2022), we also predicted the PurpleAir stations with daily models fitted using only EPA observations and using both the 2022 and newly developed smog potential models. All EPA monitoring stations within 250-km of the nearest study site were included in the GWR runs but excluded from smogP model fitting and all validation. Points outside of the study area but were included in the fitting of the daily GWR. Smog potential estimates at these points were made using predictions from the full GBM and kriged error, using an exponential variogram model fitted to this expanded dataset.

3. Results

3.1. GBM bias correction substantially improves low-cost sensor accuracy

Our data included 11 PurpleAir sensors collocated (< 120-m, but generally <3 -m) with 6 EPA reference stations, consisting of 108,000 hourly observations. Raw PurpleAir measurements exhibited systematic, concentration-dependent bias when compared against co-located EPA regulatory monitors. At lower concentrations (< 10 μg/m^3), PurpleAir sensors tended to overestimate $PM_{2.5}$, while underestimating at higher concentrations (Fig. 2). Applying our machine learning bias correction resolved this issue and substantially improved the agreement

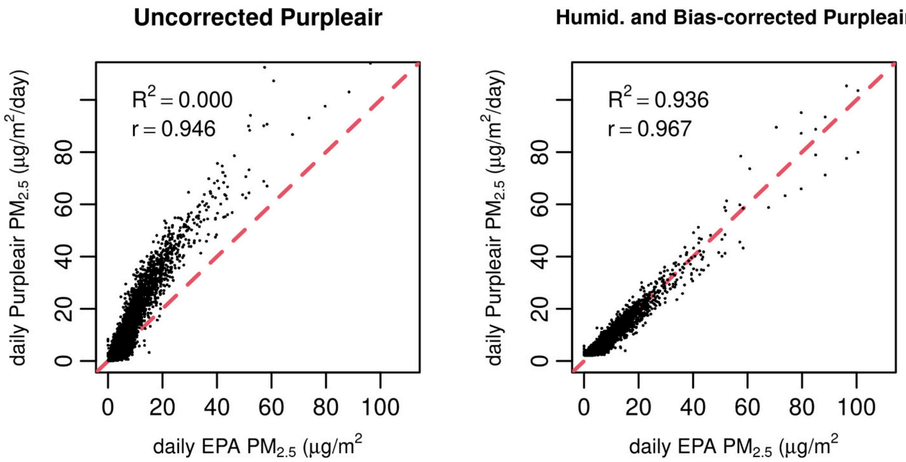


Fig. 2. Improvement of Sensor Accuracy After Bias Correction. Comparison of daily PurpleAir $PM_{2.5}$ measurements with collocated EPA regulatory monitor observations before correction (left) and after humidity and bias correction (right). The red dashed line indicates the 1:1 relationship. Correlation coefficients and R^2 values are shown for each comparison.

between low-cost sensors and the reference monitors. For hourly data, the R^2 increased from negligible to 83.8%, and for daily means, it increased from negligible to 93.6%. This compares favorably to the national correction developed by (Barkjohn et al., 2022), which achieved R^2 values of 74.9% and 82.9% for hourly and daily data, respectively, when applied to the same observations (Fig. 2).

3.2. Hybrid model captures terrain and emission-driven smog potential

Our hybrid machine learning and geostatistical model successfully predicted the spatial patterns of winter pollution potential (smogP), explaining 59.5% of the variance at the 109 station locations for which we were able to obtain empirical estimates of smogP, under strict LOOCV validation (Fig. 5). Although many monitoring locations were spatially clustered, additional evaluation at "remote" sites located >10 km from their nearest neighboring monitor (Tables 3 and 4) demonstrating that model performance remained robust even under reduced spatial redundancy, supporting the generalizability of the model beyond densely monitored areas. The underlying GBM model, which accounted for 56.4% of the variance, identified population density and a metric of terrain confinement as the dominant predictors; together, these top 2 predictors accounted for 88% of the model's relative influence (Table 2, Fig. 4, Fig. S2).

Residual errors from the full data GBM smogP fit were found to have significant spatial autocorrelation (IDW Moran's I *p*-value = 0.002). The exponential variogram model fitted to the sample semi-variogram had an effective range of 250-km, approximately the width of the study area (Fig. 3). Correcting for the model's spatially autocorrelated residuals via kriging slightly improved the final map's accuracy and spatial representation of smog potential (Figs. 4 & 5). We hypothesized that the inclusion of the dense PurpleAir network would yield improved predictions of smogP to new locations within the study area, and this was partially confirmed. When evaluated at PurpleAir-only locations,

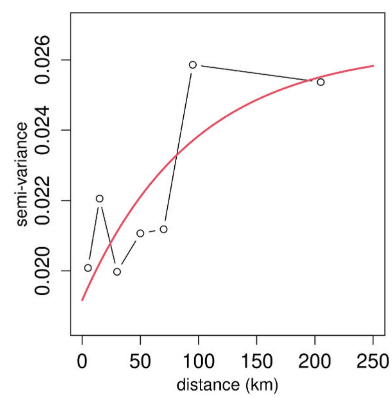


Fig. 3. Spatial Autocorrelation in Smog Potential Model Residuals. Variogram of residual errors from the static smogP GBM fit. The red line represents the fitted exponential model, which has an estimated effective range of 250-km, indicating that errors are correlated up to this distance.

the new smogP model achieved a correlation coefficient of 0.732 under strict LOOCV, compared to 0.651 for the 2022 model (Swanson et al., 2022). Importantly, model performance remained consistent at remote monitoring locations (>10 km from the nearest neighboring monitor), indicating that the observed validation skill was not solely driven by closely spaced sensor clusters (Tables 3 and 4).

As an additional assessment of our new smogP, we generated LOOCV predictions using both the current smogP surface and the smogP surface of Swanson et al. (2022), using only the permanent EPA monitoring stations to fit daily GWR models (Table 5). We found that the two versions were roughly equivalent in winter, but the older predictor did slightly better on an annual basis when constrained to sparse input data (Table 5).

Table 2
Relative influence of predictors in the smogP GBM fit.

Predictors	Relative Influence (%)	Description
Local minima function, 1000 m	47.3	Valley Bottom Position (1000-m radius)
Population density	40.6	Kernel-smoothed Population Density (2020)
Jan nighttime LST	7.3	Mean Nighttime Land Surface Temperature
RuralPopulation	4.9	Rural Housing Density

3.3. PurpleAir sensors increase prediction accuracy

The inclusion of PurpleAir sensors in daily modeling of $PM_{2.5}$ dramatically increased prediction accuracy, raising the full model ($smogP_{+}AOD$) R^2 from 0.692 (Table 4) to 0.837 (Table 3). This substantial gain is attributable to the increase in data density; the median number of observations used to fit the daily GWR increased from 11 to 98 with the inclusion of the PurpleAir data. With the high density of PurpleAir sensors, the relative importance of predictor covariates diminished; an interpolation-only model achieved an R^2 of 0.811. This

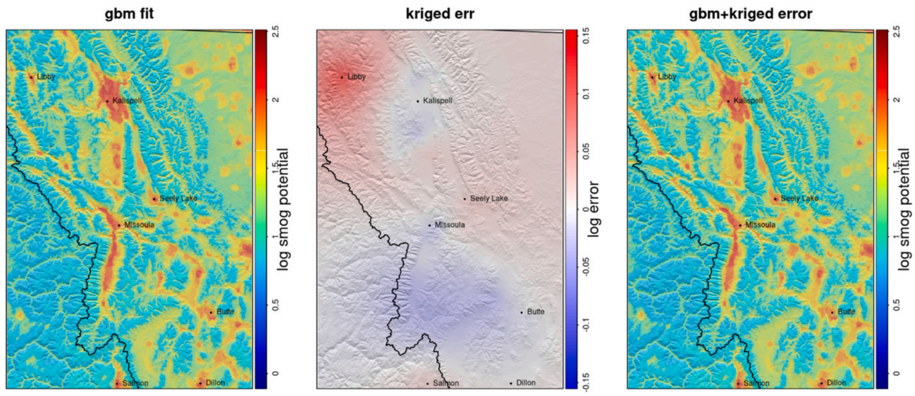


Fig. 4. Hybrid Smog Potential Surface. Static smog potential (smogP) results from the GBM. (Left) Initial GBM fit. (Center) Kriged residual error from the GBM fit, where positive values indicate locations with higher observed smog than predicted. (Right) Final combined surface (GBM fit and kriged error), showing improved spatial representation of smog potential across the region.

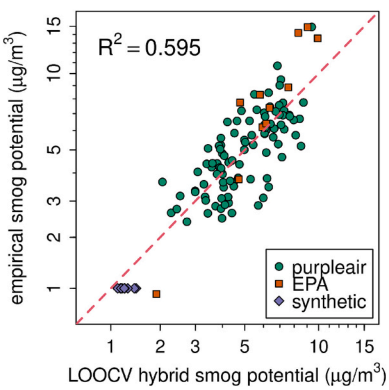


Fig. 5. Smog Potential Model Performance. Scatterplot comparing smogP model predictions (GBM fit with kriged error) against empirical smogP estimates at monitoring locations. Empirical smogP is defined as average $PM_{2.5}$ under standardized, stagnant, winter conditions. The 1:1 line is shown in black. Using strict LOOCV validation, we found that predictions of this hybrid model, excluding the synthetic points, explained 59.5% of the variance, while the GBM alone explained 56.4%.

Table 3
Full data prediction R^2. Results of daily GWR using both EPA and PurpleAir data as input, evaluated at all locations using strict LOOCV validation. 'Remote' refers to evaluation at stations with no neighbors within 10 km. Satellite AOD offers limited improvement over interpolation, even for remote locations during fire season. In contrast, the smogP predictor substantially improves winter predictions of $PM_{2.5}$.

Model	Full Year	Remote	Winter	Fire	Winter + Remote	Fire + Remote
Interpolation	0.811	0.723	0.359	0.825	0.152	0.724
AOD	0.809	0.732	0.342	0.835	0.114	0.718
smogP	0.833	0.756	0.535	0.813	0.368	0.716
AOD+smogP	0.837	0.758	0.535	0.829	0.363	0.730

result held during fire season, where covariates provided only a modest improvement in fit over interpolation alone. However, in winter, the smogP covariate contributed substantially, increasing the R^2 from 0.359 (interpolation) to 0.535 (smogP).

The full model performed very well at classifying $PM_{2.5}$ concentration by EPA categories (Table 6), with 94.5% correctly classified and a Kappa statistic of 0.741. A map of accuracy by sensor location (Fig. 7) shows higher errors in locations farther from their neighbors and indicates that winter errors have a more skewed distribution. Further analysis found that the stations with the highest winter error tended to also have large errors in the smogP model under LOO validation or were sites without enough observations to make a valid empirical estimate of smogP.

Table 4
EPA-only prediction R^2. Results of daily GWR using only EPA data as input, evaluated at all locations using strict LOOCV validation. With fewer observations (typically 10–12 within the study region) to drive daily GWR predictions, all models perform worse than when PurpleAir data is included, while the covariates become more powerful. Satellite AOD improves fire-season and full-year predictions, while the smogP predictor continues to greatly improve winter predictions.

Model	Full Year	Remote	Winter	Fire	Winter + Remote	Fire + Remote
Interpolation	0.618	0.581	0.167	0.640	0.080	0.638
AOD	0.675	0.602	0.167	0.725	0.040	0.683
smogP	0.637	0.616	0.373	0.616	0.218	0.603
AOD+smogP	0.692	0.645	0.360	0.699	0.234	0.642

Table 5
SmogP comparison R^2. Comparison of EPA-only fits evaluated at new PurpleAir locations to assess the 2022 vs. new smogP predictor. The new predictor, developed using PurpleAir and EPA data, failed to improve on the predictor used by Swanson et al., 2022.

Model	Full Year	Winter
New smogP, EPA Input	0.682	0.327
2022 smogP, EPA Input	0.711	0.326
New smogP, Full Input	0.845	0.510
2022 smogP, Full Input	0.848	0.520

Table 6
Confusion matrix for LOOCV predictions (Full AOD + smogP model). PCC = 0.945; Kappa = 0.741.

	Observed			
Predicted	Good	Moderate	Unhealthy-sens.	Unhealthy
Good	56,976	2139	12	5
Moderate	1094	5377	174	14
Unhealthy-sens.	3	94	243	45
Unhealthy	1	3	39	208

3.4. *Optimal $PM_{2.5}$ prediction strategy season-dependent*

The most striking result of our daily $PM_{2.5}$ predictions was a strong seasonal dichotomy in model performance, mirroring the region's bimodal annual pollution cycle (Figs. 1 & 6). No single-predictor model was optimal for the entire year; instead, the best approach depended on the dominant seasonal pollution source. During the winter inversion season (November–January), models that included the smogP layer were essential and consistently outperformed all other model configurations (Table 3). This demonstrates the importance of a terrain-aware model for capturing topographically trapped, local-source pollution.

Conversely, during the summer wildfire season (July–September), the simpler GWR model using only spatial interpolation of the dense sensor data was highly effective. This "null" model ($R^2 = 0.825$) was on par with the more complex covariate-based approaches (best model $R^2 = 0.835$), even in remote areas ($R^2 = 0.724$ vs. full model $R^2 = 0.730$).

Daily predictions using only the EPA stations showed stronger covariate effects, with AOD increasing prediction R^2 over the interpolation model from 0.640 to 0.725 during fire season, and smogP increasing R^2 from 0.167 to 0.373 during winter.

4. Discussion

Our study demonstrates the potential for low-cost/citizen science sensors to significantly improve $PM_{2.5}$ exposure modeling, particularly in regions with sparse regulatory monitoring data and complex mountainous terrain. However, our results also highlight the distinct seasonally driven processes that govern air pollution in mountains and the potential for novel modeling approaches to more accurately resolve the spatial patterns of pollution exposure across the landscape. We show that a static covariate based on topography and population density can substantially improve predictions of air quality in valleys associated with winter inversion events. During other parts of the year, we found that with the higher spatial density afforded by these data, relatively simple statistical interpolation performed nearly as well as models that included ancillary predictors like satellite-derived aerosol optical depth data. This finding is particularly important given that MODIS data is widely used for $PM_{2.5}$ modeling is seasonally unavailable (Fig. 1) in our study region due to snow and cloud cover and will soon be unavailable globally due to the retirement of the MODIS satellites. A caveat here is that our study spanned two relatively mild fire seasons – with more wildfire smoke we expect AOD to be more important as a predictor. Our approach provides a powerful tool for regions where valley-level pollution gradients, often invisible to coarser models, pose a significant public health risk.

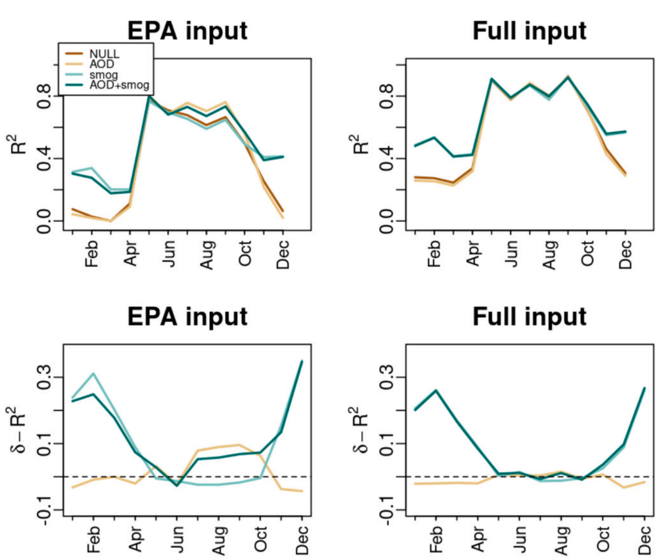


Fig. 6. Seasonal Dichotomy in Daily Model Performance. Performance of daily GWR fits under strict LOOCV validation, aggregated by month. The top row shows R^2, and the bottom shows R^2 relative to the interpolation-only model. The first column shows the results of daily models fit using only the permanent reference stations but evaluated at all sites. The second column shows daily GWR models fit using the full data and evaluated at all sites. Clear performance gains are visible for all models containing smogP during winter (November–March), while AOD improves over interpolation-only models during fire season (July–September), but only for models fit using sparse input from permanent EPA monitoring stations.

The most significant implication of our study is that for complex terrain, there is no universal "best model." The success of our smogP-driven models in winter highlights that when air quality is governed by local emissions, stagnant meteorology, and topographic trapping, a model informed by these static features is superior. In contrast, summer air quality is often dominated by large-scale, synoptic wildfire events. In these scenarios, pollution patterns are less constrained by local terrain, and the primary value of the low-cost sensor network is its ability to capture the real-time spatial footprint of a smoke plume. For air quality managers and forecasters, this suggests that operational models should be dynamic, switching from a terrain-based framework in the winter to a more flexible, interpolation-based system during wildfire season to provide the most accurate public health guidance. Our daily GWR approach facilitates this by leveraging the data itself to more heavily weight the seasonally relevant predictor.

Our new smogP predictor developed using PurpleAir and EPA data failed to outperform the predictor of Swanson et al., 2022 (Table 5), which was made using EPA-only data over a wider geographic area. The old predictor was better at estimating $PM_{2.5}$ under inversion conditions. This could be attributable to the shorter record of the new PurpleAir data leading to more uncertain estimates of smogP, and the smaller number of station locations ($n = 115$) than were used in the 2022 model ($n = 198$). With such a small sample size, every observation is important, which was highlighted in validation. A modeling run using full-data predictions of smogP outperformed the run using the 2022 predictor. The fact that the 2022 predictor, fit using data from a much broader area, outperformed a locally calibrated predictor suggests universal mechanisms by which terrain and human activity affect $PM_{2.5}$ distribution.

While this study provides a robust framework, it has limitations that open clear avenues for future research. The sensor network, though dense, is biased toward populated areas, and our analysis is limited to two years of data. Future work should expand this seasonally adaptive framework to other mountainous regions and over longer time periods. Furthermore, incorporating dynamic variables such as woodstove density data and active fire locations could further refine source attribution and predictive accuracy. Ultimately, the methodology presented here provides a clear, actionable pathway for transforming citizen-science data into a cornerstone of modern air quality management, enabling more targeted and effective public health interventions in the communities most vulnerable to the dual threats of winter inversions and wildfire smoke.

Abbreviations

$PM_{2.5}$	Fine Particulate Matter
AOD	Aerosol Optical Depth
EPA	Environmental Protection Agency
AQS	Air Quality System
smogI	Smog Intensity
smogP	Smog Potential
GBM	Generalized Boosted Model
FFS	Forward Feature Selection
LOOCV	Leave-One-Out Cross-Validation
GWR	Geographically Weighted Regression
CFSR	Climate Forecast System Reanalysis
MAE	Mean Absolute Error

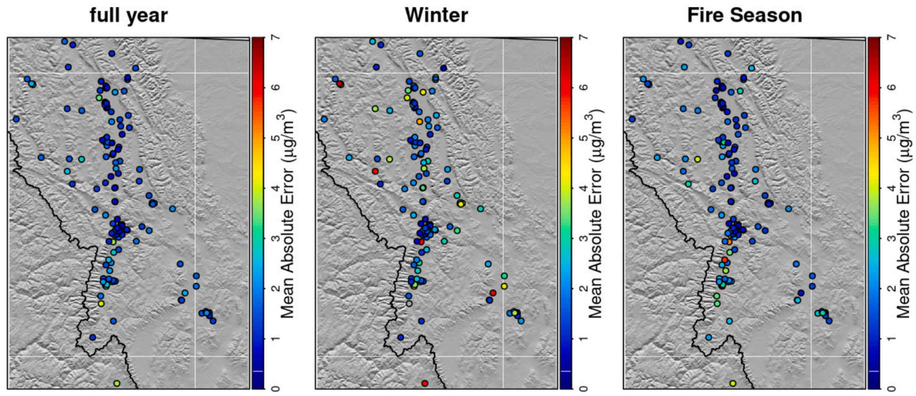


Fig. 7. Map of MAE (Mean Absolute Error) by S. Errors for the model using all data as input to the daily GWR. The largest winter errors tend to occur at locations where smogP was poorly estimated by our model. The correlation coefficient between MAE and absolute LOOCV error in the smogP model was 0.62.

CRediT authorship contribution statement

Alan Swanson: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ava Orr:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Zachary A. Holden:** Writing – review & editing, Conceptualization. **Kyle Bocinsky:** Writing – review & editing, Conceptualization. **Erin L. Landguth:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This material is based upon work supported in part by the National Science Foundation EPSCoR Cooperative Agreement OIA-2242802, the National Institute of General Medical Sciences of the National Institutes of Health (P20GM130418), and computational resources from the University of Montana's Computational Ecology Lab, Center for Population Health's Data and Modeling Core (P20GM130418), and HellGate Shared Computing Cluster (Grant No 2018112 & 1925267). The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2026.181915.

Data availability

Data will be made available on request.

References

Barkjohn, K.K., Holder, A.L., Frederick, S.G., Clements, A.L., 2022. Correction and accuracy of PurpleAir PM(2.5) measurements for extreme wildfire smoke. Sensors (Basel) 22.

Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J., Guo, Y., 2018. A machine learning method to estimate PM(2.5) concentrations across China with remote sensing, meteorological and land use information. Sci. Total Environ. 636, 52–60.

Chen, Gongbo, Yuming Guo, Xu, Yue, Shilu Tong, Gasparrini, Antonio, Bell, Michelle L., Armstrong, Ben, Schwartz, Joel, Jaakkola, Jouni J.K., Zanobetti, Antonella, Lavigne, Eric, Saldiva, Paulo Hilario Nascimento, Kan, Haidong, Royé, Dominic, Milojevic, Ai, Overcenco, Ala, Urban, Aleš, Schneider, Alexandra, Entezari, Alireza, Vicedo-Cabrera, Ana Maria, Zeka, Ariana, Tobias, Aurelio, Nunes, Baltazar, Alahmad, Barrak, Forsberg, Bertil, Pan, Shih-Chun, Íñiguez, Carmen, Ameling, Caroline, De la Cruz, César, Valencia, Christofer Åström, Houthuijs, Danny, Van Dung, Do, Samoli, Evangelia, Mayvaneh, Fatemeh, Sera, Francesco, Carrasco-Escobar, Gabriel, Lei, Yadong, Orru, Hans, Kim, Ho, Holobaca, Iulian-Horia, Kyselý, Jan, Teixeira, João Paulo, Madureira, Joana, Katsouyanni, Klea, Hurtado-Díaz, Magali, Maasikmets, Marek, Ragettli, Martina S., Hashizume, Masahiro, Stafoggia, Massimo, Pascal, Mathilde, Scortichini, Matteo, de Sousa, Micheline, Coêlho, Zanotti Stagliorio, Ortega, Nicolás Valdés, Ryti, Niilo R.I., Scovronick, Noah, Matus, Patricia, Goodman, Patrick, Garland, Rebecca M., Abrutzky, Rosana, Garcia, Samuel Osorio, Rao, Shilpa, Fratianni, Simona, Dang, Tran Ngoc, Colistro, Valentina, Huber, Veronika, Lee, Whanhee, Seposo, Xerxes, Honda, Yasushi, Guo, Yue Leon, Ye, Tingting, Wenhua, Yu, Abramson, Michael J., Samet, Jonathan M., Li, Shanshan, 2021. Mortality risk attributable to wildfire-related PM_{2·5} pollution: a global time series study in 749 locations. Lancet Planet. Heath 5 e579-e87.

Chu, Hone-Jay, Ali, Muhammad Zeeshan, He, Yu-Chen, 2020. Spatial calibration and PM2.5 mapping of low-cost air quality sensors. Sci. Rep. 10, 22079.

DeFlorio-Barker, S., Crooks, J., Reyes, J., Rappold, A.G., 2019. Cardiopulmonary effects of fine particulate matter exposure among older adults, during wildfire and non-wildfire periods, in the United States 2008-2010. Environ. Health Perspect. 127, 37006.

Di, Q., Wei, Y., Shtein, A., Hultquist, C., Xing, X., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J.T., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., 2021. Daily and Annual PM2.5 Concentrations for the Contiguous United States, 1-km Grids, v1 (2000–2016). NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, New York.

Dockery, D.W., Pope, C.A., 1994. Acute respiratory effects of particulate air pollution. Annu. Rev. Public Health 15, 107–132.

van Donkelaar, Aaron, Martin, Randall V., Li, Chi, Burnett, Richard T., 2019. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. Environ. Sci. Technol. 53, 2595–2611.

Ebisu, Keita, Malig, Brian, Hasheminassab, Sina, Sioutas, Constantinos, 2019. Age-specific seasonal associations between acute exposure to PM2.5 sources and cardiorespiratory hospital admissions in California. Atmos. Environ. 218, 117029.

Hadley, Odelle, Cutler, Anthony, Schumaker, Ruth, Bond, Robin, 2021. Wildfires and wood stoves: woodsmoke toxicity and chemical characterization study in the north-western United States. Atmos. Environ. 253, 118347.

Horne, B.D., Joy, E.A., Hofmann, M.G., Gesteland, P.H., Cannon, J.B., Lefler, J.S., Blagev, D.P., Korgenski, E.K., Torosyan, N., Hansen, G.I., Kartchner, D., Pope 3rd., C. A., 2018. Short-term elevation of fine particulate matter air pollution and acute lower respiratory infection. Am. J. Respir. Crit. Care Med. 198, 759–766.

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM(2.5) concentrations in the conterminous United States using the random forest approach. Environ. Sci. Technol. 51, 6936–6944.

Jaffe, D.A., Miller, C., Thompson, K., Finley, B., Nelson, M., Ouimette, J., Andrews, E., 2023. An evaluation of the U.S. EPA's correction equation for PurpleAir sensor data in smoke, dust, and wintertime urban pollution events. Atmos. Meas. Tech. 16, 1311–1322.

Jainonthee, C., Wang, Y.L., Chen, C.W.K., Jainontee, K., 2022. Air pollution-related respiratory diseases and associated environmental factors in Chiang Mai, Thailand, in 2011–2020′. Trop. Med. Infect. Dis. 7.

Jeong, S., Park, S.A., Park, I., Kim, P., Cho, N.H., Hyun, J.W., Hyun, Y.M., 2019. PM2.5 exposure in the respiratory system induces distinct inflammatory signaling in the lung and the liver of mice. J. Immunol. Res. 2019, 3486841.

Jerome, Friedman, H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.

Kang, J., Choi, K., 2024. Calibration methods for low-cost particulate matter sensors considering seasonal variability. Sensors (Basel) 24.

Li, Jiayu, Zhang, Huang, Chao, Chun-Ying, Chien, Chih-Hsiang, Chang-Yu, Wu, Luo, Cyuan Heng, Chen, Ling-Jyh, Biswas, Pratim, 2020. Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM2.5. Atmos. Environ. 223, 117293.

Liu, Xiaoxi, Huey, L. Gregory, Yokelson, Robert J., Selimovic, Vanessa, Simpson, Isobel J., Müller, Markus, Jimenez, Jose L., Campuzano-Jost, Pedro, Beyersdorf, Andreas J., Blake, Donald R., 2017. Airborne measurements of western US wildfire emissions: comparison with prescribed burning and air quality implications. J. Geophys. Res. Atmos. 122, 6108–6129.

Magzamen, S., Gan, R.W., Liu, J., O'Dell, K., Ford, B., Berg, K., Bol, K., Wilson, A., Fischer, E.V., Pierce, J.R., 2021. Differential cardiopulmonary health impacts of local and long-range transport of wildfire smoke. GeoHealth 5, e2020GH000330.

Maji, Kamal J., Li, Zongrun, Hu, Yongtao, Vaidyanathan, Ambarish, Stowell, Jennifer D., Milando, Chad, Wellenius, Gregory, Kinney, Patrick L., Russell, Armistead G., Odman, M. Talat, 2024. Prescribed burn related increases of population exposure to PM2.5 and O3 pollution in the southeastern US over 2013–2020. Environ. Int. 193, 109101.

Meyer, Hanna, Reudenbach, Christoph, Hengl, Tomislav, Katurji, Marwan, Nauss, Thomas, 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Software 101, 1–9.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Nana, Bernard, Raheja, Garima, Ouarma, Issoufou, Kayaba, Haro, Gounkaou, Woro Yomi, Daho, Tizane, Béré, Antoine, Mellouki, Abdelwahid, Westervelt, Daniel M., 2025. Monitoring of PM2.5 using well-calibrated low-cost sensors over one year in Burkina Faso. ACS ES&T Air 2, 40–48.

O'Neill, S.M., Diao, M., Raffuse, S., Al-Hamdan, M., Barik, M., Jia, Y., Reid, S., Zou, Y., Tong, D., West, J.J., Wilkins, J., Marsha, A., Freedman, F., Vargo, J., Larkin, N.K., Alvarado, E., Loesche, P., 2021. A multi-analysis approach for estimating regional health impacts from the 2017 Northern California wildfires. J. Air Waste Manag. Assoc. 71, 791–814.

Orr, Ava, Alden, Nisha B., Austin, Elizabeth, Jaffar, Zeina, Knudson, Jonathon, Graham, Jon, Miglaccio, Christopher T., Noonan, Curtis, Urbanski, Shawn, Wegener, Maximilian, Landguth, Erin L., 2025a. Wildfire-season fine particulate matter exposure and associations with influenza and influenza-like-illness risk in the Western USA. Environ. Health Perspect. https://doi.org/10.1021/EHP.6c00326.

Orr, Ava, Adam Claire, E., Graham, Jon, Holden Zachary, A., Hu, Lu, Jaffar, Zeina, Leary, Cindy, Migliaccio Christopher, T., Mullan, Katrina, Noonan, Curtis, Semmens Erin, O., Urbanski, Shawn, Walker, Ethan, Landguth Erin, L., 2025b. A state of the science review of wildfire-specific fine particulate matter data sources, methods, and models. Environ. Health Perspect. 133 (6), 66001. https://doi.org/10.1289/EHP15672.

PurpleAir (Ed.), 2024. PurpleAir Low-Cost PM2.5 Sensor Network Data. PurpleAir.

Safarov, R., Shomanova, Z., Nossenko, Y., Kopishev, E., Bexeitova, Z., Kamatov, R., 2025. Filling gaps in PM2.5 time series: a broad evaluation from statistical to advanced neural network models. PloS One 20, e0330211.

Swanson, Alan, Holden, Zachary A., Graham, Jon, Warren, D. Allen, Noonan, Curtis, Landguth, Erin, 2022. Daily 1 km terrain resolving maps of surface fine particulate matter for the western United States 2003–2021. Sci. Data 9, 466.

Wang, Yanwen, Du, Yanjun, Wang, Jiaonan, Li, Tiantian, 2019. Calibration of a low-cost PM2.5 monitor using a random forest model. Environ. Int. 133, 105161.

Ward, T., Lange, T., 2010. The impact of wood smoke on ambient PM2.5 in northern Rocky Mountain valley communities. Environ. Pollut. 158, 723–729.

Wu, K., Ho, H.C., Su, H., Huang, C., Zheng, H., Zhang, W., Tao, J., Hossain, M.Z., Zhang, Y., Hu, K., Yang, M., Wu, Q., Xu, Z., Cheng, J., 2022. A systematic review and meta-analysis of intraday effects of ambient air pollution and temperature on cardiorespiratory morbidities: first few hours of exposure matters to life. EBioMedicine 86, 104327.