Final Draft Report

October 13, 2025

High Spatiotemporal Resolution $PM_{2.5}$ Speciation Exposure Modeling in California

Principal Investigator:

Jun Wu

Department of Environmental & Occupational Health

856 Health Sciences Rd (Quad), #3082

University of California, Irvine, Irvine, CA 92697

Prepared by:

Qingqing He[1],
Meredith Franklin[2],
Veronica Berrocal[3],
Jun Wu[1]

[1] Department of Environmental & Occupational Health, University of California, Irvine, Irvine, CA 92697, United States

[2] Department of Statistical Sciences and School of the Environment, University of Toronto, Toronto, Canada

[3] Department of Statistics, University of California, Irvine, Irvine, CA 92697, United States

Disclaimer

Acknowledgments

We would like to express our heartfelt gratitude to all those who supported this project. In particular, we thank the graduate students, postgraduate students and research scholars whose hard work and dedication greatly contributed to the data collection, analysis, and modeling described in this report. Their efforts and insights were essential to the progress of this study. The student and scholar team included:

Mengyi Li, Postgraduate Student, Department of Environmental & Occupational Health, University of California, Irvine

Joseph Hotz, graduate Student, Department of Statistical Sciences, University of Toronto

Zichun Xu, graduate Student, Department of Statistical Sciences, University of Toronto

Table of Contents

LIST of TABLES

LIST OF FIGURES

Abstract

We developed a spatiotemporal modeling framework based on a deep-forest algorithm to estimate daily concentrations of five major $PM_{2.5}$ components—sulfate (SO_4^{2-}), nitrate (NO_3^-), elemental carbon (EC), organic carbon (OC), and mineral dust (DUST)—across the western United States from 2002 to 2019 at 1-km resolution. The framework first generated gap-free total $PM_{2.5}$ fields using MAIAC aerosol optical depth (AOD) and ancillary predictors, then incorporated these estimates and CMAQ-speciated simulations, together with meteorological and land-use variables, to predict component concentrations. Cross-validation against ground observations showed strong performance (R^2 = 0.81, 0.89, 0.75, 0.66, and 0.75; RMSE = 0.30, 0.59, 0.26, 1.52, and 0.59 μg/m³ for SO_4^{2-}, NO_3^-, EC, OC, and DUST, respectively). The results reproduced realistic spatial and temporal patterns, capturing both long-term declines and episodic wildfire extremes. We extended the record for California to 2000–2020 by integrating MERRA-2 and MERRA-2 GMI reanalysis predictors for years without CMAQ data. Persistent exposure hotspots were identified in the Los Angeles Basin and San Joaquin Valley, dominated by OC and NO_3^-. Population-weighted $PM_{2.5}$ exposure declined by about 20 % from 2000 to 2020, driven primarily by ~50 % reductions in NO_3^- and SO_4^{2-}, while OC and EC decreased modestly and DUST remained nearly unchanged. Disparity analyses using CalEnviroScreen revealed that tracts with higher burdens of asthma, cardiovascular disease, low birth weight, and child populations consistently faced higher exposures to NO_3^- and OC (0.28–0.59 μg/m³ above low-burden tracts). Decomposition analyses indicated that meteorology contributed variably but weakly (<10 % of interannual variance for most species, except DUST at ~30–40 %), whereas wildfire smoke increasingly offset regulatory gains, especially in northern California and the Sierra Nevada. These high-resolution component datasets provide a valuable foundation for characterizing spatiotemporal exposure patterns, evaluating emission-control effectiveness, and informing equitable air-quality management strategies across California.

Executive Summary

Fine particulate matter ($PM_{2.5}$) remains a major public health concern, yet most monitoring and regulation have focused on its total mass rather than chemical composition. Different $PM_{2.5}$ species arise from distinct sources, display unique atmospheric behaviors, and exert varying health effects. Despite these differences, $PM_{2.5}$ continues to be regulated on the basis of total mass, and most exposure assessments and health studies have examined associations with total $PM_{2.5}$ rather than individual components. Although interest in and concern about $PM_{2.5}$ components are increasing, variability in component-specific exposures and their health effects remains less well understood. To address this gap, this project developed high-resolution, long-term estimates of $PM_{2.5}$ chemical components and examined their implications for exposure, policy effectiveness, and environmental justice in the western United States, with particular emphasis on California.

Methodology

This project developed the first high-resolution, long-term dataset of $PM_{2.5}$ chemical components—sulfate (SO_4^{2-}), nitrate (NO_3^-), elemental carbon (EC), organic carbon (OC), and mineral dust (DUST)—across the western United States (2002–2019) and California (2000–2020). The 2002-2019 dataset across the western United States was generated by integrating multiple data sources within a two-stage advanced machine-learning framework. In the first stage, gap-free daily $PM_{2.5}$ total mass concentrations at 1-km resolution were derived from satellite aerosol optical depth (AOD) and ancillary predictors. In the second stage, these gap-free $PM_{2.5}$ estimates and speciated $PM_{2.5}$ fields from the Community Multiscale Air Quality (CMAQ) model were used as primary predictors, supplemented by meteorological variables, land cover, population density, and satellite-based wildfire detections. Using these inputs, daily 1-km concentrations of the five components were estimated with spatiotemporal deep-forest models. The models achieved strong predictive performance (cross-validation R^2 = 0.66–0.89) and reproduced realistic spatial and temporal patterns consistent with ground observations. To extend the framework to California for 2000–2020, MERRA-2 and MERRA-2 GMI reanalysis products were incorporated to replace unavailable CMAQ inputs before 2002 and after 2019, producing a continuous 21-year record of daily 1-km concentrations for all five $PM_{2.5}$ components. Using the long-term high-resolution dataset across California, we conducted long-term trend, population-weighted exposure, and environmental disparity analyses, along with a two-stage decomposition to identify the drivers of observed trends. The decomposition first applied meteorological normalization to isolate emission-driven changes from meteorology-induced variability, followed by partitioning concentrations into fire and non-fire components based on satellite-derived smoke plume observations serving as wildfire candidate gates. This framework enabled quantitative attribution of long-term $PM_{2.5}$ and component changes to meteorological, wildfire, and anthropogenic emission influences.

Key Findings

Western U.S. (2002–2019): Spatiotemporal analysis of our model predictions showed that urban populations were exposed to 1.5–2 times higher concentrations of SO_4^{2-}, NO_3^-, EC, and OC compared with rural populations, while exposures to DUST were comparable. All five species exhibited declining trends over the study period, although patterns varied by season, region, and species. Wildfire events contributed sharp episodic increases, particularly in OC, NO_3^-, and EC, superimposed on the broader downward trends. Day-to-day analysis of extreme events, such as the 2018 Camp Fire, revealed that OC and NO_3^- exposures increased approximately sixfold during the fire period, demonstrating the models' capability to capture both chronic and acute exposure patterns.

California hotspots (2000–2020): From 2000 to 2020, concentrations of $PM_{2.5}$ and its major components showed strong and persistent spatial concentration in the Los Angeles megaregion and the San Joaquin Valley (SJV), with secondary elevations in the San Francisco Bay Area and along inland transport corridors. OC and NO_3^- were the dominant contributors to these concentration hotspots, jointly accounting for nearly half of total $PM_{2.5}$, while SO_4^{2-}, and DUST played smaller yet region-specific roles. Population-weighted $PM_{2.5}$ exposures decreased by about 20 % over two decades—from approximately 14.8 μg/m³ in 2000 to 11.9 μg/m³ in 2020. Statewide component averages declined most strongly for SO_4^{2-} and NO_3^- (≈50 %), reflecting effective fuel-sulfur and NO_x controls, whereas OC and EC decreased more modestly (≈6–12 %) and DUST remained largely unchanged. However, the benefits were uneven: communities with high burdens of asthma, cardiovascular disease, child populations, or low-birth-weight prevalence consistently experienced higher exposures, especially to NO_3^- and OC, while elderly-dense tracts tended to face lower concentrations. These patterns reveal both the progress and the persistence of regional and demographic disparities in California's $PM_{2.5}$ burden.

Decomposition and Policy Implications: Decomposition analyses separating meteorological, wildfire, and anthropogenic influences show that most long-term $PM_{2.5}$ and component trends reflect emission reductions rather than weather variability, with meteorology explaining less than 10% of the variance for most species but up to 30-40% for DUST. Wildfire smoke, dominated by OC and EC, emerges as an increasingly frequent episodic driver, particularly in northern California and Sierra Nevada mountain regions, while the chronic baseline in the SJV and Southern California remains governed by anthropogenic sources. After removing meteorological and fire effects, $SO_4^{2-}$and EC exhibit strong historical declines followed by recent rebounds, NO_3^- persists as a concentrated hotspot in the SJV and coastal SoCAB, OC remains the dominant contributor in both urban and valley regions, and DUST shows high sensitivity to climatic drying and resuspension but minimal long-term change. Together, these findings confirm that California's emission-control programs have achieved major and lasting reductions but also highlight emerging challenges—plateauing progress in EC and SO_4^{2-}, persistent NO_3^- and OC burdens, climate-linked DUST variability, and escalating wildfire smoke—that must be addressed to sustain and equalize future air-quality gains.

Implications

The decomposition of California's $PM_{2.5}$ record underscores both the effectiveness of past policies and the challenges that remain. Sharp reductions in SO_4^{2-} and EC, particularly in Southern California, demonstrate the success of fuel sulfur regulations, diesel emission standards, and industrial controls, showing that technology-based policies can deliver rapid and large-scale benefits. However, persistently high NO_3^- and OC levels in the San Joaquin Valley reveal the limits of current strategies in ammonia-rich and secondary-organic-aerosol-prone environments, pointing to the need for targeted agricultural measures, better control of residential wood combustion, and reductions in aerosol precursors. DUST also remains problematic, with plateauing or rebounding trends in Fresno, Sacramento, and Los Angeles indicating that existing controls have been insufficient to curb agricultural and resuspended urban dust. Finally, the growing influence of wildfire smoke—dominated by OC and EC and concentrated in the northern Sacramento Valley and Sierra Nevada—presents a distinct and worsening challenge, emphasizing the need to link air quality management with forest and land management strategies. CalEnviroScreen-based analyses show census tracts with higher burdens of asthma, cardiovascular disease, and low birth weight consistently experience elevated exposures to NO_3^- and OC—the two dominant components in California's most polluted regions—underscoring that the remaining $PM_{2.5}$ burden disproportionately affects vulnerable populations. Addressing these disparities will require integrating air quality management with environmental justice initiatives, prioritizing emission reductions and clean-technology investments in disadvantaged communities. Together, these findings suggest that while California's emission control programs have been broadly effective, sustaining progress will require a dual focus on reducing persistent anthropogenic sources, and mitigating wildfire smoke impacts.

1. Introduction

1.1 Motivation

Fine particulate matter with an aerodynamic diameter smaller than 2.5 μm ($PM_{2.5}$) is a complex mixture of chemical constituents, including sulfate (SO_4^{2-}), nitrate (NO_3^-), elemental carbon (EC), organic carbon (OC), and mineral dust (DUST). A growing body of toxicological and epidemiological evidence suggests that these constituents differ in their sources, atmospheric behavior, and potential to cause harm to human health. For example, EC—primarily emitted from combustion sources—is strongly associated with cardiovascular effects, while secondary inorganic aerosols such as NO_3^- and SO_4^{2-} have been linked to respiratory and systemic inflammation. OC, which can originate from both primary and secondary formation processes, contains various compounds of differing toxicity, some of which are highly oxidative and can trigger cellular stress. Consequently, certain $PM_{2.5}$ species may be more harmful than others on an equal-mass basis, and their health impacts cannot be fully inferred from total $PM_{2.5}$ concentrations alone.

In addition to differing toxicities, these five major $PM_{2.5}$ constituents often display greater spatial and temporal variability than total $PM_{2.5}$ mass. This variability arises from differences in emission source distributions, atmospheric formation pathways, meteorological influences, and chemical reactivity. For instance, EC concentrations can vary sharply over short distances in urban areas due to localized traffic emissions, while secondary species such as SO_4^{2-} may have more regional patterns influenced by precursor transport and photochemical processes. Such heterogeneity means that monitoring only total $PM_{2.5}$ mass may fail to capture important local-scale differences in exposure and risk.

California presents a particularly critical case for advancing $PM_{2.5}$ speciation research. The state continues to experience some of the highest $PM_{2.5}$ levels in the United States, with persistent non-attainment in regions such as the South Coast and San Joaquin Valley air basins. The spatial patterns of SO_4^{2-}, NO_3^-, EC, OC, and DUST in California are shaped by a complex interplay of urban traffic emissions, industrial activities, agricultural operations, wildfires, and long-range transport. However, the state's ground-based $PM_{2.5}$ speciation monitoring network is sparse—Los Angeles County, for example, has only two routine monitoring sites for chemical speciation—limiting the ability to evaluate fine-scale exposure differences across communities. Given the environmental justice concerns in California, where disadvantaged populations often live closer to major emission sources, a high-resolution understanding of $PM_{2.5}$ constituent concentrations is essential for effective and equitable air quality management.

1.2 Research Objectives

The overarching goal of this project is to generate high spatiotemporal resolution estimates of five key $PM_{2.5}$ constituents—SO_4^{2-}, NO_3^-, EC, OC, and DUST—across California by integrating multiple data sources, including satellite remote sensing, chemical transport model outputs, meteorological reanalyses, land-use information, and ground-based measurements. This integrated modeling framework enables the quantification of constituent-specific patterns and trends over two decades, providing critical insights into spatial disparities and long-term progress in air quality improvement.

The specific objectives are:

(1). Quantify California's local and regional ambient $PM_{2.5}$ components[1]—SO_4^{2-}, NO_3^-, EC, OC, and DUST—for the period 2000–2020 by integrating multiple data sources, including ground-based measurements, satellite observations, simulations, and reanalysis data.

(2). Evaluate long-term trends in both total $PM_{2.5}$ mass and its major chemical components, with particular focus on areas underrepresented in the existing monitoring network.

(3). Assess spatial disparities in $PM_{2.5}$ constituent concentrations across demographic and geographic subgroups in California.

(4). Recommend pathways to refine $PM_{2.5}$ mitigation strategies that account for constituent-specific patterns and incorporate environmental justice considerations.

1.3 Hypotheses

This study is guided by the following hypotheses:

1. Greater spatial heterogeneity of constituents – The five major $PM_{2.5}$ constituents—SO_4^{2-}, NO_3^-, EC, OC, and DUST—exhibit greater spatial variability than total $PM_{2.5}$ mass, particularly in regions influenced by localized sources such as traffic corridors, industrial zones, and agricultural areas.

2. Enhanced detection through high-resolution multi-source integration – Integrating multiple data sources, including satellite products, chemical transport model outputs, meteorological reanalyses, land-use variables, and ground-based measurements, will provide high spatiotemporal resolution estimates capable of revealing spatial and temporal patterns of SO_4^{2-}, NO_3^-, EC, OC, and DUST that are not captured by the existing ground monitoring network.

3. Identification of disparities in constituent exposures – The improved high-resolution constituent data will uncover disparities in exposure to SO_4^{2-}, NO_3^-, EC, OC, and DUST among different demographic and geographic subgroups in California, providing evidence to inform more equitable and effective $PM_{2.5}$ mitigation strategies.

---

[1] Throughout this report, the terms "component" and "species" are used interchangeably, both referring to the entities discussed under "speciation" in the project proposal.

2. Literature Review and Preliminary Results

2.1 Introduction

Understanding the chemical composition of fine particulate matter ($PM_{2.5}$) is a critical step toward improving air quality management and public health protection. $PM_{2.5}$ comprises a mixture of components with distinct emission sources, atmospheric lifetimes, and physical and chemical properties (Hand et al. 2014; Seinfeld and Pandis 2016; WHO 2021). Among these, SO_4^{2-}, NO_3^-, EC, OC, and DUST are the most abundant constituents in many regions, contributing significantly to the mass and toxicity of $PM_{2.5}$. Their individual spatial and temporal variations carry important implications for exposure assessment, source attribution, and targeted emission controls. However, their quantification at fine spatial and temporal resolution over long periods is challenging, as direct chemical speciation monitoring is sparse and infrequent, and no single data source provides the necessary combination of chemical accuracy, spatial continuity, and temporal completeness. In the following sections, we review existing data sources for $PM_{2.5}$ components (Section 2.2) and recent developments in hybrid modeling approaches (Section 2.3), followed by a discussion of their implications for this study.

2.2 Data sources available for $PM_{2.5}$ components

2.2.1 Monitoring networks

Ground-based chemical speciation measurements are the foundation for $PM_{2.5}$ component research, providing the most reliable data for model training and evaluation. In the United States, three nationwide monitoring networks form the backbone of these observations. The Chemical Speciation Network (CSN) focuses on urban and suburban sites, capturing population exposure in densely inhabited areas; the Interagency Monitoring of Protected Visual Environments (IMPROVE) network targets rural and remote locations, often within national parks and wilderness areas; and the Clean Air Status and Trends Network (CASTNET) primarily measures rural background air quality on a weekly basis (Solomon et al. 2014). These networks have generated long-term datasets—CSN since the late 1990s and IMPROVE since the late 1980s—that have been used extensively for trend analysis, regulatory assessments, and epidemiological studies (Hand et al. 2014; Malm et al. 2011). However, site density is limited to a few hundred stations across the continental U.S., with large gaps in mountainous, rural, and desert regions. Sampling intervals, typically every 1–6 days, further restrict their utility for daily mapping, particularly for species with high day-to-day variability.

2.2.2 Satellite retrievals

Satellite remote sensing has transformed air quality research by offering broad spatial coverage and, in some cases, fine spatial resolution. The Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm applied to MODIS observations produces aerosol optical depth (AOD) retrievals at 1 km resolution, enabling near-global mapping of atmospheric aerosol loading

(Li et al. 2020; He et al. 2023a). MAIAC's high resolution makes it particularly valuable for capturing fine-scale gradients in urban and complex-terrain settings. However, AOD represents total columnar extinction rather than surface-level mass, and the relationship with ground-level $PM_{2.5}$ components is nonlinear and modulated by atmospheric mixing, humidity, and aerosol composition (Franklin et al. 2017; Meng et al. 2018a). The Multi-angle Imaging SpectroRadiometer (MISR) fractional AOD offers additional information on particle size and composition at 4.4 km resolution, improving physical relevance for component estimation (Geng et al. 2020), but its coarse revisit time (global coverage every nine days) and reduced spatial footprint limit its capacity for continuous daily mapping. Both products are further constrained by retrieval gaps due to clouds, snow cover, or bright surfaces.

2.2.3 Chemical transport models (CTMs)

CTMs simulate the full life cycle of atmospheric particles by integrating emissions, chemical transformations, transport, and deposition, driven by meteorological data. The Community Multiscale Air Quality (CMAQ) model, developed by the U.S. EPA, provides species-resolved $PM_{2.5}$ outputs at resolutions as fine as 12 km for regional applications (Appel et al. 2017). CMAQ has been widely used in research and regulatory contexts for its ability to generate physically consistent spatiotemporal patterns of $PM_{2.5}$ components, including SO_4^{2-}, NO_3^-, EC, OC, and dust. However, uncertainties in emission inventories, parameterizations of secondary aerosol formation, and coarse resolution for large domains (often ≥12 km) can limit agreement with local observations, especially in heterogeneous regions such as the western U.S. Other CTMs, such as GEOS-Chem and WRF-Chem, have similar limitations, with typical long-term simulations conducted at even coarser spatial scales (≥25 km).

2.2.4 Dispersion models

Dispersion modeling approaches, including Gaussian plume (e.g., AERMOD) and Lagrangian particle tracking models, are designed to predict pollutant concentrations from specific sources given meteorological inputs (Gibson et al. 2013). They are effective for estimating near-source impacts and for regulatory permitting applications but are not suited to simulating regional-scale secondary aerosol formation or the broad range of chemical processes affecting $PM_{2.5}$ components.

2.2.5 Reanalysis datasets

Atmospheric reanalysis products blend model simulations with multiple observational datasets, producing spatially and temporally complete fields of meteorological and chemical variables. MERRA-2 includes global, multi-decadal records of speciated aerosol concentrations (Randles et al. 2017). While reanalyses provide consistent, gap-free coverage and can extend component estimates into data-sparse regions, their coarse resolution (~0.5° × 0.625°) makes them insufficient for detailed exposure assessment, especially in urban and topographically complex environments.

2.3 Review of $PM_{2.5}$ component hybrid modeling approaches

Estimating $PM_{2.5}$ components at fine spatiotemporal resolution is inherently more complex than modeling total $PM_{2.5}$, given their greater variability, shorter lifetimes, and localized sources (Amini et al. 2022; Donkelaar et al. 2019; He et al. 2023b). Hybrid modeling integrates complementary data sources to overcome the limitations of individual datasets. By combining chemically accurate ground observations, spatially extensive satellite retrievals, and mechanistically rich CTM outputs—together with meteorological, land use, and population data—these approaches enable broader coverage and more accurate predictions (Donkelaar et al. 2019; Amini et al. 2022).

2.3.1 Statistical and machine-learning $PM_{2.5}$ component models and limitations

Hybrid models use diverse statistical and machine-learning frameworks to integrate multiple predictors. Land-use regression (Hoogh et al. 2013; Hsu et al. 2018) incorporates geographic covariates but is limited in temporal coverage. Geographically weighted regression (Donkelaar et al. 2024) improves spatial flexibility but is computationally demanding and sensitive to monitoring density. Nonlinear methods such as random forest (Geng et al. 2020) and backpropagation neural networks (Di et al. 2016) capture complex interactions among predictors but often rely on predictors at coarse resolution. For example, Meng et al. (2018b) estimated $PM_{2.5}$ components using coarse-resolution GEOS-Chem reanalysis component data and auxiliary predictors at ~0.25° × 0.3125°, demonstrating the feasibility of multi-source integration but also highlighting the need for high-resolution, physically relevant predictors. Ensemble approaches, such as those in Amini et al. (2022), stack multiple base learners to improve accuracy, achieving resolutions as fine as 50 m in urban areas at the annual scale. Despite these advances, most existing models achieve either high spatial resolution or high temporal resolution, but rarely both, due to resolution constraints in key predictors and computational costs.

2.3.2 Challenges in estimating components in California

California presents a particularly complex environment for $PM_{2.5}$ component modeling due to its combination of diverse emission sources, frequent wildfire activity, complex terrain, and meteorological variability. The state's urban areas, such as the Los Angeles–Long Beach–Anaheim metropolitan region, experience intense traffic emissions and industrial activities, contributing to elevated levels of NO_3^-, OC, and EC. The San Joaquin Valley, bounded by mountains, suffers from poor dispersion conditions that exacerbate NO_3^- and carbonaceous aerosol accumulation. Episodic wildfire events—especially in northern and central California—introduce large, variable emissions of OC and EC, often overwhelming typical seasonal patterns. Additionally, DUST emissions in the southeastern deserts contribute significantly to DUST concentrations, especially under dry and windy conditions. Sparse monitoring coverage in mountainous and rural regions, combined with retrieval gaps in satellite AOD data due to cloud, smoke, or snow cover, further complicates modeling. Previous studies have consistently reported lower predictive performance for $PM_{2.5}$ components in western United States, particularly for California, compared to other parts of the United States (Meng et al. 2018b; Geng et al. 2020), underscoring the need for models capable of integrating high-resolution predictors that can resolve both the spatial heterogeneity of emission sources and the temporal variability driven by episodic events.

2.4 Preliminary analysis

2.4.1 Correlation coefficient analysis based on monitoring observation data

Previous efforts to estimate $PM_{2.5}$ component concentrations over large geographic areas have often been constrained by limited spatiotemporal resolution and suboptimal modeling accuracy, even when incorporating multiple data sources such as satellite AOD and CTM outputs. To inform the design of our modeling framework, we conducted a preliminary analysis examining the statistical relationship between observed total $PM_{2.5}$ mass concentrations and observations of the five major components.

The degree of linear association between total $PM_{2.5}$ and each component was quantified using the Pearson correlation coefficient. As summarized in Table 2.1, the five target components are key chemical constituents of $PM_{2.5}$ mass and exhibit moderate to strong correlations with total $PM_{2.5}$ concentrations (r = 0.33–0.85). These correlations reflect the shared influence of emission sources and atmospheric processes on both total $PM_{2.5}$ and its components. Based on this evidence, total $PM_{2.5}$ was selected as a primary predictor for component modeling.

Table 2.1. Correlations between $PM_{2.5}$ observations and species observations across the western U.S., 2000-2020.

SO_4^{2-}	NO_3^-	EC	OC	DUST
0.50	0.67	0.74	0.85	0.33

In addition to its chemical relevance, total $PM_{2.5}$ mass can be estimated at high spatial and temporal resolution. Specifically, daily, 1 km gap-free $PM_{2.5}$ fields can be derived from satellite AOD and other auxiliary predictors, providing spatially continuous and temporally consistent coverage across the study domain. This dual advantage—strong chemical linkage to the components and high-resolution availability—positions total $PM_{2.5}$ as an effective anchor variable for integrating other predictors, including CTM-speciated outputs, meteorological parameters, and land-use characteristics.

2.4.2 Expanding modeling region

During the implementation stage, we found that California's $PM_{2.5}$ component monitoring network contains only a limited number of stations, particularly for speciated measurements (Figure 2.1). The sparse spatial coverage limits the representativeness of the training and validation data if modeling were restricted solely to California. To enhance model robustness and capture a broader range of emission source types, meteorological regimes, and terrain features, we expanded the modeling domain to encompass the entire western United States. This larger study area (Figure 2.1) includes additional monitoring sites from the CSN and IMPROVE networks, increasing both the diversity and volume of training data. Expanding the domain also helps the model learn from regions with similar pollution characteristics and transport patterns to California, ultimately improving prediction accuracy within the state.

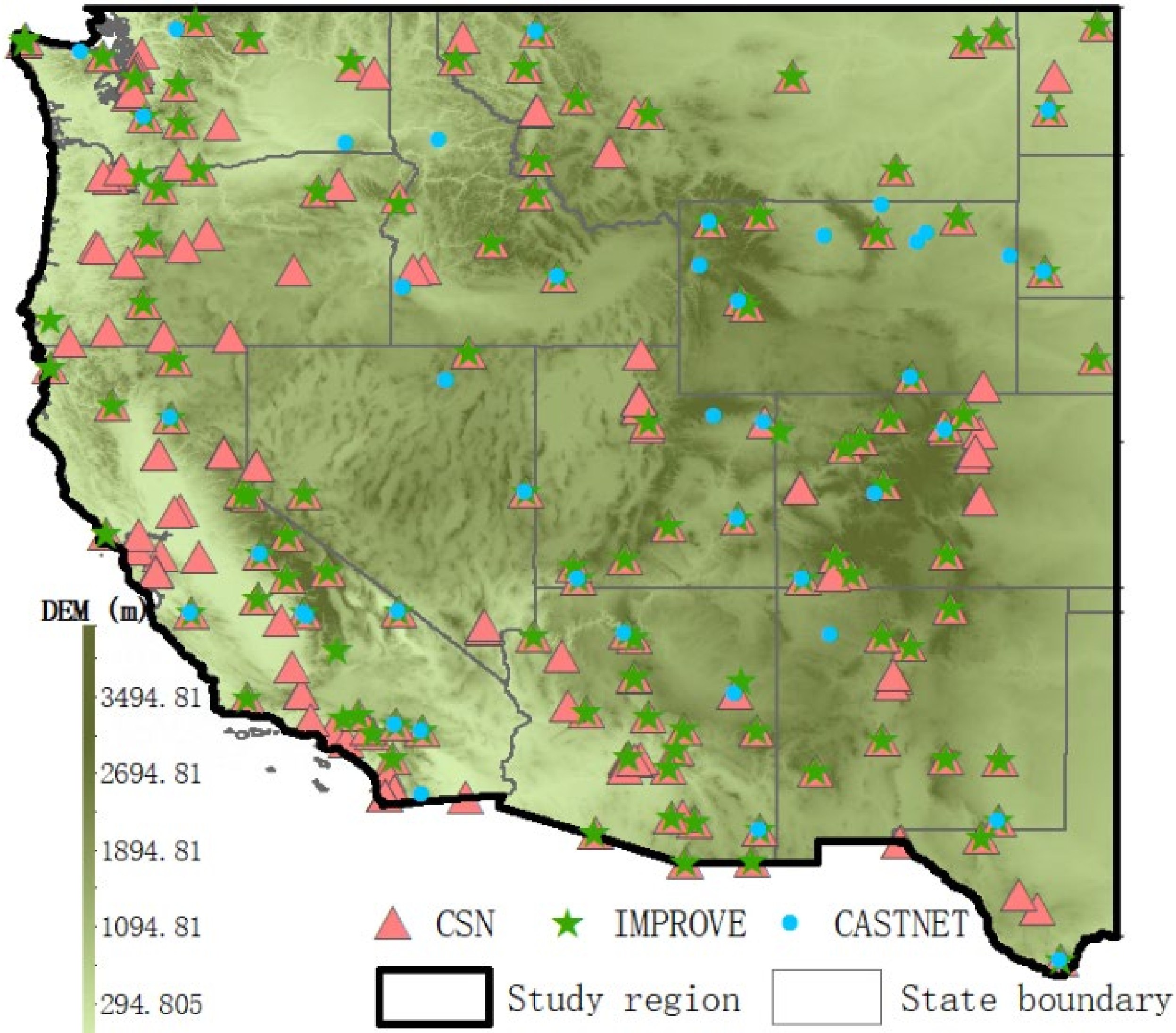

Figure 2.1. Study region and spatial distribution of $PM_{2.5}$ component monitoring network used in this study.

2.5 Summary

Existing studies on $PM_{2.5}$ component estimation have significantly advanced our understanding of spatial and temporal patterns, yet datasets offering both high spatial and high temporal resolution remain exceptionally scarce. Most prior efforts have achieved either fine spatial detail, often limited to annual or seasonal averages, or daily to sub-daily resolution at much coarser spatial scales. As a result, no long-term, spatially continuous record currently exists for $PM_{2.5}$ chemical components that simultaneously resolves fine-scale spatial variability and daily temporal dynamics across large and complex regions such as California. Two major limitations underlie this gap:

- First, many hybrid models still depend on coarse-resolution predictors, such as CTM or reanalysis outputs at ≥12 km, which limit the ability to resolve fine-scale gradients in complex environments. Integrating high-resolution, physically and chemically relevant predictors—particularly daily, 1 km total $PM_{2.5}$ estimates supplemented by CTM-speciated outputs and other covariates—can address this gap.
- Second, many modeling algorithms are not optimized to capture both gradual seasonal variations and rapid changes from episodic events like wildfires. Employing advanced spatiotemporal modeling approaches—such as deep learning, deep forest, or stacking ensemble techniques—can better capture these dynamics, especially in data-sparse and heterogeneous regions.

Through this exploratory analysis, we identified these two focal areas as potential improvement directions for developing more accurate, spatially resolved, and temporally continuous $PM_{2.5}$ component datasets for California and other regions with complex atmospheric environments. Therefore, we chose to develop a spatiotemporal deep-learning model where $PM_{2.5}$ total mass serves as a primary predictor to estimate daily, gap-free $PM_{2.5}$ SO_4^{2-}, NO_3^-, EC, OC, and DUST at a high spatial resolution of 1 km over the western United States. This approach addresses both identified gaps: (1) using high-resolution, gap-filled total $PM_{2.5}$ as a chemically relevant predictor (correlation coefficients $r = 0.33$-0.85 with components), and (2) employing a spatiotemporal deep-forest algorithm capable of capturing complex spatiotemporal patterns including episodic wildfire events.

3. Data Sources and Preprocessing

This project integrates a diverse suite of ground-based measurements, satellite remote sensing products, chemical transport model outputs, meteorological reanalysis datasets, and geographical variables to produce daily, gap-free estimates of five major $PM_{2.5}$ components—SO_4^{2-}, NO_3^-, EC, OC, and DUST—at a 1-km spatial resolution across the western United States for the period 2000–2020. In this section, we describe the data sources and preprocessing procedures used to harmonize datasets with varying spatial and temporal resolutions onto a uniform 0.01° (~1 km) grid, using satellite-derived AOD as the spatial reference. The modeling domain was expanded beyond the core study area for reasons detailed in Section 2.4.2. The resulting spatial extent, encompassing the western United States, is shown in Figure 3.1.

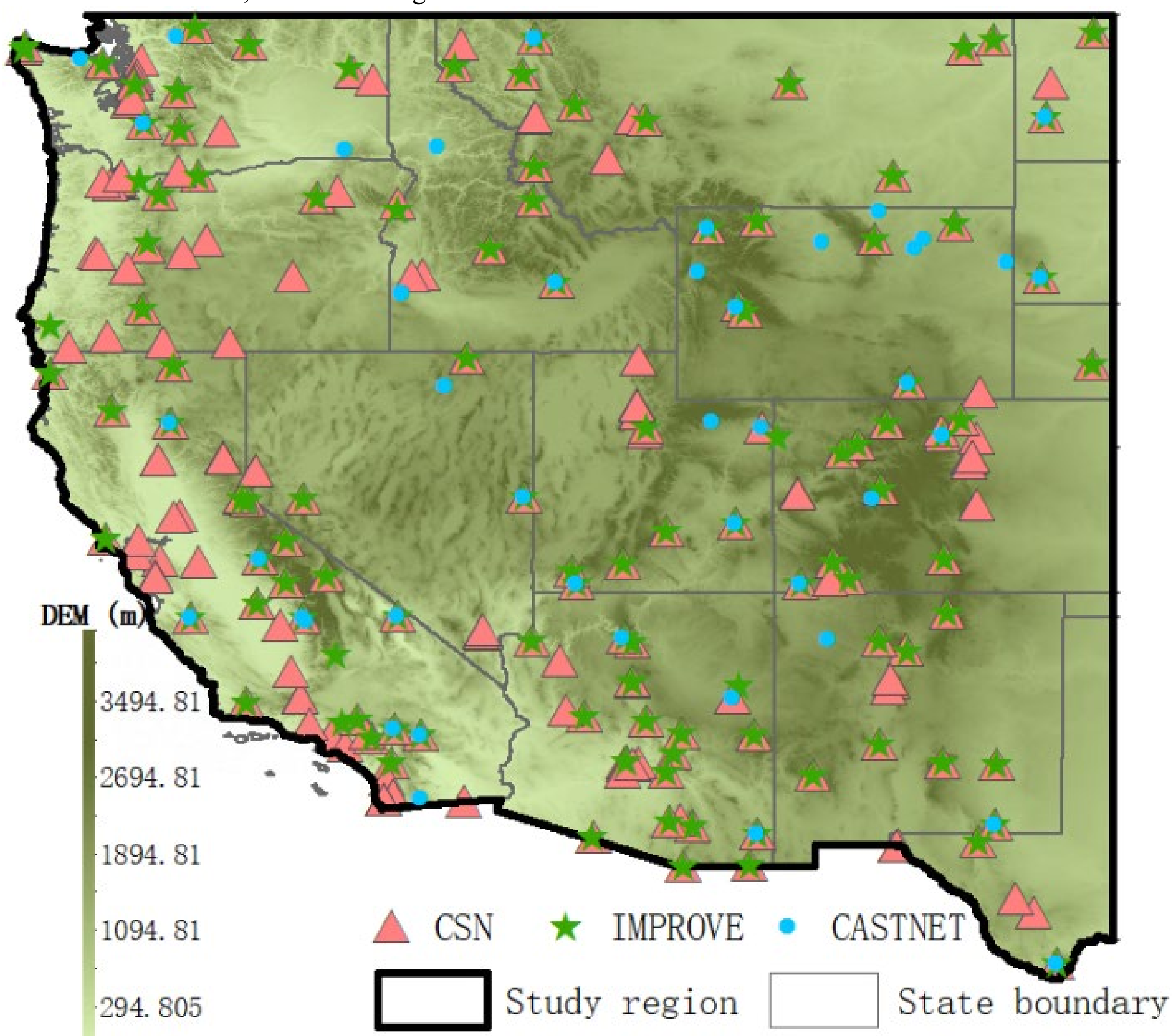

Figure 3.1. Modeling region and spatial distribution of $PM_{2.5}$ component monitoring network used in this study.

3.1 Ground-level $PM_{2.5}$ total mass and speciated measurements

We obtained daily $PM_{2.5}$ component concentrations from three major U.S. monitoring networks: (1). Chemical Speciation Network (CSN) – Operated by the U.S. Environmental Protection Agency (EPA) (https://www.epa.gov/amtic/chemical-speciation-network-csn). CSN sites are primarily located in urban and suburban areas. Sampling frequency: 1–6 days.

(2). Interagency Monitoring of Protected Visual Environments (IMPROVE) – Managed by a federal-state partnership (https://vista.cira.colostate.edu/Improve/). IMPROVE sites are generally located in remote or rural locations, including national parks and wilderness areas. Sampling frequency: 1–3 days.

(3). Clean Air Status and Trends Network (CASTNET) – Operated by EPA to monitor rural air quality (https://www.epa.gov/castnet). CASTNET measurements have weekly resolution and were used only for independent validation.

In the modeling, CSN and IMPROVE measurements served as the training and cross-validation datasets, while CASTNET data provided independent evaluation of SO_4^{2-} and NO_3^- estimates. Measurements of SO_4^{2-} and NO_3^- are comparable between CSN and IMPROVE (Hand et al., 2012; Solomon et al., 2014) and were directly combined. EC and OC were harmonized following Malm et al. (2011) and Meng et al. (2018a) to correct for analytical differences. Specifically, for CSN EC and OC measurements, data processing differed by method. When analyzed using the Thermal Optical Reflectance (TOR) technique, blank corrections were applied directly to the measurements. For samples analyzed with the Thermal Optical Transmittance (TOT) method, EC concentrations were scaled by a factor of 1.3 to align with IMPROVE EC values. OC concentrations, on the other hand, were adjusted using Equation below (Malm et al., 2011).

$$OC_{adj} = ((OC_{CSN} - 0.\times EC_{CSN}) - A/M$$

The coefficients A and M used in this equation are provided in Table 3.1. The equation was derived from paired CSN and IMPROVE measurements of PM2.5 OC and EC at collocated sites during 2005–2006. In this context, A accounts for the monthly positive artifact caused by filter adsorption of semivolatile organic compounds (SVOCs), while M represents the multiplicative negative artifact associated with volatilization losses of collected OC (Malm et al., 2011).

Table 3.1. A and M values for CSN OC conversion.

M (unitless)	1.2
A_{Jan} ($\mu g/m^3$)	1.1
A_{Feb} ($\mu g/m^3$)	1.3
A_{Mar} ($\mu g/m^3$)	1.2
A_{Apr} ($\mu g/m^3$)	1.4
A_{May} ($\mu g/m^3$)	1.6

A_{Jun} ($\mu g/m^3$)	1.7
A_{Jul} ($\mu g/m^3$)	1.8
A_{Aug} ($\mu g/m^3$)	1.9
A_{Sep} ($\mu g/m^3$)	1.5
A_{Oct} ($\mu g/m^3$)	1.2
A_{Nov} ($\mu g/m^3$)	1.0
A_{Dec} ($\mu g/m^3$)	1.1

DUST concentrations were calculated from elemental measurements using the IMPROVE formula:

$$DUST = 2.20 \times Al + 2.49 \times Si + 1.63 \times Ca + 2.42 \times Fe + 1.94 \times Ti$$

Daily $PM_{2.5}$ total mass concentrations from approximately 1,300 U.S. EPA monitoring sites (Figure 3.2) were used as the target variable in the modeling of gap-free $PM_{2.5}$ total mass, which served as a primary predictor for $PM_{2.5}$ component estimation. To reduce edge effects associated with the spatial boundaries of $PM_{2.5}$ estimates, the $PM_{2.5}$ modeling domain was expanded beyond the component modeling region to include entire states along the western edge of the study area (Figure 3.2).

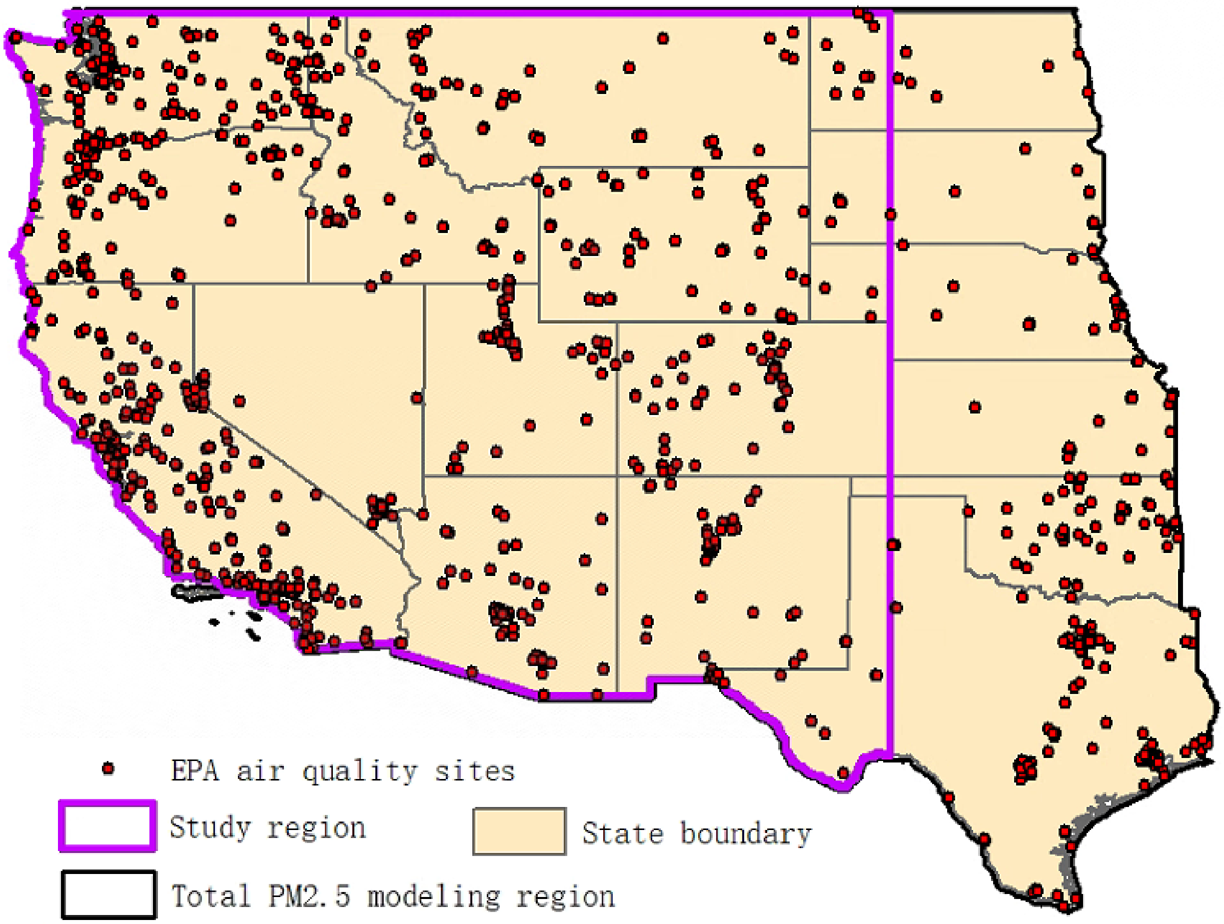

Figure 3.2. Spatial distribution of total $PM_{2.5}$ observation sites used in this study with extended modeling region for total $PM_{2.5}$ modeling.

3.2 Satellite aerosol optical depth (AOD) and gap filling

We used the MODIS Collection 6.1 MAIAC AOD product (MCD19A2) (https://lpdaac.usgs.gov/products/mcd19a2v061/) retrieved from the MODIS sensors aboard NASA's Terra and Aqua satellites (1 km daily resolution over land). To address missing retrievals from clouds, snow, or bright surfaces (availability <50% in many areas; Figure S3), we reconstructed gap-free daily AOD using a random forest imputation framework described in Section 5.

3.3 CMAQ speciated $PM_{2.5}$ simulations

We obtained weekly outputs from the Community Multiscale Air Quality (CMAQ) model (https://www.epa.gov/cmaq) at 12-km resolution for SO_4^{2-}, NO_3^-, NH_4^+, EC, and OC. CMAQ explicitly simulates emissions, chemical transformation, and transport (Appel et al., 2017). These outputs were interpolated to the 1-km modeling grid via inverse distance weighting.

3.4 Meteorological datasets

Three sources of meteorological data were used:

(1) Daily maximum (Tmax) and minimum (Tmin) temperature, shortwave radiation (SRAD), snow water equivalent (SWE), daylight duration (DayL), and water vapor pressure (WVP) were obtained from DayMet (1km × 1km) (Thornton et al. 2022);

(2) daily maximum (Rmax) and minimum (Rmin), relative humidity, mean vapor pressure deficit (VPD), mean wind speed at 10 m (WS10M), and wind direction (WD) were obtained from gridMET (~4 km × 4 km) (Abatzoglou 2013);

(3) planetary boundary layer height (PBLH), total cloud area fraction (CLDTOT), 2-m eastward (U2M) and northward (V2M) wind, surface air temperature (TLML), surface pressure (PS) and total precipitation (PRECTOT) were obtained from MERRA-2 GMI "Daily Average Diagnostics" (~55.5 km × 50 km) reanalysis data (https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI).

Datasets with spatial resolutions coarser than 1 km were downscaled to a 1-km grid using bilinear interpolation for MERRA2-GMI and inverse distance weighting for gridMET.

3.5 Auxiliary data

We incorporated multiple geospatial datasets:

- We obtained 30-m resolution impervious surface and land cover data from the National Land Cover Database (NLCD) for years 2001, 2004, 2006, 2008, 2011, 2013, 2016, and 2019 (https://www.mrlc.gov). We then aggregated the 30-m data to our 1-km grid by calculating the fraction of road density (RD), developed areas of varying intensity (NLCD_D), forest (NLCD_F), grassland (NLCD_G), and wetlands (NLCD_W) within each 1-km grid cell.
- Annual 1-km population (POP) data were downloaded from the LandScan Global database developed by Oak Ridge National Laboratory (https://landscan.ornl.gov/).
- Elevation (ELE) data were sourced from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Map (GDEM) version 2 (https://asterweb.jpl.nasa.gov/gdem.asp). From this database, to represent geographical topography, we calculated slope, aspect, hillshade (HS), roughness, and topographic position index (TPI) using ArcGIS 10.8 Spatial Analyst Tools.
- 1-km resolution MODIS 16-day normalized difference vegetation index (NDVI) data (MOD13A2.061) were obtained to reflect vegetation surface changes.

Recognizing the frequent occurrence of wildfires in the western United States, we also obtained fire point detection data from the Hazard Mapping System (HMS, https://www.ospo.noaa.gov/products/land/hms.html) at daily time resolution for years 2004 to 2019. To integrate this dataset into our model, we applied Gaussian kernel density estimation to detected fires, generating daily spatially continuous fire intensity maps by estimating the density of fire points within a 1-km radius. The bandwidth was set automatically based on standard deviation and sample size of the data.

3.6 Summary

This section focused on data sources used for project. All datasets, their sources, resolutions, and preprocessing steps are summarized in Table 3.2.

Table 3.2. Data sources used in this project and spatiotemporal information.

Variable source	Variable	Abbreviation	Spatial scale	Temporal scale	Time span	Preprocessing method
MODIS C61 MAIAC aerosol product (MCD19A2)	Aerosol optical depth	AOD*	0.01°× 0.01°	daily	2002 - 2019	----
CMAQ	EC NH4 NO3 OC SO4	EC_CMAQ NH4_CMAQ NO3_CMAQ OC_CMAQ SO4_CMAQ	12×12 km	weekly	2002 - 2019	IDW
Daymet (https://daymet.ornl.gov/overview)	Day length Precipitation Shortwave Snow water equivalent Maximum air temperature Minimum air temperature Water vapor pressure Daily total radiation	DayL SRAD SWE Tmax Tmin WVP Trad	1km	daily	2002 - 2019	Eq. S1
gridMet (https://www.climatologylab.org/gridmet.html)	Precipitation amount Maximum relative humidity Minimum relative humidity Wind direction Mean vapor pressure deficit Wind speed	 Rmax Rmin* WD VPD WS10M*	0.0416°×0.0416°	Daily	2000 - 2020	IDW

MERRA-2 GMI "Daily Average Diagnostics" (https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI)	Total cloud area fraction	CLDTOT*	0.5°×0.625°	Daily	2000 - 2020	BR
	Planetary boundary layer height	PBLH*				
	Total precipitation	PRECTOT*				
	Surface air temperature	TLML*				
	Surface pressure	PS*				
	2-Meter eastward wind	U2M				
	2-Meter northward wind	V2M				
National Land Cover Database (NLCD) (https://www.mrlc.gov)	Developed, Open Space and Low/Meduim/High Intensity land cover	NLCD_D*				Extracting fraction of each class based on area
	Deciduous forest, Evergreen Forest, Mixed Forest, and Shrub/Scrub	NLCD_F*				
	Herbaceous and Hay/Pasture	NLCD_G*				
	Woody wetlands and Emergent herbaceous wetlands	NLCD_W*				
	Primary, Secondary, Tertiary, and Thinned road land cover	RD				
	Urban impervious surfaces	UMS				
ASTER DEM (https://doi.org/10.5067/ASTER/ASTGTM.003)	Elevation	ELE*	30m	----	----	Resampling Spatial Analyst in ArcGIS 10.8 based on elevation
	Slope					
	Aspect					
	Hillshade	HS*				
	Roughness					
	Topographic position index	TPI*				
MODIS Terra 16-day vegetation	Normalized difference vegetation index	NDVI*	1km	16-day	2002 - 2019	

product (MOD13A2.061)						
LandScan Global (https://landscan.ornl.gov/)	Population density	POP*	1km	1 years	2002 - 2019	
Hazard Mapping System Fire Detection (https://www.ospo.noaa.gov/products/land/hms.html)	Active fire detection	fireKD*	Points	Daily	2003 - 2019	Kernel density estimation

Trad =(srad (W/m2) × dayl (s/day)) / 1,000,000; IDW: inverse distance weighted interpolation: BR: bilinear resampling. * indicates this variable is used for modeling total $PM_{2.5}$.

4. Effect of Incorporating MISR Fractional AOD on $PM_{2.5}$ Component Estimation

4.1 Introduction

Previous research has demonstrated the potential of MISR (Multi-angle Imaging SpectroRadiometer) fractional aerosol optical depth (AOD) to improve $PM_{2.5}$ chemical component estimation by providing size-resolved aerosol properties linked to different chemical species (Meng et al., 2018a; Geng et al., 2020). MISR's unique multi-angle viewing geometry allows retrieval of fractional AOD for multiple aerosol size bins, thereby offering additional microphysical information beyond bulk total AOD from other satellite instruments. Such information is particularly relevant for species discrimination, as it can help distinguish between sulfate-rich fine aerosols, coarse-mode dust, and carbonaceous particles.

This section evaluates the influence of MISR fractional AOD on $PM_{2.5}$ component modeling for the western United States during 2004–2019. We assessed MISR's contribution both for (1) models developed only at locations and times where MISR fractional AOD is available (Section 4.3), and (2) gap-free high-resolution species modeling in which MISR data are combined with other predictors to fill missing coverage (Section 4.4).

4.2 MISR Fractional AOD Data and Calculation Method

The MISR instrument, aboard NASA's Terra satellite, observes the atmosphere at nine along-track viewing angles in four spectral bands (blue, green, red, near-infrared). This multi-angle capability enables retrieval of aerosol microphysical properties, including fractional AOD for predefined aerosol components characterized by size, shape, and single-scattering albedo.

In this study, we used the MISR Version 23 (V23) Aerosol Physical and Optical Properties product and focused on eight fractional AODs: AOD1, AOD2, AOD3, AOD6, AOD8, AOD14, AOD19, and AOD21. These correspond to aerosol components most relevant to dominant particle types in the western United States, such as sulfate-like fine aerosols, carbonaceous aerosols, and coarse-mode dust.

Fractional AOD values were calculated following Liu et al. (2007a, 2009):

Fractional AOD_i ($i = 1–8$)
$= (\Sigma_{j=1}^{74}\ \alpha \times AOD\_mixture_j \times Fraction\_component\ i\ in\ mixture_j)$ / (No. of successful mixtures)

where $AOD\_mixture_j$ is the total AOD of mixture j at 558 nm, $Fraction\_component\ i\ in\ mixture_j$ is the fractional contribution of component i to mixture j, and $\alpha = 1$ if mixture j is successfully retrieved, otherwise $\alpha = 0$.

The calculated fractional AODs were reprojected and collocated to the 1-km modeling grid, and subsequently used in a series of experiments to evaluate their influence on $PM_{2.5}$ component modeling. In these experiments, the fractional AODs were incorporated directly as predictors alongside meteorological, land cover, and other ancillary variables.

4.3 $PM_{2.5}$ Species Estimation Using MISR Fractional AODs

For the MISR-available subset, two sets of species-specific models were developed: one including MISR fractional AOD variables and one excluding them, to evaluate the incremental contribution of MISR's aerosol-type information. Both models used the same set of predictors to ensure a controlled comparison. Both models used the same non-MISR predictors, which comprised:

- $PM_{2.5}$ total mass concentration from the main component modeling framework (see Section 6).
- Meteorological variables: daily maximum temperature, snow water equivalent, vapor pressure, minimum relative humidity, planetary boundary layer height, total cloud area fraction, total precipitation, and horizontal wind components (U2M and V2M), as well as other atmospheric indicators derived from DayMet, gridMET, and MERRA-2.
- Land-use and land-cover metrics: including fractions of selected National Land Cover Database (NLCD) classes such as open/developed areas, land cover changes, and vegetation indices (NDVI), and elevation.
- Ancillary variables: day-of-year, and year.

The details including data sources and preprocessing approaches are detailed in Section 3. Performance evaluation using 10-fold cross-validation (CV) is summarized in Table 4.1. Inclusion of MISR fractional AODs produced small but consistent gains for certain species. In sample-based CV, R^2 increased from 0.78 to 0.79 for SO_4^{2-}, from 0.84 to 0.86 for NO_3^-, and from 0.67 to 0.68 for EC, while OC and DUST remained unchanged at 0.63 and 0.64, respectively. In site-based CV, which assesses spatial predictive skill, R^2 increased by 0.02 for SO_4^{2-} (0.68 to 0.70), NO_3^- (0.76 to 0.78), EC (0.61 to 0.63), and DUST (0.46 to 0.48), with OC again showing no change (0.58). These results indicate that MISR fractional AODs enhance spatial discrimination of certain $PM_{2.5}$ species, particularly those with distinct optical properties in MISR's aerosol component set, though the magnitude of improvement remains modest given the strong baseline predictor set that already incorporates total $PM_{2.5}$ concentration and multiple physical drivers.

Table 4.1. The CV results of models with and without MISR fractional AODs as predictor across western U.S., 2000-2020.

Model	Evaluation method	SO_4^{2-}	NO_3^-	EC	OC	Dust
Include MISR	Sample-based R^2	0.79	0.86	0.68	0.63	0.64
	Site-based R^2	0.70	0.78	0.63	0.58	0.48
Exclude MISR	Sample-based R^2	0.78	0.84	0.67	0.63	0.64

	Site-based R^2	0.68	0.76	0.61	0.58	0.46

4.4 Impact on gap-free, High-Resolution Modeling

To assess MISR's role in a gap-free, high-resolution daily species estimation framework, we designed two groups of experiments:

(1). Combined-model approach: For MISR-available areas/dates, models used MISR fractional AODs together with the same other predictor set listed in Section 3.3; for MISR-unavailable areas/dates, models used CMAQ-simulated species concentrations together with the same predictor set.

(2). CMAQ-only approach: For the entire study domain and all dates, models used CMAQ-simulated species concentrations together with the same predictor set.

Results for the 2004–2019 period (Table 4.2) show that the combined-model approach did not outperform the CMAQ-only approach. For example, for SO_4^{2-}, the combined model achieved an R^2 of 0.82 with an RMSE of 0.29 μg/m³, compared to 0.83 and 0.28 μg/m³ for the CMAQ-only model. For NO_3^-, both approaches achieved $R^2 = 0.90$ with RMSE differing by only 0.01 μg/m³ (0.57 vs. 0.56). Similar equivalence was observed for EC, OC, and dust.

Table 4.2. Sample-based 10-fold CV results of models with and without MISR fractional AODs as predictors

Model	Metrics	SO_4^{2-}	NO_3^-	EC	OC	Dust
① Areas with MISR (MISR used as predictors)	R^2	0.80	0.86	0.66	0.65	0.67
	RMSE	0.36	0.68	0.34	1.49	0.80
	Sample No.	14716	14659	12604	12601	14599
② Areas without MISR (CMAQ species used as predictors)	R^2	0.82	0.90	0.75	0.65	0.77
	RMSE	0.28	0.56	0.24	1.48	0.53
	Sample No.	236806	235851	211737	211625	235331
Combined (①+②)	R^2	0.82	0.90	0.75	0.65	0.77
	RMSE	0.29	0.57	0.25	1.48	0.55
	Sample No.	251522	250510	224341	224226	249930
Entire areas (CMAQ species used as predictors)	R^2	0.83	0.90	0.75	0.65	0.78
	RMSE	0.28	0.56	0.25	1.48	0.54
	Sample No.	251522	250510	224341	224226	249930

The absence of performance gains is primarily due to MISR's sparse observational coverage (<10% of daily samples across the study domain), which substantially reduces the effective sample size for the MISR-based portion of the combined model. Consequently, predictions in most of the domain are driven by the CMAQ-based model, and the limited MISR contribution does not materially change gap-free model accuracy.

4.5 Summary

In summary, MISR fractional AODs—derived from eight microphysically distinct aerosol components in the MISR V23 product—enhanced spatial predictive performance for SO_4^{2-}, NO_3^-, EC, and DUST in MISR-available samples when used together with total $PM_{2.5}$ concentration and other meteorological and land surface predictors, improving sample-based CV R^2 by 0.01-0.02 for most species. However, their sparse spatiotemporal coverage (<10% of daily samples across the study domain) prevented these gains from improving gap-free, daily estimation. Consequently, MISR fractional AODs were excluded from the final gap-free, high-resolution species models in this project phase due to their limited contribution under current observational constraints. For future applications where continuous daily coverage is not required—such as episodic exposure assessments or validation studies—MISR fractional AODs may still provide valuable microphysical information for component discrimination.

5. MAIAC AOD Imputation

5.1 Introduction

Producing accurate, high-resolution estimates of $PM_{2.5}$ components requires complete daily $PM_{2.5}$ total mass data. This, in turn, depends on full-coverage aerosol optical depth (AOD) information because AOD is physically related to $PM_{2.5}$: both measure aerosol loading in the atmosphere, and AOD captures the total column extinction of sunlight due to aerosols, which is strongly linked to surface-level $PM_{2.5}$ through shared emission sources, transport, and removal processes. AOD is also available at a daily time step and fine spatial resolution (1 km), making it an essential input for our $PM_{2.5}$ total mass modeling. However, these satellite aerosol retrievals contain large spatial and temporal gaps due to persistent cloud cover, snow, bright surfaces, and algorithm constraints. In the western U.S., more than half of daily grid cells lack valid retrievals (Fig 5.1). To ensure that $PM_{2.5}$ total mass modeling has the required gap-free AOD inputs, this section focused on the development of an AOD imputation process for generating a complete, gap-free, high-resolution AOD dataset for 2000–2020.

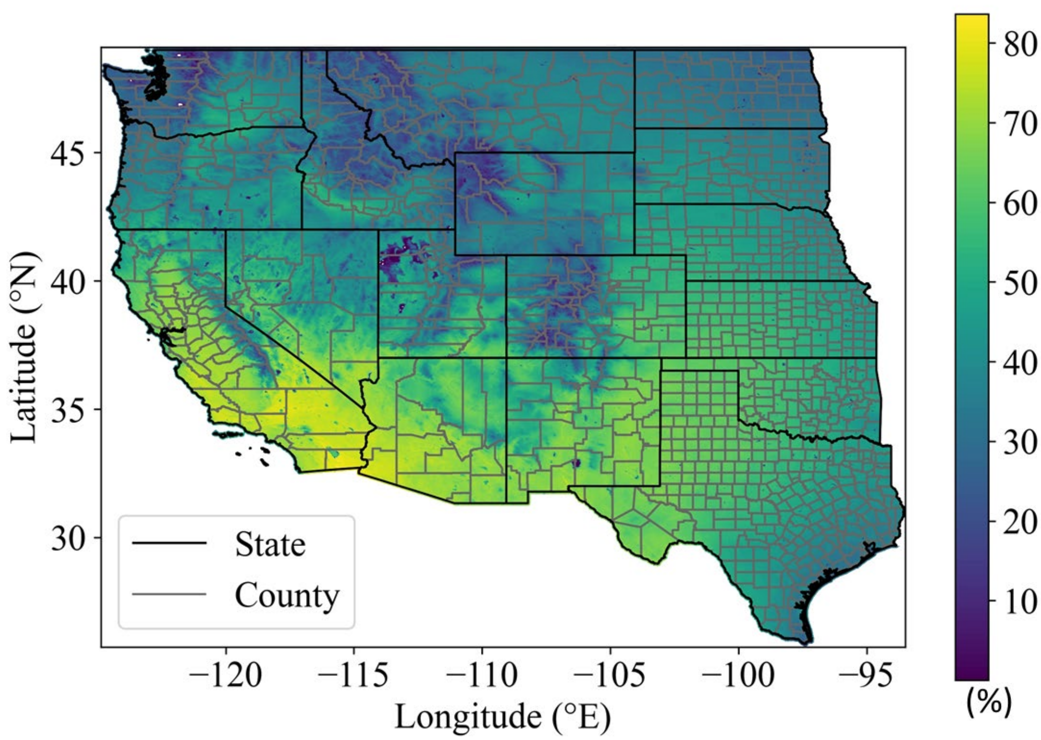

Figure 5.1. Spatial distribution of the availability proportion (%) of daily MAIAC AOD merged from Terra and Aqua observations across the western United States from 2002 to 2019.

5.2 Data for Imputation

The Multi-Angle Implementation of Atmospheric Correction (MAIAC) product provides high resolution (daily, 1-km) AOD retrievals from MODIS Terra and Aqua observations. We used the daily MAIAC AOD product (MCD19A2, Collection 6.1) from MODIS Terra and Aqua as the dependent variable in our imputation models. To maximize observational coverage, same-day Terra and Aqua retrievals were merged using a linear regression approach, following previous studies. The resulting merged MAIAC AOD dataset still contained gaps but provided greater coverage than either satellite alone. Figure 5.2 shows the spatial distribution of the multi-year mean merged MAIAC AOD across the western United States.

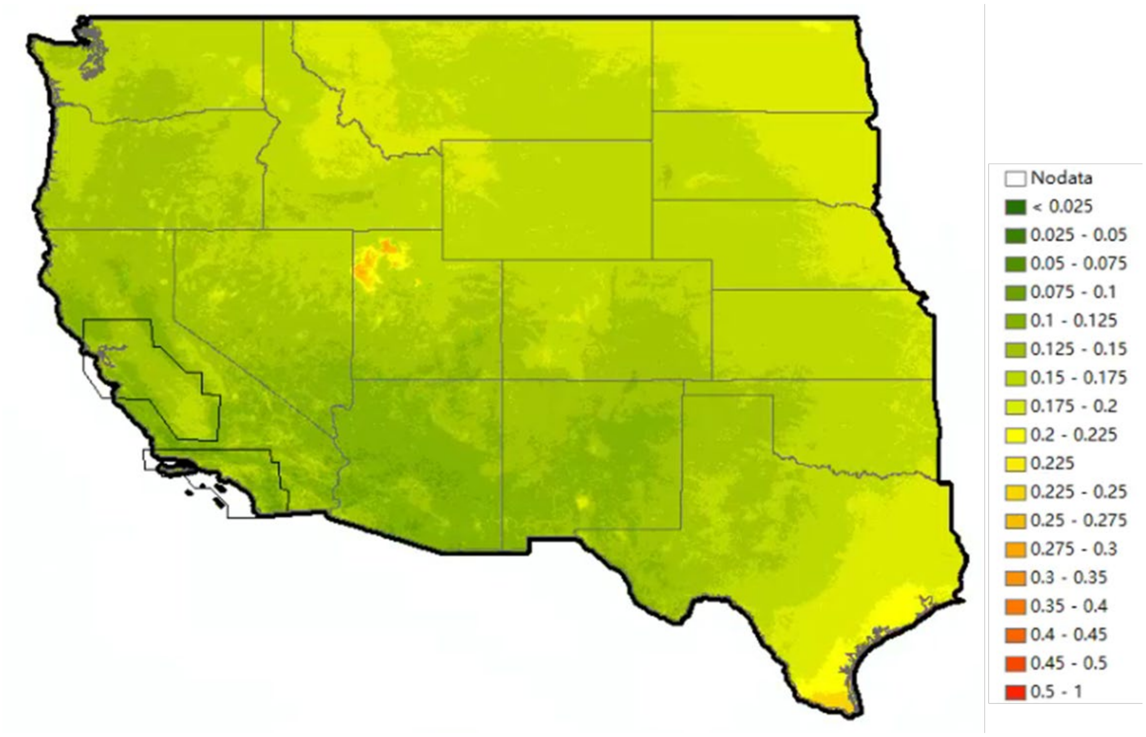

Figure 5.2. Spatial distribution of daily mean MAIAC AOD merged from Terra and Aqua observations across the western United States from 2000 to 2020.

For imputing missing MAIAC AOD values, we used multiple publicly available datasets as explanatory variables. These included satellite aerosol products, reanalysis meteorological fields, and static geophysical data, each with defined spatial and temporal resolutions. Aerosol data:

- MERRA-2 GMI AOD from the 'Aerosol Diagnostics' product (~0.5° × 0.625° spatial resolution, daily).
- Meteorological data (from MERRA-2 GMI 'Daily Average Diagnostics'):
  - 2-m air temperature (daily, ~0.5° × 0.625°)
  - 2-m specific humidity (daily, ~0.5° × 0.625°)
  - 2-m eastward wind component (u-wind, daily, ~0.5° × 0.625°)
  - 2-m northward wind component (v-wind, daily, ~0.5° × 0.625°)
  - Total cloud fraction (daily, ~0.5° × 0.625°)
- Land surface and geophysical data:
  - NDVI from MODIS/Terra vegetation indices (MOD13A2), 1 km resolution, 16-day interval; monthly composites were used.

- Elevation from ASTER GDEM (30 m native resolution, aggregated to 1 km).
- Longitude and latitude of each 1-km grid cell centroid.
- Time index (continuous daily count from the start of the study period).

The data sources and preprocess are detailed in Section 3.

5.3 Modeling Approach and Validation

The overall methodologies followed previous studies (He et al. 2023a; Li et al. 2020).
Previous studies have shown that monthly AOD patterns are promising predictors for representing the spatial and long-term trends of AOD variability in high-resolution imputation (Li et al., 2020). Therefore, where available, the monthly mean MAIAC AOD was included as a predictor to provide additional context on broader temporal patterns. This variable was only available for pixels meeting the >50% valid-day criterion within a given month. Two model branches were therefore implemented:

(1). mAOD models — used for grid cells with valid monthly MAIAC AOD, including 12 predictors (MERRA-2 GMI AOD, meteorological variables, NDVI, elevation, coordinates, time index, and monthly MAIAC AOD).

(2). non-mAOD models — used for grid cells without valid monthly MAIAC AOD, excluding the monthly AOD predictor.

We applied a spatiotemporal random forest modeling approach for each day from 2000 to 2020. A time-stratified sampling method was employed to help each imputation model capture short-term variations in AOD. The model was trained on the 1 km × 1 km grid using three rolling-day samples, with the middle day as the target day. A temporal dummy variable, coded as [1, 2, 3], was used to indicate the position of each day within the three-day window. We did not use a larger wind since previous study tested models using five- and seven-day rolling windows (with the third or fifth day as the target), but sensitivity analyses showed that these did not improve performance and increased computation time.

Approximately 20% of valid daily merged MAIAC AOD retrievals were withheld for hold-out validation, ensuring that performance metrics reflected the ability of the models to generalize to unseen data.

5.4 Results

The AOD imputation models performed well throughout the 2000–2020 period. In hold-out validation, the correlation coefficient (r) between predicted and observed MAIAC AOD ranged from 0.861 to 0.997, with an average of 0.938 (Fig. 5.3). These results demonstrate that the model accurately reconstructed daily AOD patterns across diverse spatial and temporal contexts.

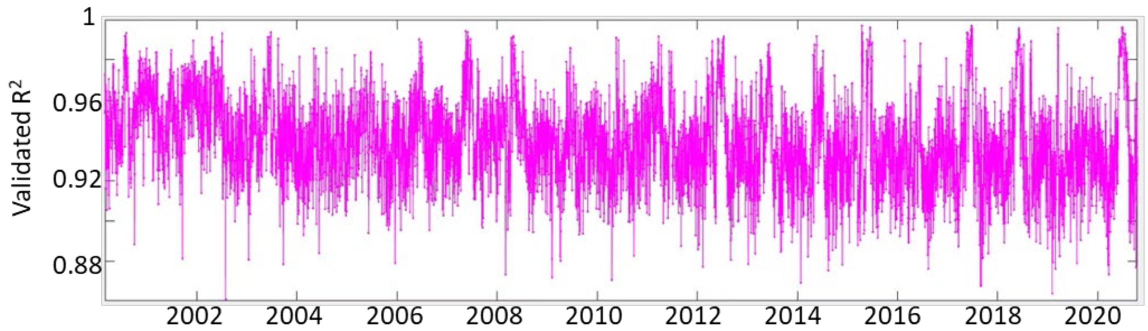

Figure 5.3. Time series of validated R^2 of AOD imputation models from 2000 to 2020.

Spatially, the imputed full-coverage AOD fields closely matched the original merged MAIAC retrievals where available, while filling in large areas with missing data due to cloud cover, snow, or retrieval algorithm limitations. Figure 5.4 compares the merged MAIAC retrievals and our imputed full-coverage AOD for a representative day. This comparison illustrates that our imputation process effectively extends the original retrievals to produce a continuous, high-resolution dataset, while preserving the fine-scale spatial structure of aerosol distributions.

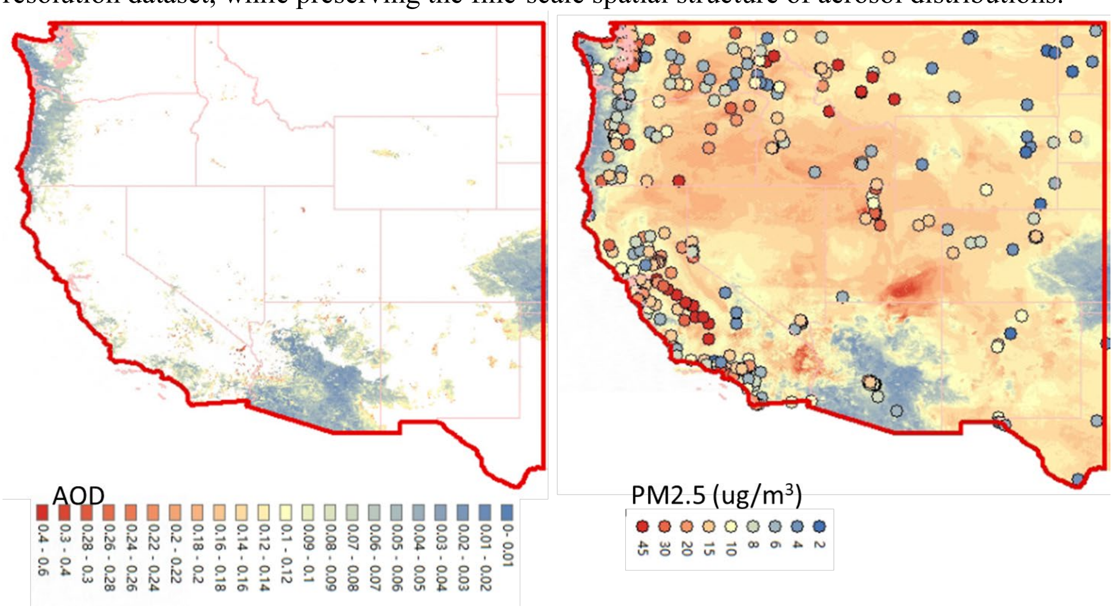

Figure 5.4. Terra-Aqua merged AOD (left) and modeled full-coverage AOD (right) on 02 Jan. 2016.

To construct the final fused AOD dataset, we began with the merged MAIAC retrievals and replaced all missing values with their corresponding imputed estimates. This ensured that observed retrievals were preserved where available, and modeled values were only used to fill gaps. The resulting fused dataset provides a complete, daily 1-km AOD time series for the entire 2000–2020 period (Fig. 5.5), forming a consistent and physically meaningful satellite-based predictor for subsequent $PM_{2.5}$ total mass modeling.

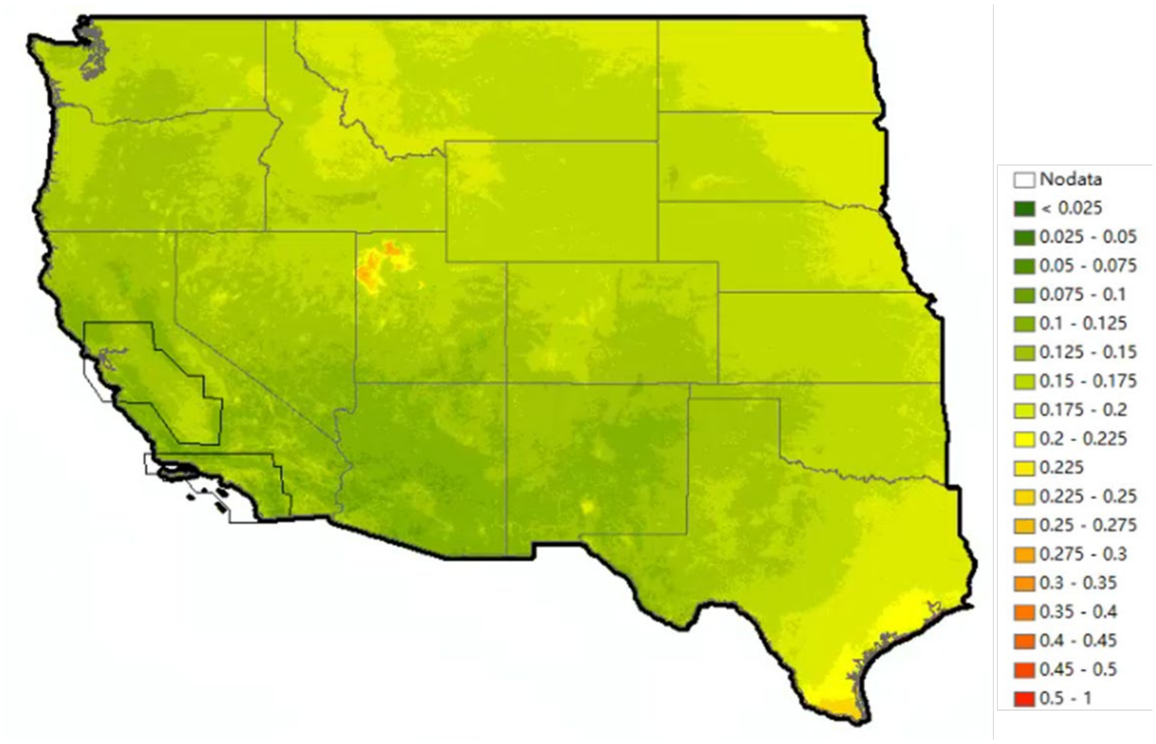

Figure 5.5. Spatial distribution of imputed full-coverage AOD across the western United States from 2000 to 2020.

5.5 Summary

While MAIAC AOD retrievals remain incomplete, they are far more abundant than surface $PM_{2.5}$ observations. This makes AOD imputation inherently easier than $PM_{2.5}$ total mass modeling (achieving higher modeling R^2 of 0.861-0.997 in Fig. 5.3), as the model can be trained on a much larger number of valid samples. In contrast, $PM_{2.5}$ total mass modeling is limited by the relatively small number of ground-based measurements, which constrains the model's ability to learn detailed spatial and temporal patterns. Through the integration of multi-source aerosol, meteorological, and geophysical data in a spatiotemporal random forest framework, we generated a gap-free, daily 1-km MAIAC AOD dataset for 2000–2020 across the western United States. The inclusion of both spatial and temporal context, along with monthly AOD climatology where available, enabled the model to achieve high accuracy in hold-out validation while maintaining the observed spatial and temporal variability of AOD. This fused AOD dataset, combining observed MAIAC retrievals with model-imputed values to ensure complete coverage, served as a critical input to the $PM_{2.5}$ total mass modeling in this project. By providing continuous, high-resolution aerosol information over two decades, the dataset supports more accurate and consistent estimation of $PM_{2.5}$ total mass and its chemical components, enabling robust assessments of spatial patterns, temporal trends, and potential exposure risks.

6. $PM_{2.5}$ Component Modeling during 2002-2019

6.1 Introduction

The major components of $PM_{2.5}$ exhibit substantial differences in their formation mechanisms, emission sources, and atmospheric lifetimes, which lead to strong spatiotemporal heterogeneity. Accurately estimating their concentrations requires integrating multiple data sources and advanced spatiotemporal modeling techniques. As discussed in Section 2, it is necessary to first obtain gap-free daily $PM_{2.5}$ total mass estimates at high spatial resolution to help the model quantify the spatiotemporal variability of major $PM_{2.5}$ components. In this section, we first generated daily, 1-km total $PM_{2.5}$ concentrations across the western United States as a foundational predictor for modeling five components. Our component modeling approach therefore followed a two-stage framework: first estimating total $PM_{2.5}$ from satellite-derived aerosol optical depth (AOD), meteorology, and land cover data, and then using these estimates with CMAQ outputs and additional meteorological variables, and geographical features to predict individual components. Since we only obtain CMAQ-speciated outputs from 2002 to 2019, this section models components from this period. The predictor data, estimation modeling approach, evaluation method and modeling results and discussions are shown in this section.

6.2 Data and Preprocessing

The modeling relied on multiple datasets (Section 3 and Table 3.1), harmonized to a 1-km resolution across the 2002–2019 study period. $PM_{2.5}$ and speciated component measurements were obtained from CSN, IMPROVE, and CASTNET networks (Fig. 3.1), with CSN and IMPROVE data used for model training/validation and CASTNET reserved for independent evaluation of SO_4^{2-} and NO_3^- estimates. Measurement harmonization steps included adjustments for EC and OC differences between networks, and calculation of DUST from elemental composition following IMPROVE protocols.

Satellite AOD was derived from the MODIS MAIAC product and gap-filled to full coverage using a random forest–based imputation approach (detailed in Section 5). CMAQ simulations provided weekly speciated $PM_{2.5}$ concentrations at 12-km resolution, interpolated to 1 km. Additional meteorological data were drawn from DayMet, gridMET, and MERRA-2 GMI; land cover and population from NLCD and LandScan; and topographic metrics from ASTER GDEM. Wildfire activity was represented using daily HMS detections processed with kernel density estimation.

6.3 Modeling $PM_{2.5}$ Components

The modeling framework is outlined in Fig. 6.1 and detailed below.

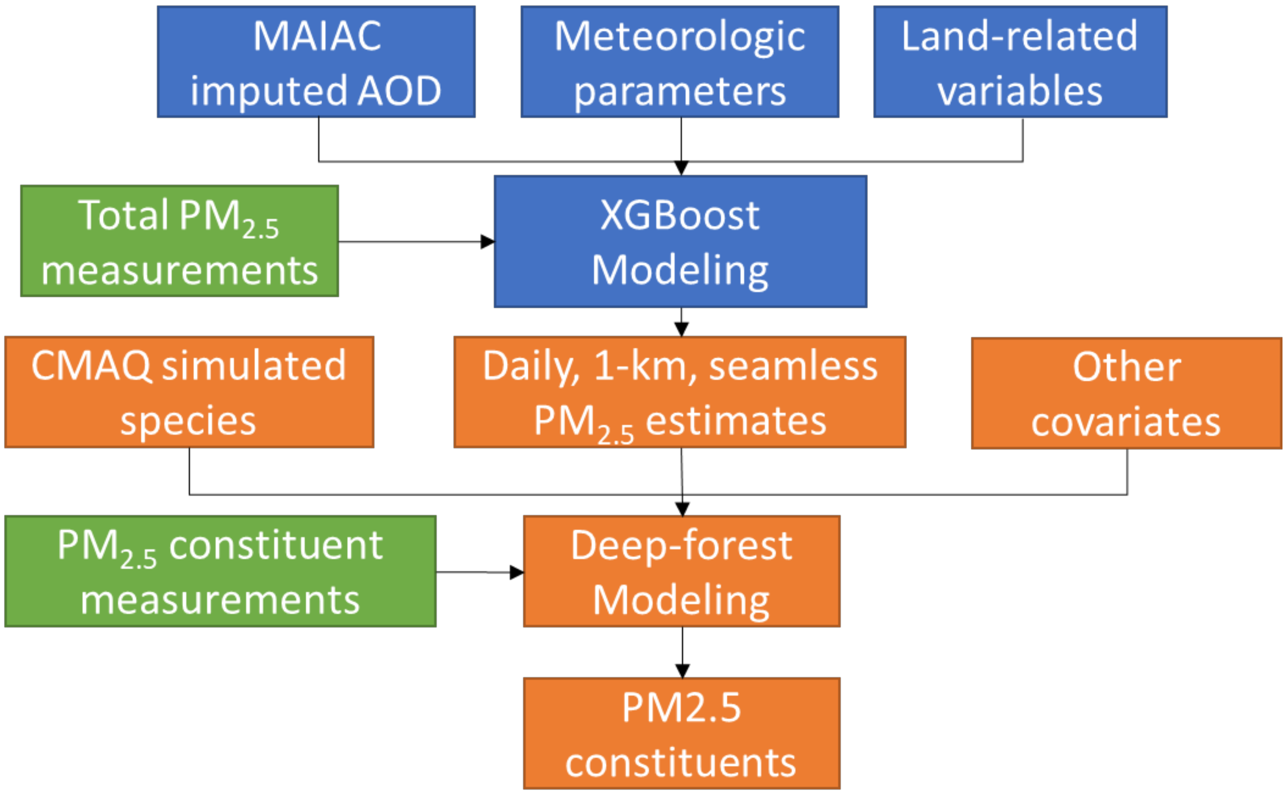

Figure 6.1. Modeling framework of high-resolution $PM_{2.5}$ constituent modeling.

6.3.1 AOD-based Total $PM_{2.5}$ Estimation

Our approach for modeling total $PM_{2.5}$ mass concentrations ($PM_{2.5}$_est) was to train annual XGBoost models on the daily $PM_{2.5}$ concentrations measured at EPA monitoring sites with imputed AOD as the primary predictor, alongside meteorological and land cover covariates. XGBoost (Extreme Gradient Boosting) (Chen and Guestrin 2016) is an efficient and scalable implementation of gradient boosting decision trees, widely used for structured data modeling due to its high predictive accuracy, ability to handle missing data, and built-in regularization to prevent overfitting. Unlike the deep learning algorithm employed for component modeling, XGBoost was chosen here for its computational efficiency in handling large datasets while maintaining comparable model performance. The dataset for total $PM_{2.5}$ was approximately ten times larger than that of an individual $PM_{2.5}$ component due to the EPA sampling schedule. Our preliminary analyses using convolutional neural networks and insights from previous studies (Yang et al. 2022) indicated that deep learning models would not yield significantly better predictions but would substantially increase computational costs. Additionally, preliminary analyses indicated that training models on annual data outperformed a single multi-year model, likely because yearly models better account for year-specific trends and intra-annual variations in emissions, meteorology, and other influencing factors. To mitigate edge effects caused by the spatial boundaries of $PM_{2.5}$ estimates, we expanded the $PM_{2.5}$ modeling area beyond the study region, including entire states along the western edge of the study domain (Fig. 3.2). The model incorporates 18 variables detailed in Table 3.1: the imputed gap-free, 1 km AOD; seven meteorological parameters; and 10 auxiliary variables. Since the HMS fire data were only available from April 2003 onward, the fire density variable was included only in annual models for 2004–2019. We also added an indicator variable for day of the year to help capture day-to-day variations. We also constructed a temporal dummy variable, day of year, to indicate the seasonal variations in the data.

A Bayesian optimization approach was used to tune five key hyperparameters for each model: n_estimators, max_depth, learning_rate, subsample, and colsample_bytree. To mitigate overfitting, optimal hyperparameters were selected based on spatial cross-validation performance. Additionally, early stopping was implemented during training using a 30% validation subset, terminating training when validation performance no longer improved. Final model performance was assessed using 10-fold cross-validation. The XGBoost for total $PM_{2.5}$ yielded a 10-fold CV R^2 of 0.81 and an RMSE of 3.43 μg/m3 across the study period (Fig. 6.2). These results indicate that total $PM_{2.5}$ mass concentration estimates (referred to as $PM_{2.5}$est hereafter) align well with monitoring measurements, thus providing confidence in their use as a predictor in $PM_{2.5}$ component models.

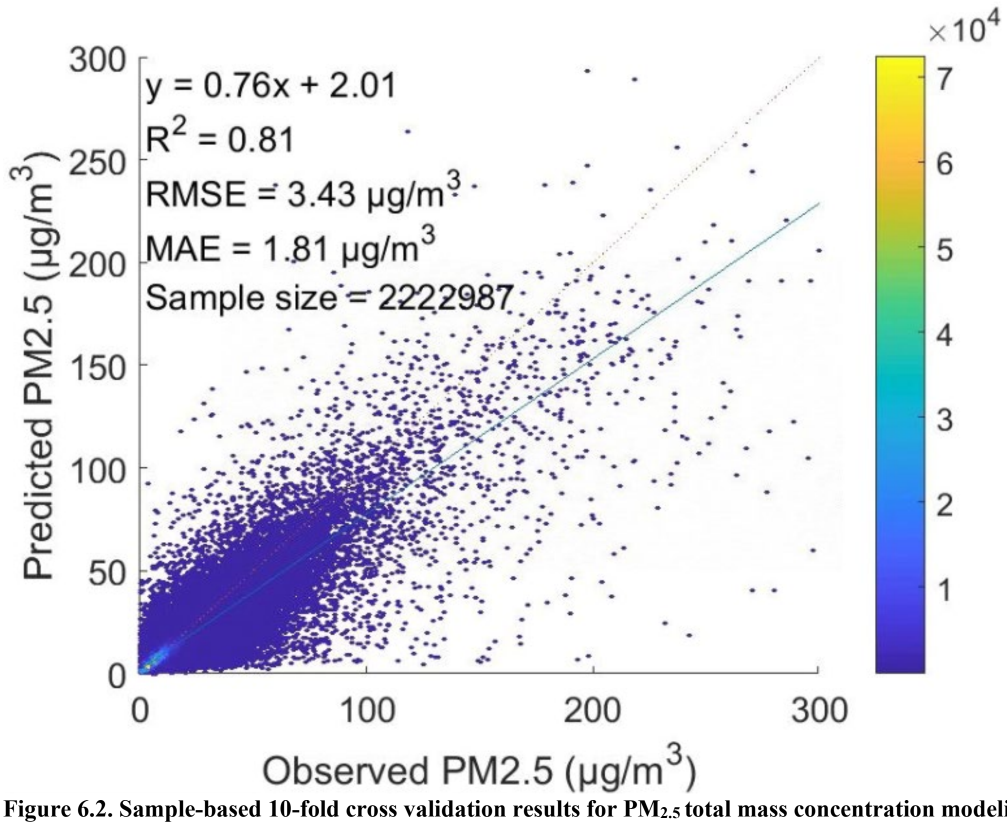

Figure 6.2. Sample-based 10-fold cross validation results for $PM_{2.5}$ total mass concentration modeling.

6.3.2 Feature Selection

We applied Pearson correlation analysis to remove redundant predictors with high intercorrelation ($|r| > 0.6$ for meteorological/geographical variables and $|r| > 0.8$ for CMAQ outputs) but weak correlation with observed component concentrations. This process resulted in 21 predictors in the models for SO_4^{2-}, NO3−, OC, and DUST, and 20 predictors for EC. The predictors included: five CMAQ species outputs, 17 meteorological variables, and five geographical features. The specific predictor variables used for each model are provided in Table 6.1.

Table 6.1. The final set of features used for each component modeling and their correlations with ground-level observations.

Category		SO_4^{2-}		NO_3^-		EC		OC		DUST
	Feature	obs	Feature	obs	Feature	obs	Feature	obs	Feature	obs
Total $PM_{2.5}$	$PM_{2.5}$ est	0.467	$PM_{2.5}$ est	0.636	$PM_{2.5}$ est	0.716	$PM_{2.5}$ est	0.728	VPD	0.416
CMAQ	SO_4_CMAQ	0.596	NO_3_CMAQ	0.606	EC_CMAQ	0.576	EC_CMAQ	0.347	SO_4_CMAQ	0.200
	NH_4_CMAQ	0.413	EC_CMAQ	0.416	NO_3_CMAQ	0.357	OC_CMAQ	0.305	EC_CMAQ	0.095
	EC_CMAQ	0.336	SO_4_CMAQ	0.225	OC_CMAQ	0.251	NO_3_CMAQ	0.214	OC_CMAQ	0.065
	OC_CMAQ	0.186	OC_CMAQ	0.158	SO_4_CMAQ	0.231	SO_4_CMAQ	0.155	NH_4_CMAQ	0.059
Meteorological & geographical variables	Tmin	0.369	NLCD_D	0.352	NLCD_D	0.454	NLCD_D	0.254	$PM_{2.5}$ est	0.309
	ELE	-0.288	ELE	-0.247	ELE	-0.236	TLML	0.185	SRAD	0.296
	NLCD_D	0.278	PBLH	-0.194	WS10M	-0.162	ELE	-0.158	PBLH	0.283
	Trad	0.246	WS10M	-0.158	Tmax	0.137	WS10M	-0.140	CLDTOT	-0.189
	CLDTOT	-0.167	DayL	-0.125	PBLH	-0.109	CLDTOT	-0.086	PRECTOT	-0.144
	WS10M	-0.164	NDVI	-0.090	NLCD_G	-0.097	Rmin	-0.082	NLCD_D	0.140
	PRECTOT	-0.151	NLCD_G	-0.072	DayMet_swe	-0.081	PRECTOT	-0.074	NDVI	-0.122
	Rmin	-0.135	PRECTOT	-0.065	PRECTOT	-0.073	SWE	-0.073	V2M	0.083
	SWE	-0.102	SWE	-0.060	U2M	-0.060	NDVI	0.059	U2M	0.077
	V2M	0.074	Rmin	0.060	WVP	0.056	Trad	0.055	NLCD_G	-0.061
	PBLH	0.062	NLCD_W	-0.049	CLDTOT	-0.052	U2M	-0.051	WVP	0.061
	NDVI	-0.054	Tmin	0.020	NLCD_W	-0.050	PBLH	-0.044	SWE	-0.054
	NLCD_G	-0.048	U2M	-0.018	WD	-0.035	NLCD_G	-0.043	ELE	-0.051
	NLCD_W	-0.044	WD	-0.012	NDVI	-0.030	NLCD_W	-0.031	WS10M	-0.024
	U2M	0.020	CLDTOT	-0.010	V2M	-0.025	V2M	-0.027	NLCD_W	-0.020
	WD	-0.004	V2M	-0.010			WD	-0.008	WD	0.001

6.3.3 Spatiotemporal deep-forest model for $PM_{2.5}$ Component estimation

Our $PM_{2.5}$ component estimation framework employs a spatiotemporal deep-forest model. Deep forest (Zhou and Feng 2019) is a decision tree-based deep learning algorithm having a layered structure of decision forests. This approach merges the benefits of traditional ensemble methods, like random forests, with a multi-level hierarchical design. Unlike neural network-based deep learning models, which often demand extensive hyperparameter tuning, deep forest delivers strong predictive performance with much less tuning required. Given the substantial heterogeneity in $PM_{2.5}$ components due to their complex sources and variations relative to total $PM_{2.5}$, along with the high computational demands posed by the large spatial domain, the long study period, and the high spatiotemporal resolution, we employed deep forest for its powerful, effective, and efficient data-mining capabilities. Comparisons with other models, including random forest and related ensemble methods, are provided in Section 6.5.2. To enhance the model's spatial and temporal predictive power, we constructed additional spatiotemporal heterogeneity features, obtaining the spatiotemporal deep-forest model. Spatial constructed features included Haversine distances—which measure the great-circle distance between two points on a sphere based on their latitude and longitude—from each grid cell to the upper-left (HavdLU), upper-right (HavdRU), lower-left (HavdLD), lower-right (HavdRD), and central points (HavdC) of the study region (see the detailed equations for those constructed features in previous studies (Wei et al. 2020). Temporal variability was accounted for by including the year and day of the year (DOY) as additional features.

We trained a spatiotemporal deep-forest model for each $PM_{2.5}$ component using measured $PM_{2.5}$ component concentrations from the CSN and IMPROVE networks as the dependent variables and AOD-derived $PM_{2.5}$ and CMAQ speciation outputs as the primary predictors, along with meteorological variables and geographical features. Each $PM_{2.5}$ component model was trained on samples from all years combined, rather than individually for each year, as the sample size from a single year was insufficient to build a robust model. The hyperparameters were automatically adjusted by the built-in self-adaptive mechanism during training (Zhou and Feng 2019).

6.3.4 $PM_{2.5}$ Component Modeling Evaluation

Model evaluation employed random (sample-based), spatial (site-based), and temporal (day-based) 10-fold CV (Table 6.2). For each CV type, all samples, sites, or days were randomly and equally divided into 10 subsets. In each iteration, data from 9 subsets were used for model training, while the remaining subset was used for validation. This process was repeated 10 times to ensure that all subsets were used for validation (see Table S4 for a summary of the three CV types). To assess model accuracy, we calculated the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE) by comparing the model estimates (predictions) to the observed measurements. Additionally, to comprehensively assess model outputs, we used weekly measurements from the CASTNET network to independently validate SO_4^{2-} and NO3− estimates. Daily estimates were averaged across the start and end days of the CASTNET sampling window (i.e., "DayON" and "DayOFF" fields) to obtain weekly-aggregated data. These three statistical indicators were calculated both for all estimate-measurement pairs and specifically for pairs at CASTNET sites not located near CSN and IMPROVE networks.

Table 6.2. Summary of CV methods for constituent model validation.

CV approach	Description

Random CV or sample-based 10-fold CV	Site-day samples were randomly divided into 10 folds, with 9 folds used for model training and the remaining fold used for validation.
Spatial CV or site-based 10-fold CV	Site-day samples were randomly grouped into 10 folds based on site ID, with 9 folds used for model training and the remaining fold used for validation.
Temporal CV or day-based 10-fold CV	Site-day samples were randomly grouped into 10 folds based on date, with 9 folds used for model training and the remaining fold used for validation.

6.3.5 Model Interpretation Using SHAP

To interpret model predictions, Shapley Additive Explanations (SHAP) were applied, quantifying the magnitude and direction of each predictor's contribution (Figs. 3, S8). $PM_{2.5}$_est consistently ranked as the most influential predictor across all components, followed by the corresponding CMAQ species output and key meteorological variables such as temperature, planetary boundary layer height, and vapor pressure deficit.

6.4 Results and Discussions

6.4.1 Descriptive Statistics of Modeling Data

Descriptive statistics of the concentrations measured at CSN and IMPROVE networks in the western United States from 2002 to 2019 are summarized in Table 6.3. The 18-year average concentrations across the study region were 0.67, 0.62, 0.25, 1.19, and 0.77 μg/m³ for SO_4^{2-}, NO_3^-, EC, OC, and DUST, respectively. Over time, concentrations exhibited decreasing trends, declining from 0.86, 0.91, 0.36, 1.66, and 0.96 μg/m³ in 2002 to 0.48, 0.49, 0.26, 0.91, and 0.64 μg/m³ in 2019. In addition, the southwestern United States experienced relatively higher pollution levels, particularly in California, where NO_3^-concentrations were approximately five times higher, and the other four components were 1.5 to 2 times higher compared to the northwest.

Table 6.3. Descriptive statistics of $PM_{2.5}$ component observations from CSN and IMPROVE networks in western United States from 2002 to 2019.

Year	SO_4^{2-}		NO_3^-		EC		OC		DUST	
	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std
2002	0.86	0.88	0.91	2.67	0.36	0.64	1.66	3.42	0.96	1.46
2003	0.81	0.80	0.72	1.87	0.36	0.58	1.60	3.05	0.83	1.48
2004	0.88	1.02	0.85	2.25	0.35	0.59	1.39	2.51	0.76	1.07
2005	0.90	0.97	0.79	2.24	0.41	0.66	1.41	2.15	0.70	1.13
2006	0.77	0.79	0.70	2.00	0.39	0.67	1.43	2.35	0.85	1.19
2007	0.81	0.85	0.78	2.27	0.23	0.51	1.16	2.23	0.82	1.25
2008	0.76	0.70	0.63	1.69	0.20	0.43	1.11	2.53	0.89	1.36
2009	0.72	0.70	0.65	1.79	0.22	0.55	1.09	2.59	0.84	1.20
2010	0.61	0.58	0.54	1.44	0.14	0.26	0.74	1.27	0.74	1.27
2011	0.65	0.60	0.56	1.67	0.15	0.28	0.82	1.60	0.71	1.19
2012	0.62	0.54	0.50	1.38	0.16	0.33	1.01	2.46	0.89	1.28

2013	0.61	0.55	0.63	2.04	0.14	0.28	0.81	1.55	0.74	1.17
2014	0.55	0.51	0.51	1.63	0.13	0.24	0.77	1.31	0.76	1.17
2015	0.53	0.48	0.49	1.42	0.14	0.33	1.04	2.52	0.62	0.83
2016	0.45	0.47	0.42	1.25	0.18	0.35	0.95	1.55	0.65	0.88
2017	0.50	0.47	0.47	1.50	0.26	0.63	1.60	4.68	0.67	0.93
2018	0.48	0.46	0.52	1.53	0.30	0.69	1.52	4.01	0.80	1.12
2019	0.48	0.45	0.49	1.27	0.26	0.44	0.91	1.06	0.64	0.96
18 yrs	0.67	0.70	0.62	1.82	0.25	0.51	1.19	2.60	0.77	1.18

Ave and Std represent the mean and standard deviation, respectively.

6.4.2 $PM_{2.5}$ Component Modeling Results

6.4.2.1 Cross-validation results

Across the five $PM_{2.5}$ components, there was a total of 255,691 to 283,206 site- and day-specific samples. Figure 6.3 presents the daily random, spatial, and temporal CV results for the five $PM_{2.5}$ components, along with maps showing model R^2 and RMSE at each monitoring site based on the random CV results. Generally, the daily estimates aligned well with ground-level measurements, showing high sample-based CV R^2 values of 0.81, 0.89, 0.75, 0.66, and 0.75 and low RMSE values of 0.30, 0.59, 0.26, 1.52, and 0.59 μg/m³ for SO_4^{2-}, NO_3^-, EC, OC, and DUST, respectively. Approximately 78%, 63%, 66%, 82%, and 51% of stations demonstrated strong correlations with ground-level measurements, as indicated by random CV R^2 values exceeding 0.6 (Fig. 6.3b). Spatial and temporal CV results reveal similar error patterns as the random CV, though with decreased performance, particularly in spatial CV, where R^2 values dropped to 0.76, 0.85, 0.65, 0.60, and 0.61 for the five components (Fig. 6.3a). We aggregated the three types of CV results over time for the five species. As shown in Table 6.4, the monthly- and yearly-aggregated CV results indicate that uncertainties in components estimates significantly decrease when estimates are aggregated: at the monthly scale, R^2 values increased (RMSE decreased) to 0.89–0.97 (0.11–0.64 μg/m³) for random CV, 0.70–0.93 (0.20–0.95 μg/m³) for spatial CV, and 0.89–0.97 (0.11–0.64 μg/m³) for temporal CV.

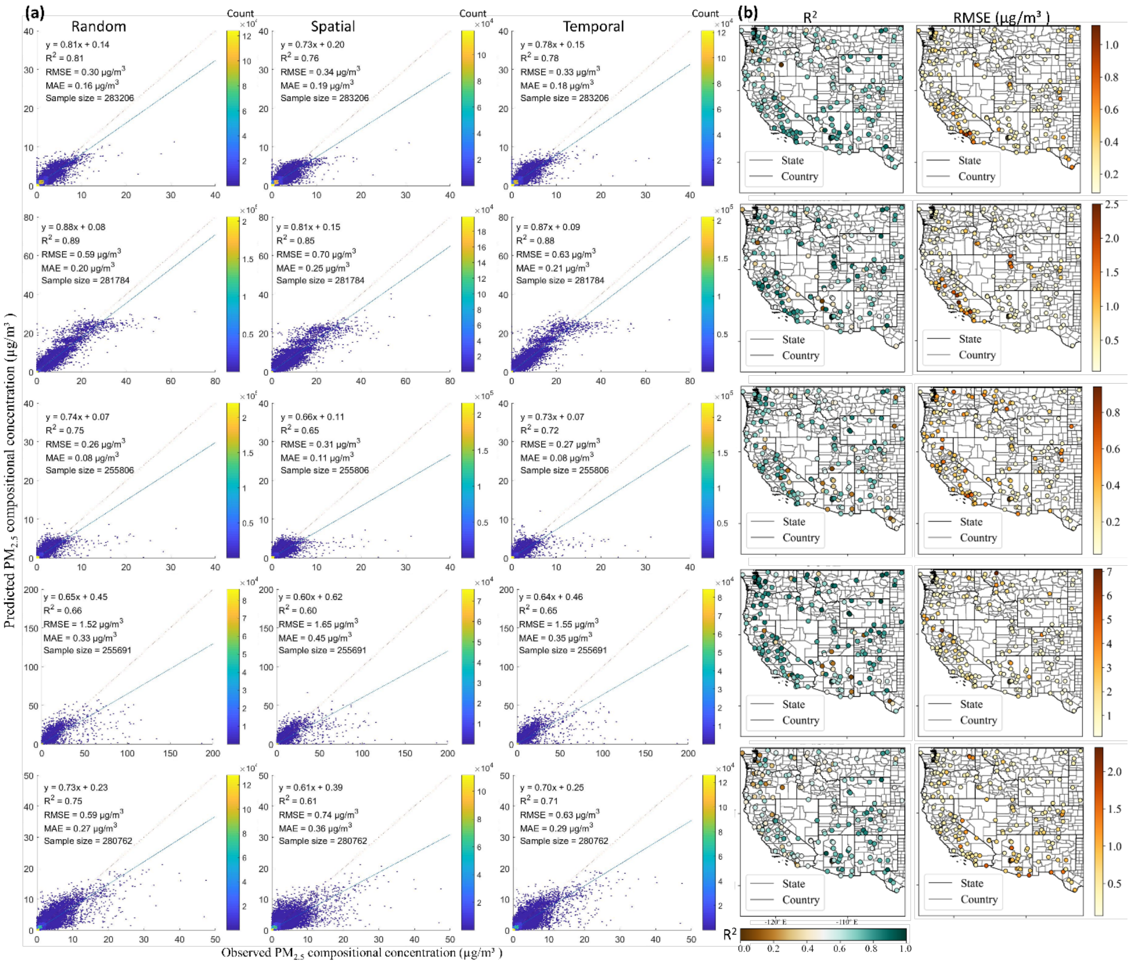

Figure 6.3. Validation results for $PM_{2.5}$ components models: (a) Scatterplots showing 10-fold CV results using random (sample-based), spatial (site-based), and temporal (day-based) partitions; (b) Spatial maps depicting the model performance (R^2 and RMSE) based on daily $PM_{2.5}$ component estimates compared to observations at each monitoring site over the period 2002–2019.

Table 6.4. Summary statistics of time-aggregated validation results for constituent modeling

10-fold cross-validation											
CV type	Metric	Monthly (Grid cell-month pairs)					Yearly (Grid cell-year pairs)				
		SO_4^{2-}	NO_3^-	EC	OC	DUST	SO_4^{2-}	NO_3^-	EC	OC	DUST
Random	R^2	0.95	0.97	0.94	0.89	0.91	0.95	0.99	0.98	0.97	0.97
	RMSE	0.13	0.25	0.11	0.64	0.27	0.10	0.10	0.05	0.20	0.12
	MAE	0.08	0.10	0.04	0.19	0.15	0.04	0.05	0.02	0.11	0.08
Spatial	R^2	0.88	0.93	0.80	0.76	0.70	0.90	0.95	0.85	0.83	0.72
	RMSE	0.20	0.42	0.20	0.95	0.49	0.15	0.26	0.15	0.52	0.36
	MAE	0.11	0.19	0.09	0.37	0.28	0.08	0.14	0.08	0.30	0.22
Temporal	R^2	0.94	0.97	0.94	0.89	0.89	0.94	0.99	0.98	0.97	0.97
	RMSE	0.14	0.27	0.11	0.64	0.29	0.12	0.10	0.05	0.21	0.13

	MAE	0.08	0.11	0.05	0.20	0.16	0.04	0.05	0.03	0.11	0.08
Pair size		32646	32350	28513	28501	32486	2814	2781	2454	2453	2796
CASTNET-based independent validation at yearly level											
Metric	R^2	RMSE	MAE		Metric	R^2	RMSE	MAE		Pair size	
SO_4^{2-}	0.81	0.09	0.06		NO_3^-	0.78	0.16	0.10		111	

The regression lines between model estimates and observed concentrations from the three cross-validation approaches suggest minor underestimation, particularly for SO_4^{2-} and NO_3^-, with slopes of 0.81 and 0.88, respectively. The magnitudes of these slopes are comparable to those reported in other PM component modeling studies (Liu et al. 2022). To further assess whether the potential underestimation affects the fraction of total $PM_{2.5}$ mass captured by the model, we compared the sum of our final predictions for the five targeted species with both (1) the sum of their observed concentrations (Fig. 6.4(a)) and (2) the total $PM_{2.5}$ mass measured at monitoring sites (Fig. 6.4(b)). For reference, we also compared total $PM_{2.5}$ observations with the sum of the five component observations, which yielded an R^2 of 0.92, RMSE of 2.66 μg/m³, and a regression slope of 0.66 (Fig. 6.4(c)). Our predicted component sum showed strong agreement with the observed component sum (R^2 = 0.98, RMSE = 0.62 μg/m³, slope = 0.99), and tracked total $PM_{2.5}$ concentrations with comparable accuracy to the observed component sum (R^2 = 0.91 vs. 0.92, RMSE = 2.66 μg/m³ and slope = 0.66 in both cases), indicating that the model captures the expected fraction of total $PM_{2.5}$ accounted for by the five major components. These results indicate that the minor underestimation observed in individual species does not compound when components are aggregated, suggesting that our component models do not systematically underpredict total component mass.

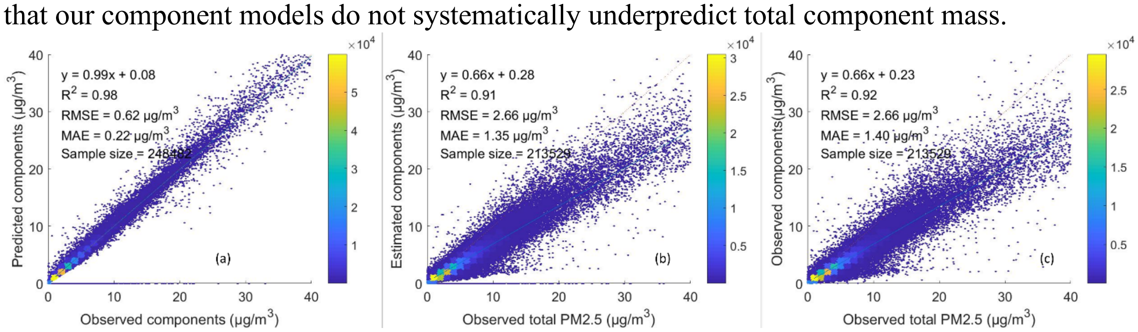

Figure 6.4. Scatterplots showing (a) observed vs. predicted sum of the five targeted $PM_{2.5}$ components, (b) observed total $PM_{2.5}$ mass vs. predicted sum of the five targeted components, and (c) observed total $PM_{2.5}$ mass vs. observed sum of the five targeted components.

6.4.2.2 Independent validation results

Figure 6.5 presents the independent validation results of daily SO_4^{2-} and NO_3^- estimates against CASTNET measurements. For all estimate-measurement comparisons, including at CASTNET sites near (>25 km) CSN and IMPROVE locations, our estimates show good agreement, with R^2 and RMSE values of 0.89 and 0.19 μg/m³ for SO_4^{2-} and 0.62 and 0.34 μg/m³ for NO_3^-, respectively (Fig. 6.5a). When considering only sites located at least 25 km from CSN and IMPROVE locations, our estimates still demonstrate reasonable agreement, with R^2 and RMSE values of 0.74 and 0.17

μg/m³ for SO_4^{2-} and 0.43 and 0.36 μg/m³ for NO_3^- (Fig. 6.5b). Time series plots of weekly estimates and measurements at selected monitoring sites—located at least 25 km from a CSN or IMPROVE site and spanning all study years—show similar temporal patterns, with a strong correlation (r=0.85) for SO_4^{2-} (Fig. 6.5c) and a lower correlation for NO_3^- (r = 0.42). Compared to SO_4^{2-}, the lower performance observed for NO_3^- is primarily attributed to the greater discrepancies between CASTNET and CSN/IMPROVE measurements for this species. Since CASTNET data were not used for model training, inter-network differences likely contribute to the weaker validation results. The observed inter-network correlation coefficients are 0.92 for SO_4^{2-} and 0.83 for NO_3^-, indicating larger inconsistencies for NO3−. Additionally, the lower model performance for NO_3^- is partially due to its lower ambient concentrations (0.19 μg/m³ for NO_3^- vs. 0.44 μg/m³ for SO_4^{2-} in screened observations), which increases relative uncertainty in both measurements and estimates. At lower concentrations, the signal-to-noise ratio diminishes, making small deviations more impactful on correlation metrics. At the annual aggregated scale, observed and predicted concentrations exhibit strong agreement, with R^2 increasing to 0.81 for SO_4^{2-} and 0.78 for NO_3^- and RMSE decreasing to 0.09 μg/m³ and 0.16 μg/m³, respectively (Table 6.4). Overall, together with the CV results presented in Section 6.4.2.1, the validation results demonstrate that our $PM_{2.5}$ components models robustly capture long-term variations in SO_4^{2-} and NO_3^- concentrations across diverse spatial contexts.

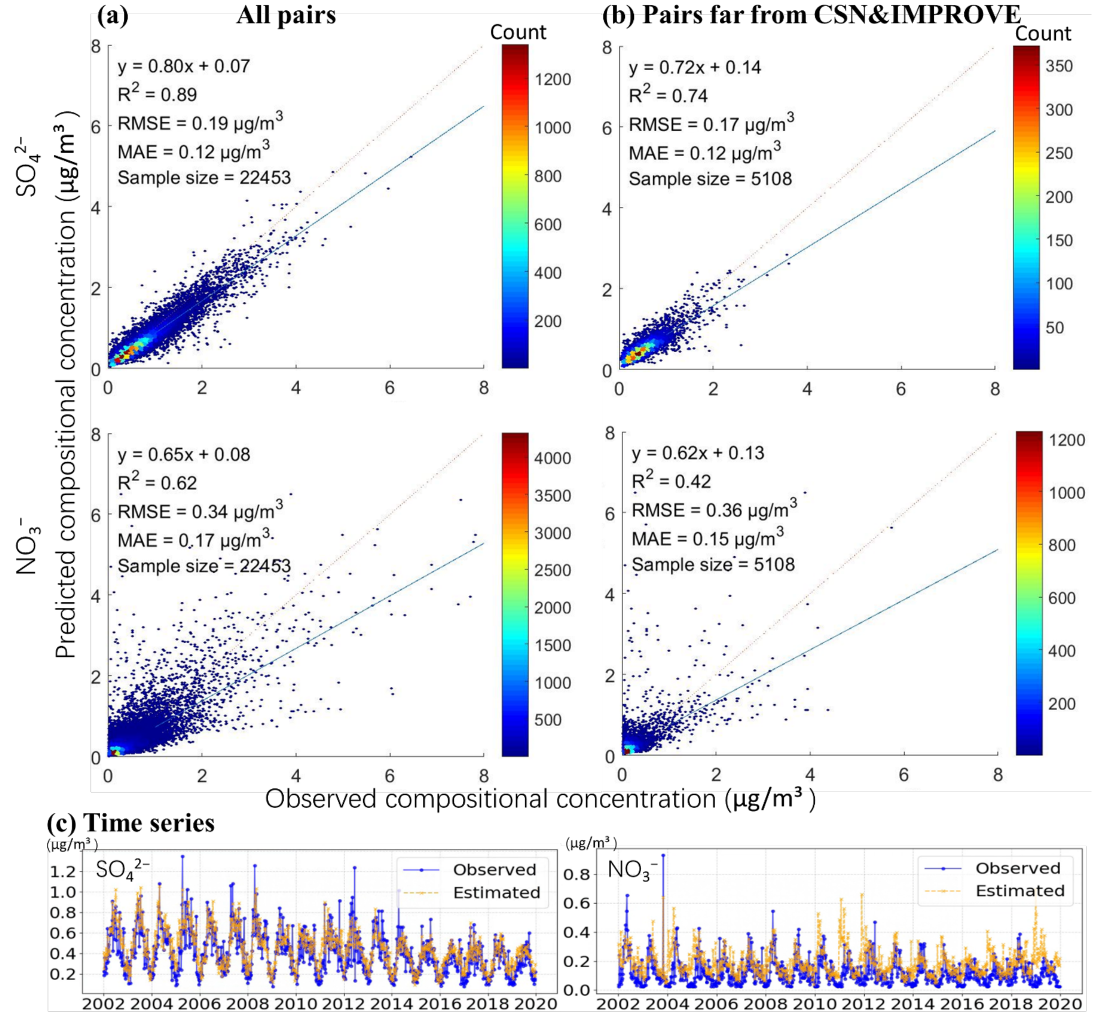

Figure 6.5. Scatterplots comparing observed and estimated $PM_{2.5}$ component concentrations (a) at all CASTNET monitoring sites and (b) at sites 25-km distant from CSN and IMPROVE networks; (c) Time series of observed versus estimated component concentrations at CASTNET sites located at least 25 km from CSN and IMPROVE networks.

6.4.3 $PM_{2.5}$ Component Modeling Results

In Fig. 6.6, we present the SHAP values to illustrate both the local and average contributions of each predictor to each $PM_{2.5}$ component. The predicted daily, high-resolution $PM_{2.5}$ total mass consistently ranked first, showing the highest average contribution across all monitoring stations over the 18-year study period. Its SHAP values were generally positive at higher $PM_{2.5}$ levels, indicating a strong positive influence. This was followed by the CMAQ species output corresponding to the target component; for example, the $SO_4^{2-}PM_{2.5}$ concentration from CMAQ ranked second. Meteorological parameters—especially temperature and planetary boundary layer height (PBLH)—are important predictors for mapping high-resolution $PM_{2.5}$ component concentrations. Typically, temperature is particularly important for the SO_4^{2-} and OC components, where it is the most influential meteorological variable, ranking third and sixth in SHAP contributions, respectively. PBLH is the most important meteorological predictor for NO_3^- and EC, ranking fourth for both. VPD is especially important for the DUST component, with the second-highest SHAP contribution.

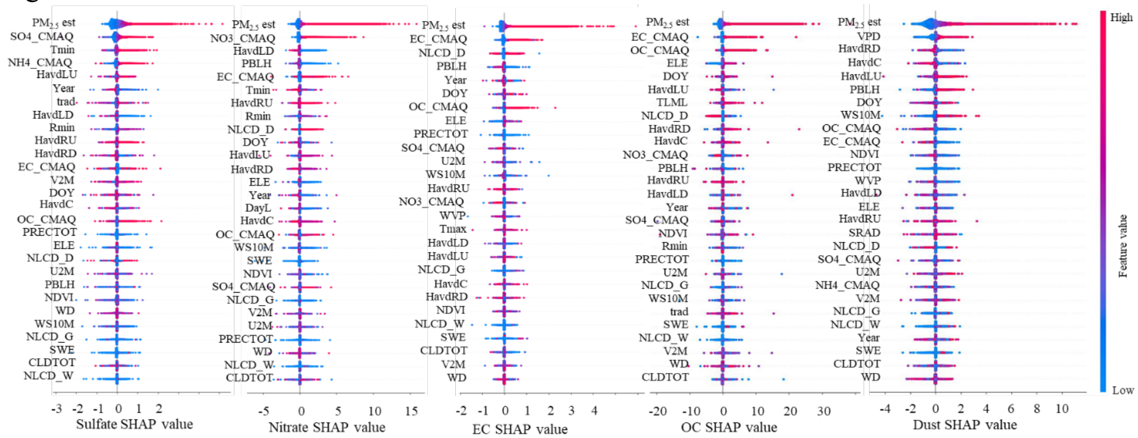

Figure 6.6. Local contribution of each predictor to each $PM_{2.5}$ component, quantified using the SHAP method. Colors range from red to blue, representing high to low normalized values of each predictor, respectively. The full names corresponding to each abbreviation are detailed in Section 3 and also summarized in Table 3.1.

6.5 Comparisons and advantages of the proposed $PM_{2.5}$ component modeling

6.5.1 Comparisons with prior studies

We developed a spatiotemporal deep-forest model that produces the first $PM_{2.5}$ component dataset for the western United States with both fine spatial (1 km) and temporal (daily) resolution over an extended period (2002–2019). The performance of our component models is superior to, or at least comparable with, those of existing models used in the U.S (Table 6.5). For example, the hybrid geoscience-statistical model by Donkelaar et al. (2024) reported 10-fold CV R^2 values of 0.77, 0.76, and 0.48 for SO_4^{2-}, NO_3^-, and DUST, respectively, at the bi-weekly level over North America. Similarly, the ensemble machine-learning models by Amini et al. (2022) achieved test R^2 values of 0.86–0.96 for SO_4^{2-}, NO_3^-, EC, and OC at a coarser temporal resolution of the annual level across the U.S. Geng et al. (2020) predicted daily ground-level $PM_{2.5}$ SO_4^{2-}, NO_3^-, OC, and EC concentrations in California at a 1-km spatial resolution, reporting out-of-bag R^2 values of 0.72, 0.70, 0.68, and 0.70, respectively. Di et al. (2016) applied a backpropagation neural network to predict $PM_{2.5}$ components at a 1 × 1 km resolution in the northeastern United States, achieving R^2 values around 0.70–0.80 for individual components. By comparison, even when modeling across the western United States—a region where previous studies have typically reported lower modeling accuracy (Donkelaar et al. 2019), our model produced estimates with simultaneously high spatial and temporal resolution (daily, 1-km resolution) and demonstrated improved accuracy at both the daily (CV R^2 = 0.66–0.89; Fig. 6.3) and annual levels (CV R^2 = 0.95–0.99; Table 6.4). Given differences in validation methods, modeling domains, and study periods between our work and prior studies, we find that our model is at least comparable in terms of estimation performance alone, though we while not claiming it definitively outperforms them.

Table 6.5. Comparison of $PM_{2.5}$ component modeling studies conducted in the United States

Study	Spatial Resolution	Temporal Resolution	Geographic Coverage	Time Span	Notes
Donkelaar et al. (2024)	1-km	Bi-weekly	North America	2000-2022	10-fold CV R^2: SO_4^{2-} (0.77), NO_3^- (0.76), DUST (0.48)
Amini et al. (2022)	50-m for urban, 1-km for rural	Annual	United States	2000-2019	Test R^2: 0.86–0.96 for SO_4^{2-}, NO_3^-, EC, OC
Geng et al. (2020)	1 km	Daily	California	2005-2014	Out-of-bag R^2: SO_4^{2-} (0.72), NO_3^- (0.70), OC (0.68), EC (0.70)
Di et al. (2016)	1 km	Daily	Northeastern United States	2000-2010	$R^2 \approx 0.70–0.80$ for individual components
Meng et al. 2018a	4.4 km	Daily (global revisit every 9 days)	Southern California	2001–2015	GAMs explained 55–66% daily variability for SO_4^{2-}, NO_3^-, OC, EC
Meng et al. 2018b	0.25º × 0.3125º	Daily	Conterminous United States	2005–2015	Random forest; OOB R^2: SO_4^{2-} (0.86), NO_3^- (0.82), OC (0.71), EC (0.75)
Our study	**1 km**	**Daily**	**Western United States**	**2002–2019**	**Spatiotemporal deep-forest: random CV R^2: SO_4^{2-} (0.81), NO_3^- (0.89), EC (0.75), OC (0.66), DUST (0.75); 0.95–0.98 at the annual level**

6.5.2 Advantages and comparisons with other machine-learning models

A key reason for the improved performance of our species-level models lies in the physical and chemical relevance of the predictors, particularly the inclusion of total ground-level $PM_{2.5}$ mass concentrations ($PM_{2.5}$_est). $PM_{2.5}$ is a chemically complex mixture composed of multiple species, and its total mass often covaries with these components through shared emission sources and atmospheric transformation processes (Hand et al. 2014). The high-resolution $PM_{2.5}$_est, derived from satellite AOD and calibrated using an extensive ground monitoring network, directly captures surface-level aerosol mass and provides a chemically meaningful anchor for estimating the behavior of individual components, with moderate correlation coefficients ranging from 0.31 to 0.73 (Table 2.1). In contrast, AOD reflects total column aerosol loading and lacks vertical specificity, resulting in weaker and less consistent associations with near-surface species concentrations. This is supported by model performance comparisons in Tables 6.4 and 6.6, where the inclusion of the imputed AOD alone provided only modest improvements (e.g., random CV R^2 increases of 0.1–0.5 across species). In contrast, replacing AOD with $PM_{2.5}$_est resulted in consistently higher predictive accuracy, and SHAP analysis (Figs. 6.6) further confirms that $PM_{2.5}$_est is the most influential predictor across all five components, contributing 24–37% of the model's explanatory power. These findings validate our two-stage strategy of first estimating $PM_{2.5}$ total mass and then using it as a key predictor for species modeling.

Table 6.6. Cross validation results of constituent models with various model structures.

Model	Metric	SO_4^{2-}			NO_3^-			EC			OC			DUST		
CV method		Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp
Ref	R^2	0.81	0.76	0.78	0.89	0.85	0.88	0.75	0.65	0.72	0.66	0.60	0.65	0.75	0.61	0.71
+$PM_{2.5est}$	RMSE	0.30	0.34	0.33	0.59	0.70	0.63	0.26	0.31	0.27	1.52	1.65	1.55	0.59	0.74	0.63
Ref	R^2	0.74	0.70	0.68	0.78	0.73	0.74	0.69	0.56	0.66	0.58	0.49	0.54	0.57	0.40	0.51
+AOD	RMSE	0.36	0.38	0.39	0.85	0.96	0.93	0.28	0.35	0.30	1.70	1.86	1.77	0.77	0.94	0.83
Ref	R^2	0.73	0.69	0.67	0.76	0.70	0.71	0.67	0.53	0.63	0.53	0.44	0.48	0.57	0.40	0.51
	RMSE	0.37	0.39	0.40	0.90	1.01	0.99	0.30	0.36	0.31	1.79	1.97	1.88	0.78	0.94	0.83

The *Ref* model incorporated meteorological predictors, CMAQ predictors, and spatiotemporally constructed features used for each constituent modeling. The specific predictors for each component model are detailed in Table 3.1. AOD is the imputed gap-free AOD. Rd, Sp, and Tp represent the random, spatial, and temporal CV, which is described in Section 6.3.4. The unit for RMSE is μg/m³.

The CMAQ-speciated outputs further enhanced model performance by providing species-specific chemical information. In addition, CMAQ offers relatively high spatial resolution (~12 km), enabling finer spatial representation of species concentrations compared to MERRA-2 and GEOS-Chem, which have coarser resolutions of 0.625° × 0.5° and 0.25° × 0.3125°, respectively, and were used in previous studies (e.g., Meng et al. 2018b). As shown in Table 6.7, incorporating CMAQ data improved validation R^2 by 0.01–0.11 across species. SHAP analysis further reveals that the CMAQ species corresponding to the target component typically rank as the second most important predictors after $PM_{2.5}$_est. These findings reinforce that model performance is closely tied to the physical and chemical alignment between predictors and target species. Interestingly, in the OC model, EC_CMAQ ranks higher than OC_CMAQ in SHAP importance, unlike the other species models where the corresponding CMAQ species is the dominant chemical predictor. This result likely reflects the co-emission of EC and OC from combustion sources, as well as CMAQ's limited ability to capture secondary organic aerosol (SOA) formation processes (Appel et al. 2017; Woody et al. 2016). As a result, EC_CMAQ may serve as a stronger proxy for combustion-related OC variability than OC_CMAQ itself. SHAP results also highlight the importance of meteorological predictors that influence relevant atmospheric processes—particularly secondary formation and vertical mixing and geographical variables. As groups, meteorological conditions accounted for 8–23% of the total predictive importance, with the significance of individual predictors within these groups varying considerably across models (Fig. 6.6). Temperature is particularly important for the SO_4^{2-} and OC components, ranking third and sixth in SHAP contributions, respectively. This likely reflects the strong temperature dependence of secondary aerosol processes: for SO_4^{2-}, temperature influences photochemical reaction rates and gas–particle partitioning of sulfur species, while for OC, it affects both the volatility of semi-volatile organic compounds and the rate of secondary organic aerosol formation (Carlton et al. 2009). PBLH emerges as the most important meteorological predictor for NO_3^- and EC, consistent with its role in controlling vertical mixing and dilution. For NO_3^-, a shallow boundary layer under cold, stagnant conditions enhances NO_3^-

accumulation via gas–particle partitioning (Guo et al. 2017; Seinfeld and Pandis 2016). For EC, which is primarily from combustion sources, boundary layer dynamics strongly regulate surface concentrations by modulating the extent of dispersion. VPD is especially important for DUST, reflecting its link to surface dryness and soil moisture availability, both of which influence the emission potential of windblown DUST particles (Prospero and Lamb 2003).

Table 6.7. Cross validation results of constituent models with various model structures.

Model	Metric	SO_4^{2-}			NO_3^-			EC			OC			DUST		
CV method		Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp
Ref	R^2	0.75	0.59	0.73	0.88	0.75	0.87	0.72	0.59	0.70	0.63	0.54	0.61	0.70	0.51	0.68
	RMSE	0.35	0.45	0.36	0.63	0.92	0.66	0.27	0.33	0.28	1.59	1.77	1.63	0.64	0.83	0.67
Ref	R^2	0.80	0.74	0.77	0.89	0.83	0.88	0.75	0.62	0.72	0.65	0.59	0.64	0.74	0.60	0.71
+ST	RMSE	0.31	0.36	0.34	0.59	0.76	0.65	0.26	0.32	0.27	1.53	1.67	1.57	0.60	0.75	0.64
Ref	R^2	0.79	0.70	0.76	0.89	0.84	0.88	0.73	0.64	0.71	0.65	0.59	0.63	0.72	0.55	0.69
+CMAQ	RMSE	0.32	0.38	0.34	0.61	0.73	0.64	0.27	0.31	0.28	1.55	1.67	1.58	0.62	0.80	0.66

The *Ref* model incorporated the 1-km $PM_{2.5}$ estimates and meteorological predictors used for each constituent modeling. ST represents the spatiotemporally constructed predictors and CMAQ represents the CMAQ species predictors. The other abbreviations are the same as Table 6.6.

Another key factor contributing to the strong performance of our component modeling is the use of a spatiotemporal deep-forest algorithm, which effectively captures both spatial and temporal variability inherent in the data. Compared to traditional tree-based algorithms such as random forest and CatBoost, deep forest is better suited for complex tasks involving limited sample sizes and higher spatiotemporal heterogeneity—as is the case for $PM_{2.5}$ chemical components, which are more variable and sparsely monitored than total $PM_{2.5}$ mass. In our comparison, deep forest outperformed random forest and CatBoost, with average cross-validation R^2 values higher by 0.06 and 0.04, respectively, across various validation schemes (Table 6.8). Its performance is also comparable to more complex ensemble approaches like the stacking model used in Di et al. (2019). To enhance the model's ability to generalize across space and time, we incorporated explicit spatiotemporal features—such as Haversine distances and day of year (DOY)—into the deep forest framework, ultimately developing a spatiotemporal deep-forest model (Section 6.3.3). This was particularly important given the sparse and infrequent nature of chemical speciation monitoring (sampling intervals ranging from 1 to 7 days). Sensitivity analyses confirmed the importance of these features: removing them led to a decline in spatial predictive performance, with spatial CV R^2 values reduced by 0.03–0.15 (Table 6.7). Furthermore, we found that individual tree-based models such as XGBoost were prone to generating spatially erratic or unrealistic component predictions when spatial features like latitude, longitude, or Haversine distances were included, a phenomenon also reported in prior $PM_{2.5}$ modeling studies (e.g., Yang et al. 2022). In contrast, the cascade structure of the deep forest model, which ensembles multiple layers of learners, helps alleviate this issue by introducing regularization and enabling hierarchical feature extraction. This design allows for more stable incorporation of spatial features and effectively reduces the risk of overfitting or abnormal predictions in our spatiotemporal deep-forest model.

Table 6.8. The cross-validation results of component models with machine-learning algorithms

Models	Metric	SO_4^{2-}			NO_3^-			EC			OC			DUST		
	CV	Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp	Rd	Sp	Tp
XGB	R^2	0.74	0.67	0.72	0.86	0.83	0.85	0.69	0.58	0.69	0.61	0.56	0.61	0.64	0.44	0.62
	RMSE	0.36	0.40	0.37	0.69	0.76	0.70	0.28	0.33	0.29	1.63	1.73	1.63	0.71	0.89	0.73
CAT	R^2	0.76	0.69	0.74	0.87	0.83	0.87	0.72	0.62	0.70	0.63	0.57	0.62	0.67	0.55	0.66
	RMSE	0.34	0.39	0.36	0.65	0.74	0.67	0.27	0.32	0.28	1.59	1.71	1.61	0.68	0.79	0.69
Two-	R^2	0.80	0.74	0.78	0.89	0.85	0.88	0.74	0.65	0.72	0.65	0.58	0.63	0.74	0.60	0.71
stage1	RMSE	0.31	0.36	0.33	0.61	0.70	0.63	0.26	0.31	0.27	1.55	1.71	1.58	0.61	0.76	0.64
Two-	R^2	0.78	0.74	0.77	0.86	0.85	0.85	0.69	0.70	0.70	0.66	0.66	0.64	0.67	0.62	0.66
stage2	RMSE	0.33	0.35	0.34	0.69	0.71	0.70	0.28	0.28	0.28	1.53	1.53	1.57	0.68	0.73	0.68

XGB refers to the XGBoost model, and CAT represents the CatBoost model. Two-stage 1 is a stacking model where the first stage includes Random Forest, XGBoost, and CatBoost, and the second stage uses Random Forest. Two-stage 2 is another stacking model, with a first stage composed of Gradient Boosting, Neural Network, and Random Forest, followed by a second stage using a Generalized Additive Model (GAM), as described in Di et al. (2019). All other abbreviations are consistent with those used in Table 6.6. All models in this table were trained on the same sample dataset used for our deep-forest models to ensure a fair comparison.

6.6 Limitation and Future Directions

Despite advancements in modeling $PM_{2.5}$ components, uncertainties and limitations remain. The much sparser ground-level $PM_{2.5}$ composition monitoring network, compared to that of total $PM_{2.5}$, provides less spatial representation, leading to relatively large uncertainties in unmonitored areas. Although independent validation against CASTNET measurements generally suggests reasonable accuracy, inter-network inconsistencies between CASTNET and CSN/IMPROVE may affect the interpretability of these results. Additionally, CASTNET provides measurements not for OC and EC, restricting the scope of external validation. Because our spatial resolution is set to 1 km and monitoring sites are scarce near high-emission areas in small-scale regions (e.g., roadsides), our component models are likely to underestimate extremely high concentrations. In addition, given that observational data for separating primary and secondary OC are limited, we did not disentangle them in our modeling process. Potential biases from CMAQ simulations may occur in areas with complex terrain or unique emission sources, and the model may struggle to fully capture extreme events or rapid $PM_{2.5}$ component changes. Additionally, the CMAQ species dataset used in this study is limited to the period from 2002 to 2019 and is available only at a weekly resolution. Future work should focus on extending component estimates to more recent years, potentially by incorporating alternative reanalysis products when CMAQ data are unavailable.

6.7 Summary

This section developed a two-stage spatiotemporal deep-forest framework to estimate daily, 1-km concentrations of five major $PM_{2.5}$ components across the western U.S. over 2002–2019. The models integrated gap-free AOD, CMAQ speciated outputs, meteorology, land cover, topography, wildfire activity, and spatiotemporal features. The framework achieved strong predictive performance (sample-based CV R^2 = 0.66-0.89), substantially outperforming alternative machine learning approaches including random forest and CatBoost, and producing estimates comparable to or better than previous studies while providing simultaneous high spatial (1-km) and temporal (daily) resolution. Validation demonstrated strong predictive performance, and SHAP analysis identified the estimated high-resolution $PM_{2.5}$ total mass concentration ($PM_{2.5}$_est) as the most influential predictor (accounting for 24-37% of model explanatory power) and chemically relevant CMAQ predictors as dominant influences. Independent validation against CASTNET measurements confirmed model generalizability, with yearly R^2 values of 0.81 for SO_4^{2-} and 0.78 for NO_3^-. The resulting dataset captures long-term trends, regional disparities, and episodic events such as wildfires, providing a valuable resource for exposure assessment and policy evaluation.

7. Spatiotemporal Patterns of $PM_{2.5}$ Components over Western US

In this section, we analyze spatiotemporal patterns and population exposure for daily, 1-km $PM_{2.5}$ total mass and five chemically resolved components across the western United States from 2002 to 2019. We summarize multi-year and seasonal spatial distributions, quantify long-term trends using deseasonalized monthly anomalies, and assess exposure gaps between urban and rural populations using population-weighted metrics. We also highlight short-term dynamics via a wildfire case study (the 2018 Camp Fire) to elucidate day-to-day variability and species composition during extreme events. Results are presented as long-term means and pixel-level trends (Section 7.2), and regional exposure summaries (Section 7.3) and day-to-day variations (Section 7.4).

7.1 Statistical analysis methods

We conducted analyses at both the pixel level and the region-aggregated level. For the latter, we aggregated grid-cell estimates by predefined regions to construct time series and assess region-specific trends. The western United States was divided into two subregions: (1) Northwestern U.S. — Washington, Oregon, Idaho, Montana, and Wyoming; (2) Southwestern U.S. — California, Nevada, Utah, Colorado, Arizona, and New Mexico.

We also stratified the study area into urban and rural categories. Urban areas were defined according to the 2010 U.S. Census Bureau urbanized area boundaries (https://www.census.gov/cgi-bin/geo/shapefiles/index.php), while rural areas encompassed all locations outside these boundaries.

7.1.1 Seasonal and multi-year summaries

We computed multi-year (2002–2019), annual, and seasonal means on the 1-km grid. Seasons are defined as winter (December, January, February), spring (March, April, May), summer (June, July, August), and autumn (September, October, November).

7.1.2 Long-term trend estimation

To assess temporal changes at the local scale, we conducted pixel-level long-term trend analyses using monthly anomaly time series. Monthly anomalies were computed by subtracting the multi-year mean for each calendar month from the corresponding monthly mean concentration, thereby removing the influence of the seasonal cycle. This approach ensures that the estimated slopes reflect genuine long-term changes rather than seasonal fluctuations. We then applied ordinary least squares to the anomaly series to estimate linear trends and their statistical significance ($p < 0.05$), following Weatherhead et al. (1998). Trends are reported as $\mu g/m^3$ per year, and statistical significance was assessed at the $p < 0.05$ level. The resulting spatial maps of trends provide insight into both the magnitude and direction of change at the grid-cell level, allowing for direct comparison across pollutants and regions.

7.1.3 Population-weighted exposures

We adopted the conventional population-weighted mean concentration approach to quantify exposure hotspots. Specifically, for each grid cell, we multiplied the pollutant concentration by the concurrent population from the 1-km LandScan dataset. The total weighted sum across a given region was then divided by the total population of that region to obtain the population-weighted mean concentration.

$$\bar{C}_{pop} = \left(\sum_{i=1}^{n} (C_i \times P_i) \right) / \left(\sum_{i=1}^{n} P_i \right)$$

Where:

$\bar{C}_{pop}$ = population-weighted mean concentration for the region;

C_i = pollutant concentration in grid cell i;

P_i = population in grid cell i;

n = number of grid cells in the region.

7.2 Spatial distribution and long-term trends

Figure 7.1 presents the spatial distributions of multi-year average concentrations of the five derived components, alongside observed values at corresponding grid cells, demonstrating good agreement between model estimates and measurements. The multi-year estimated concentrations are 0.78, 0.44, 0.22, 1.31, and 1.33 μg/m³ for SO_4^{2-}, NO_3^-, EC, OC, and DUST, respectively, with substantial spatial variability. DUST is predominantly concentrated along the southern California and western Arizona boundary, an arid and semi-arid region frequently affected by dust events (Aryal and Evans 2022). The other four components— SO_4^{2-}, NO_3^-, EC, and OC—show high concentrations in the urbanized southern California region, centered in the Los Angeles–Long Beach–Anaheim metropolitan area, a densely populated urban agglomeration with intensive human activity. NO_3^-, OC, and EC also show elevated levels in California's Central Valley, particularly in the San Joaquin Valley, a region notorious for consistently exceeding federal particulate pollution standards due to a combination of complex anthropogenic emissions, frequent wildfires, and unfavorable topographical conditions that limit pollutant dispersion (Casey et al. 2024; P et al. 2022). Additionally, high EC and OC concentrations are observed in parts of the northwestern U.S., including areas of Washington and Oregon.

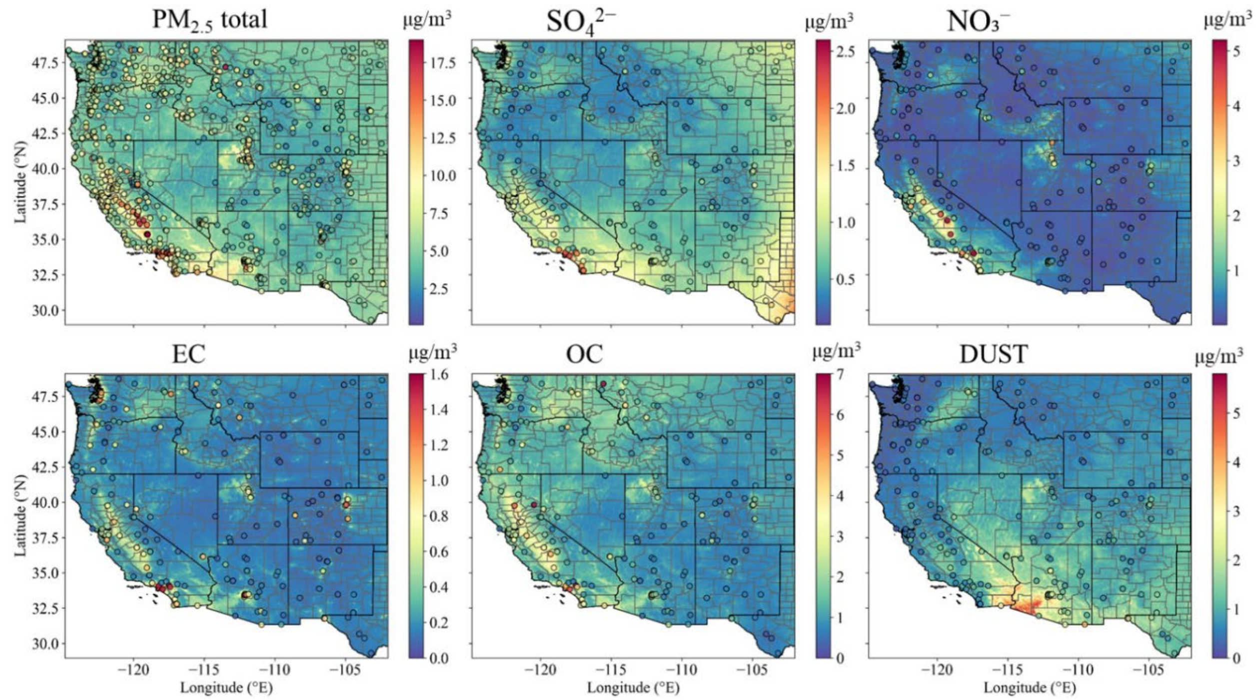

Figure 7.1. Spatial distribution of multi-year mean $PM_{2.5}$ and its components for 2002-2019 over western United States, shown alongside monitoring observations.

Seasonal patterns in total $PM_{2.5}$ concentrations across the study region revealed distinct peaks in summer and autumn, with average levels reaching 6.34 μg/m³ and 4.58 μg/m³, respectively. These seasonal peaks were primarily driven by elevated concentrations of OC (1.99 μg/m³ in summer and 1.40 μg/m³ in autumn), DUST (1.30 μg/m³ and 0.85 μg/m³), and SO_4^{2-} (0.83 μg/m³ and 0.61 μg/m³) (Figure 7.2). In contrast, winter exhibited the lowest average $PM_{2.5}$ concentration at 3.36 μg/m³, with OC (0.77 μg/m³) and NO_3^- (0.70 μg/m³) contributing to the total mass. During winter, EC is a significant contributor in regions such as Southern California, urban areas of western Washington, the Central Valley, and central Arizona, where transportation emissions and episodic biomass burning are dominant sources (Faraz Enayati Ahangar 2021; Sorooshian et al. 2011).

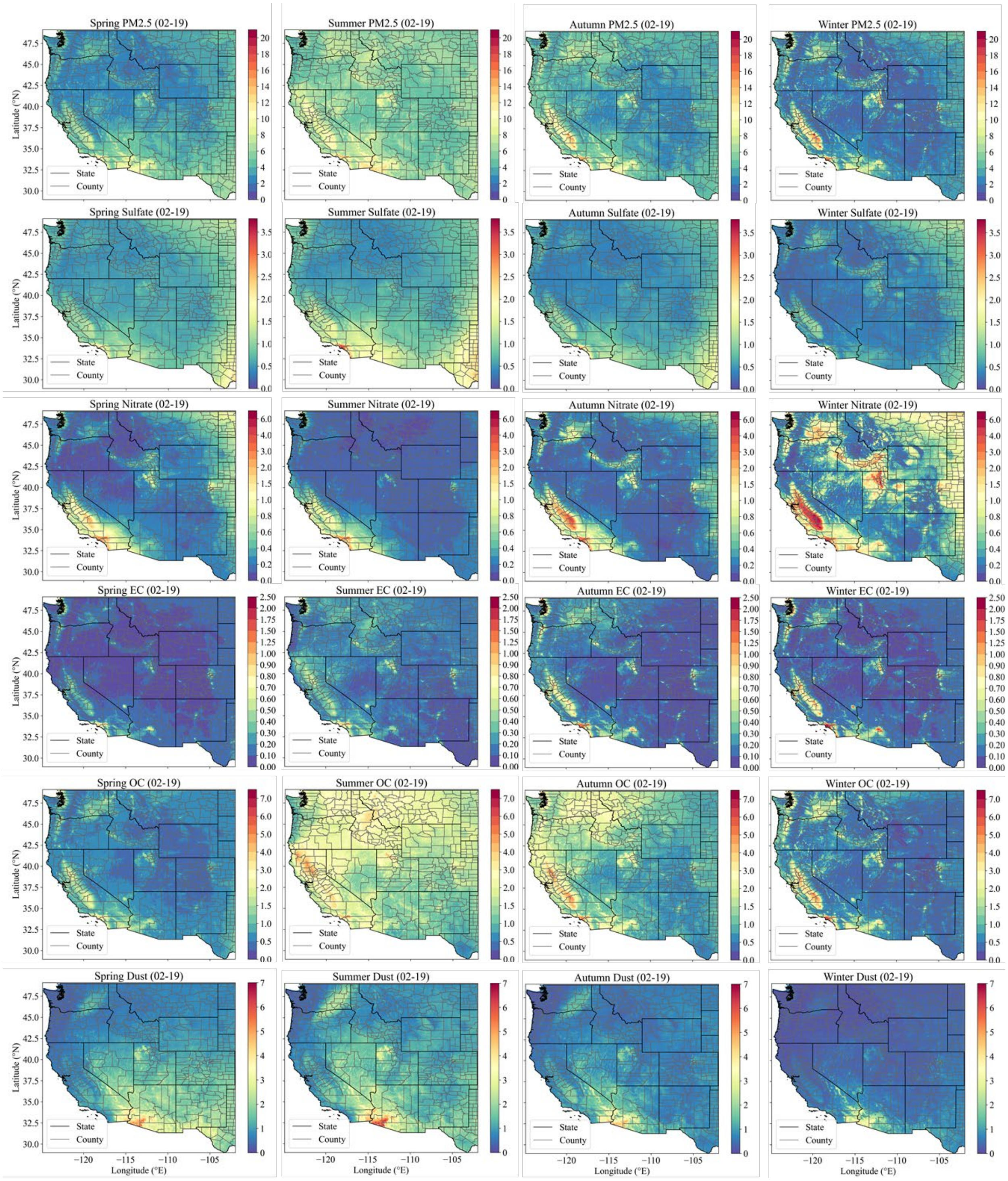

Figure 7.2. Seasonal average total and compositional $PM_{2.5}$ concentrations for 2002-2019.

As shown in the pixel-level trend figure, from 2002 to 2019, steady declines were observed in the concentrations of total $PM_{2.5}$ and the five major components across the western U.S., with significant spatial variations (Fig. 7.3). The statistically significant decline in total $PM_{2.5}$ (slope <-0.05 μg/m³/year, $p < 0.05$) is most pronounced in the southwestern U.S. and northwestern coastal areas, particularly in southern California's urban regions and Central Valley, which exhibited steeper negative trends (slope <-0.20 μg/m³/year, $p < 0.05$). Similarly, EC and OC had comparable declining patterns, though with smaller magnitudes (slope < -0.003 μg/m³/year for EC and slope <-0.02 μg/m³/year for OC, $p < 0.05$). SO_4^{2-} and NO_3^- also exhibited widespread declines across the study region, with steeper trends in southern California's urban areas (slope <- 0.045 μg/m³/year for SO_4^{2-} and slope < -0.15 μg/m³/year for NO_3^-, $p < 0.05$). These substantial reductions over densely populated areas can be largely attributed to significant decreases in anthropogenic emissions, driven by diesel emission controls, vehicle NOx standards, and industrial regulations under major air quality policies such as the Clean Air Act amendments and California Air Resources Board (CARB) PM Reduction Programs (Kotchenruther 2020; Mailloux et al. 2022). DUST also shows a widespread declining trend, except in areas with higher levels, such as the southern California and western Arizona boundary, which warrant greater attention and targeted mitigation measures.

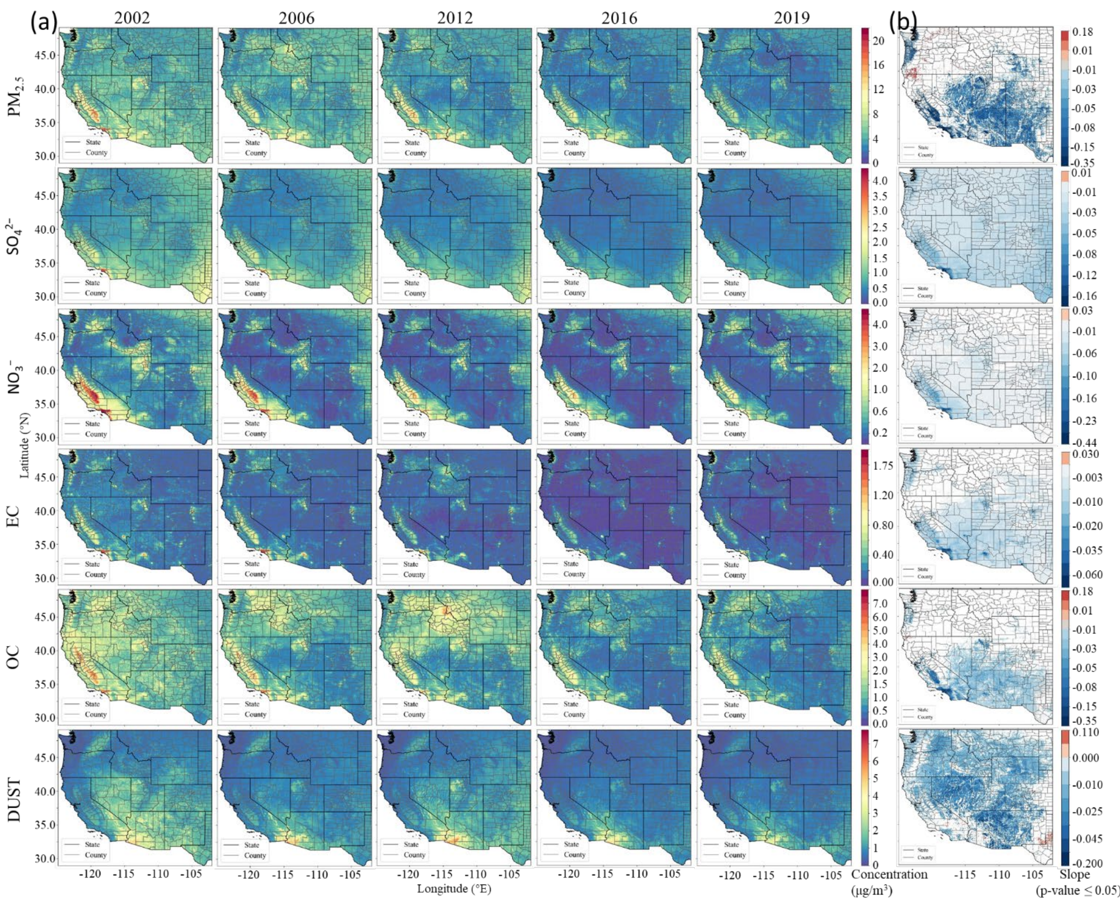

Figure 7.3. (a) Spatial distribution of annual mean total $PM_{2.5}$ mass and its components, and (b) pixel-specific long-term trends of monthly anomalies for the period 2002-2019 across the western United States.

7.3 Population exposure variations

Figure 7.4 illustrates regional variations in population-weighted exposure to $PM_{2.5}$ and its species from 2002 to 2019, revealing substantial spatial heterogeneity. On average, urban populations experienced ~1.5 times higher exposure than rural populations for SO_4^{2-} (1.24 vs. 0.84 μg/m³) and OC (2.97 vs. 2.01 μg/m³), ~2 times higher for NO_3^- (1.83 vs. 0.93 μg/m³) and EC (0.87 vs 0.41 μg/m³), and comparable levels for DUST (1.09 vs. 1.08 μg/m³) (Figure 7.4(a)). Monthly population-weighted mean concentrations based on CMAQ-simulated species also showed that EC and NO_3^- levels were substantially higher in urban areas compared to rural areas, ranging from 0.73 to 1.84 μg/m³ versus 0.31 to 1.29 μg/m³, respectively. In the northwestern U.S., the average monthly population-weighted mean $PM_{2.5}$ slightly decreased from 8.28 μg/m³ in 2002–2004 to 7.23 μg/m³ in 2017–2019 (Figure 7.4(a)). The month with the highest concentration of total $PM_{2.5}$ in this region varies annually, occurring primarily in winter but sometimes in August. These peaks in total $PM_{2.5}$ are largely driven by the OC component. In contrast, the southwestern U.S. experienced a significant decrease in $PM_{2.5}$ concentrations, with average monthly population-weighted means declining from 12.32 μg/m³ in 2002-2004 to 9.03 μg/m³ in 2017-2019, primarily due to reductions in OC and NO_3^- levels, partly driven by tailpipe emission regulations.

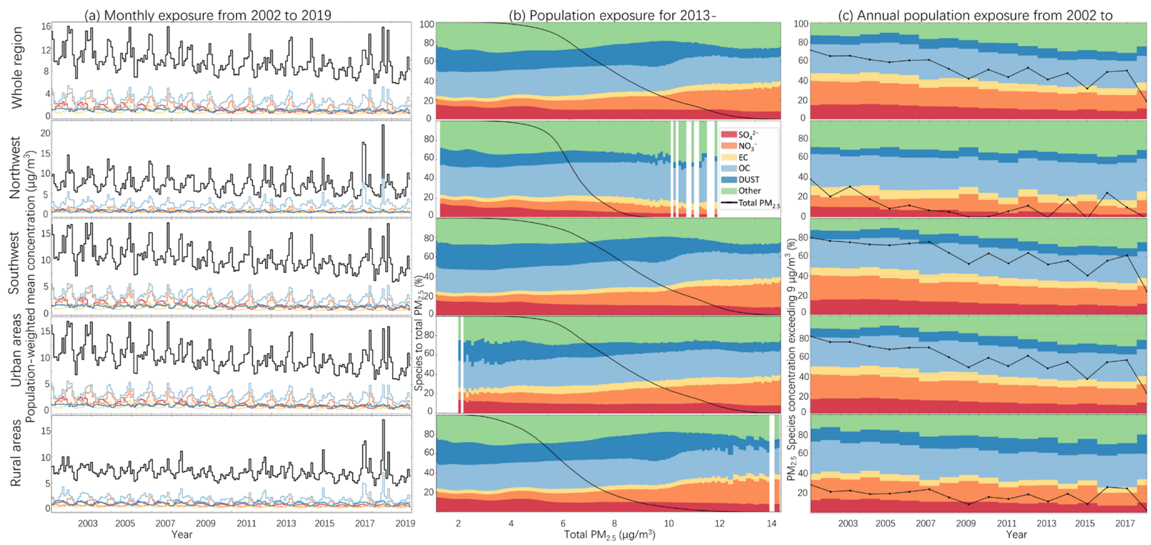

Figure 7.4. Regional variation in population-weighted exposure to $PM_{2.5}$ and its five components from 2002 to 2019 across various time scales. $PM_{2.5}$ total (black), SO_4^{2-} (red), NO_3^- (orange), EC (yellow), OC (light blue), DUST (blue), and 'other' (green, defined as the difference between total $PM_{2.5}$ and the sum of the five target species) are represented by color: (a) Monthly population-weighted mean concentrations from 2002 to 2019; (b) Proportions of each species versus total $PM_{2.5}$ for 2013–2019, with the black line showing cumulative proportions of the population exposed to each $PM_{2.5}$ level (this time frame was selected because the last U.S. national air quality standard for annual $PM_{2.5}$ was set at 12 μg/m³ in 2012); (c) Annual population-weighted concentrations of species exceeding the 2024 U.S. national annual $PM_{2.5}$ mass concentration limit (9 μg/m³), with the black line indicating the percentage of the population exceeding the annual limit.

Figure 7.4(b) concentrations within the context of the 2012 regulatory framework. Similar to patterns observed across North America (Donkelaar et al. 2019), the relative contribution of the DUST component decreased significantly as $PM_{2.5}$ levels increase across the four regions (Fig.

7.4(b)). In the southwest, NO_3^- has a minor contribution at low $PM_{2.5}$ levels but becomes a major contributor at higher levels, which is in line with previous studies (Donkelaar et al. 2019). In the northwest, OC was the largest contributor among the five target species, with its relative contribution increasing as total $PM_{2.5}$ levels rose, though the population exposed to elevated levels was very small. Figure 7.4(c) illustrates the annual changes in population exposure to $PM_{2.5}$ levels exceeding the latest 2024 U.S. National Ambient Air Quality Standard (annual average concentration of 9 μg/m³). The northwestern is approaching compliance with the stricter standard, with the proportion of the population exposed to non-compliant $PM_{2.5}$ levels remaining around 1% in many years since 2010. However, the southwestern region remains far from meeting the new standard, with a large proportion of the population (~24% in 2019) still exposed to above-standard levels, despite significant improvements over time. The impact of wildfires is evident across the study region, with seasonal OC enhancements varying annually. This effect was particularly pronounced in 2017 and 2018, when substantially elevated OC and total $PM_{2.5}$ levels were observed, highlighting the influence of wildfire events during these years.

7.4 Day-to-day variability

Our estimates effectively captured daily variations in $PM_{2.5}$ component concentrations, providing valuable insights into atmospheric pollution and chemical composition dynamics, supporting targeted environmental management, and aiding in the mitigation of health impacts from short-term exposure. Using the Camp Fire—a severe wildfire in California from November 8–25, 2018—as an example, our estimates show that population-weighted mean $PM_{2.5}$ concentrations began rising on November 10, 2018, rapidly peaking at 77.55 μg/m³ the same day. Levels remained elevated until reaching a maximum of 110.41 μg/m³ on November 16, 2018, before subsequently decreasing to a background level (5.47 μg/m³) by November 22, 2018, following rainfall on November 21, 2018 (Fig. 7.5(a)). These trends align well with observed data (Fig. 7.5(c)), highlighting the accuracy of our daily-scale estimates.

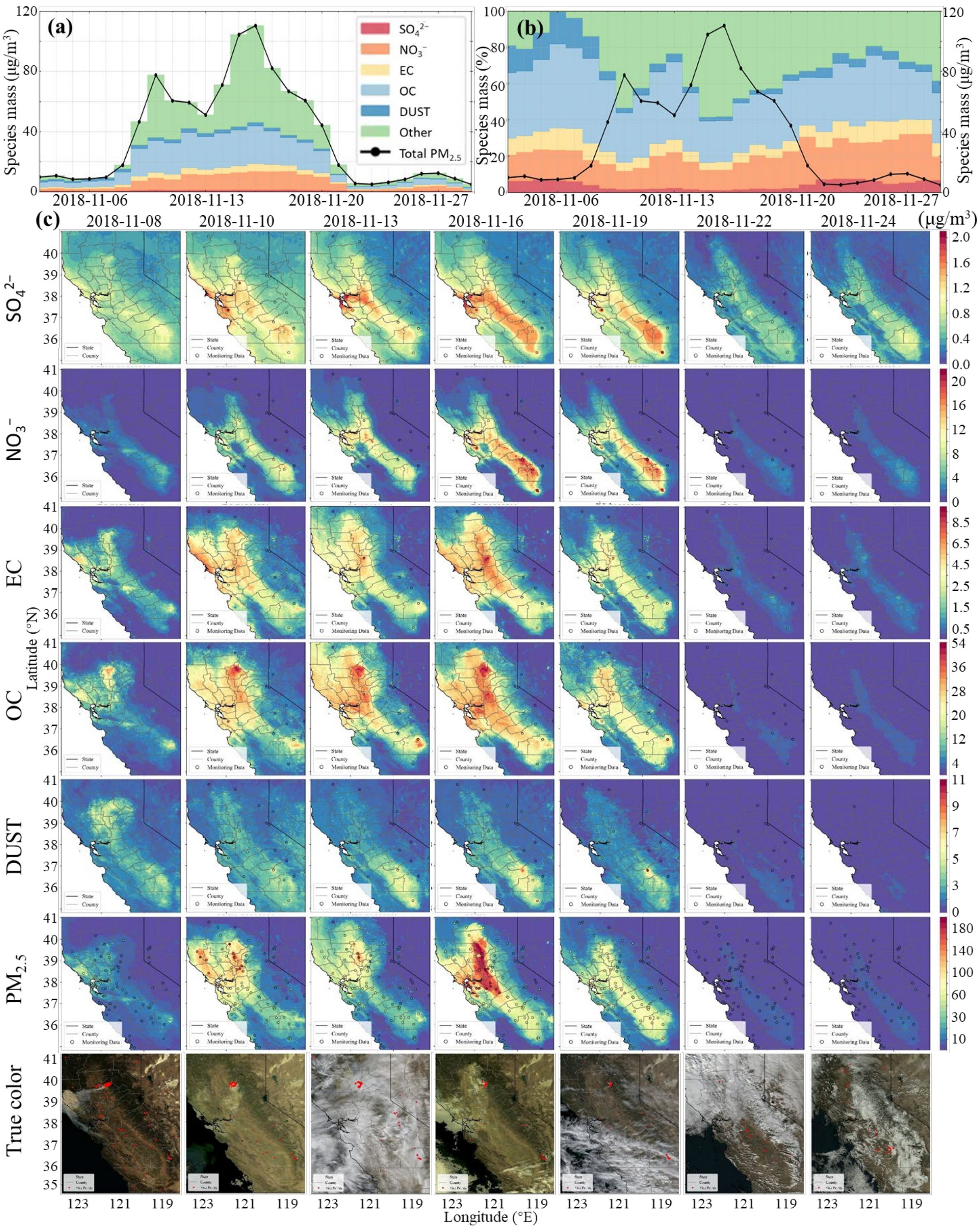

Figure 7.5. Population-weighted daily mean estimated (a) total $PM_{2.5}$ and component concentrations and (b) proportions in California and (c) the corresponding concentration spatial distributions and true color Aqua images with active fire as red dots during the Camp Fire (November 8-24, 2018).

The $PM_{2.5}$ component estimates indicate that concentrations of OC and EC, direct emissions from biomass combustion, substantially increased, particularly over the Central Valley, closer to the fire source (Fig. 7.5(c) and Fig. 7.6). In contrast, NO_3^- increases were primarily observed in the San Joaquin Valley, while SO_4^{2-} and DUST showed only moderate increases. Overall, exposure to OC and NO_3^- significantly increased across the Central Valley during the Camp Fire, with levels averaging ~6 times higher (18.9 and 9.2 μg/m³, respectively) compared to pre-fire levels (3.8 and 1.5 μg/m³). Increases in EC exposure were also notable. Figure 7.5(b) illustrates that OC, the dominant component, does not scale proportionally with total $PM_{2.5}$ during the Camp Fire, suggesting that other $PM_{2.5}$ components, such as metals (e.g., lead and zinc), potassium ion (K^+), and ammonium (CARB 2021), may have significantly increased. These other components, which spiked due to the burning of buildings and infrastructure, also should be a focus for environmental management due to their potential health and environmental impacts (Potter et al. 2021; Sicheng Li 2023).

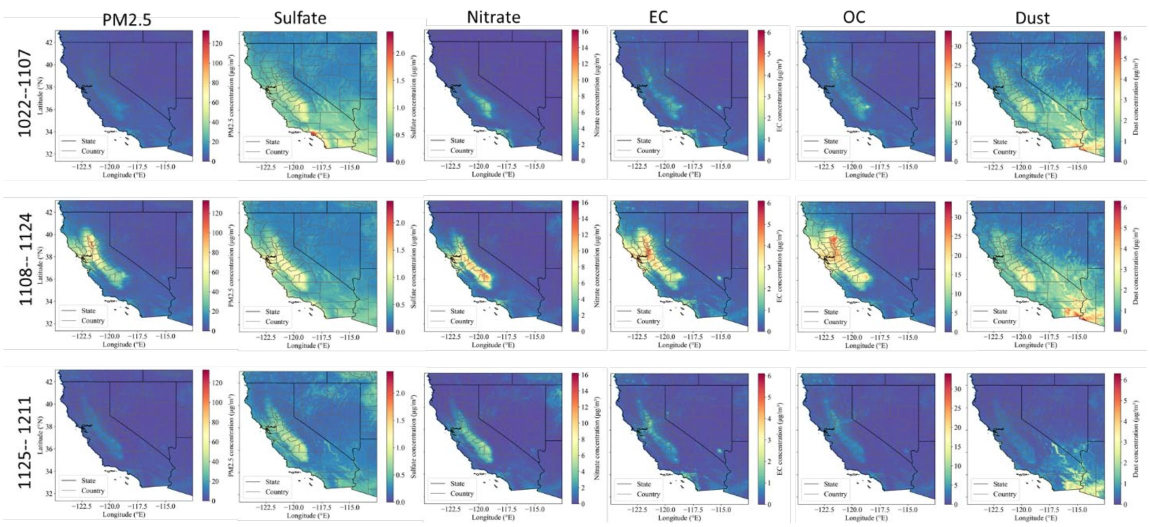

Figure 7.6. Spatial distributions of the five $PM_{2.5}$ component concentrations before (upper panel), during (middle panel), and after (lower panel) Camp fire.

7.5 Discussion

Figure 7.7 displays component concentrations from 2002 to 2019 across the western U.S. based on monitoring data, revealing variation patterns consistent with those estimated by our component models at the monitored grid cells (Fig. 7.3). Figure 7.8 shows the annual concentrations of CMAQ-simulated species and their long-term trends, derived using the same methodology as in Figure 7.3. Overall, the 12-km CMAQ outputs capture broad-scale trend patterns similar to our high-resolution estimates, indicating widespread declines across the study region, although differences in magnitude and spatial detail are evident.

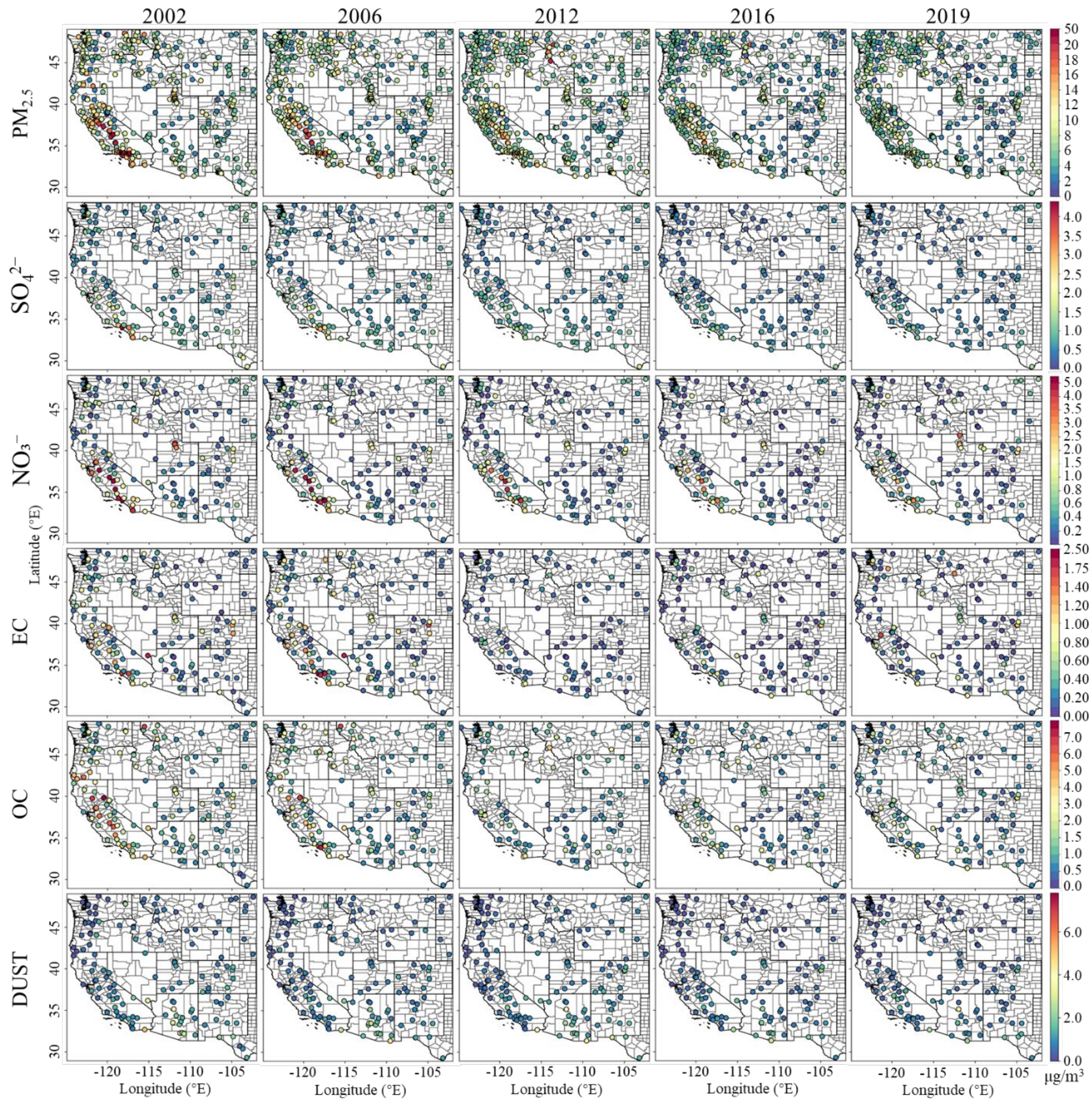

Figure 7.7. Spatial distribution of annual mean total $PM_{2.5}$ mass and its components for the period 2002-2019 across the western United States based on monitoring data.

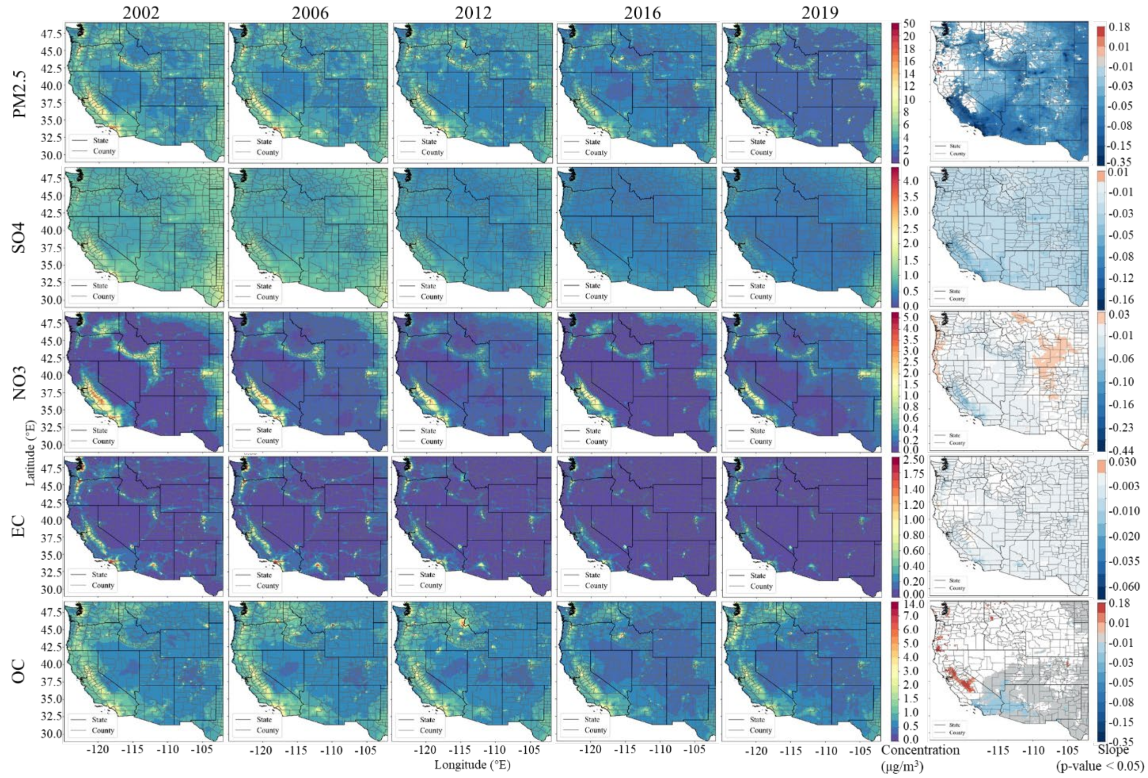

Figure 7.8. (a) Spatial distribution of annual mean total $PM_{2.5}$ mass and its components, and (b) pixel-specific long-term trends of monthly anomalies for the period 2002-2019 across the western United States based on CMAQ-speciated simulation data.

Although the CMAQ-simulated species exhibit broadly similar spatiotemporal patterns to our high-resolution estimates, notable differences in magnitude and spatial detail are evident (Fig. 7.3 vs. 7.8). Specifically, CMAQ tends to underestimate concentrations, particularly in high-concentration regions identified by our model. In addition, CMAQ shows increasing trends in NO_3^- over the northwestern part of the study region and in OC over central California—patterns not observed in either ground-based measurements (Fig. 7.7) or our estimates. In contrast, our model predictions align more closely with observed trends at monitoring sites, as demonstrated in Figs. 7.1, 7.3, and 7.7. These discrepancies are likely due to the relatively lower estimation accuracy of CMAQ. Validation against CSN and IMPROVE observations (Fig. 7.9) confirms CMAQ's lower accuracy, with R^2 values of 0.39, 0.55, 0.40, and 0.16 and RMSEs of 0.53, 1.45, 0.41, and 2.36 μg/m³ for SO_4^{2-}, NO_3^-, EC, OC, respectively.

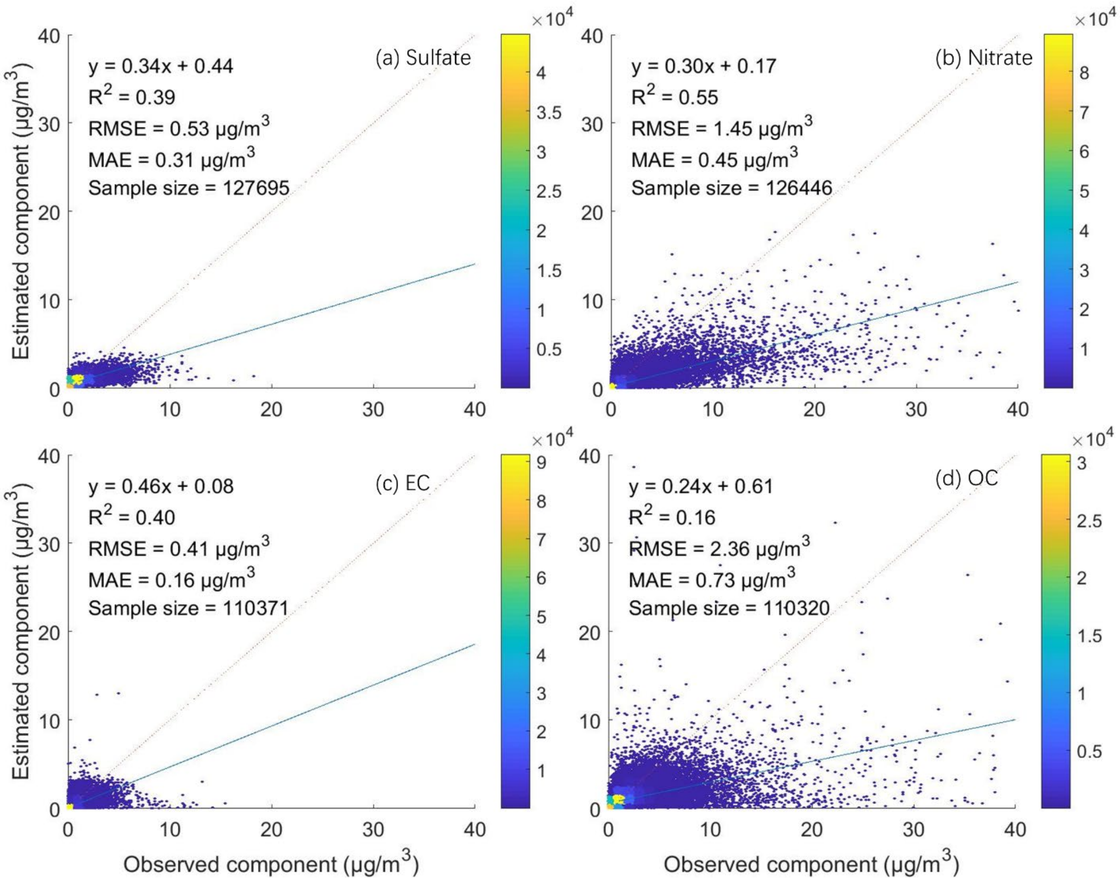

Figure 7.9. Scatterplots comparing observed and CMAQ-speciated weekly $PM_{2.5}$ component concentrations from 2002 to 2019.

We further observed that CMAQ-based population-weighted concentrations tend to underestimate exposure levels, with underestimation patterns varying by species and between urban and rural settings (Figure 7.3 vs. 7.10). This underestimation is partly attributed to CMAQ's coarse spatial resolution. Previous studies have shown that using coarse-resolution $PM_{2.5}$ data can systematically underestimate exposure (He et al. 2021). SO_4^{2-} and NO_3^- components show greater underestimation in urban areas, likely due to limitations in resolving high-density emission sources and the more intense secondary formation processes driven by elevated precursor levels in urban environments. In contrast, EC underestimation is more pronounced in rural areas, likely reflecting missing or underestimated rural combustion sources such as residential biomass burning and wildfires. OC underestimation was relatively balanced between urban and rural areas, consistent with its mixed original from both primary emissions and secondary organic aerosol formation.

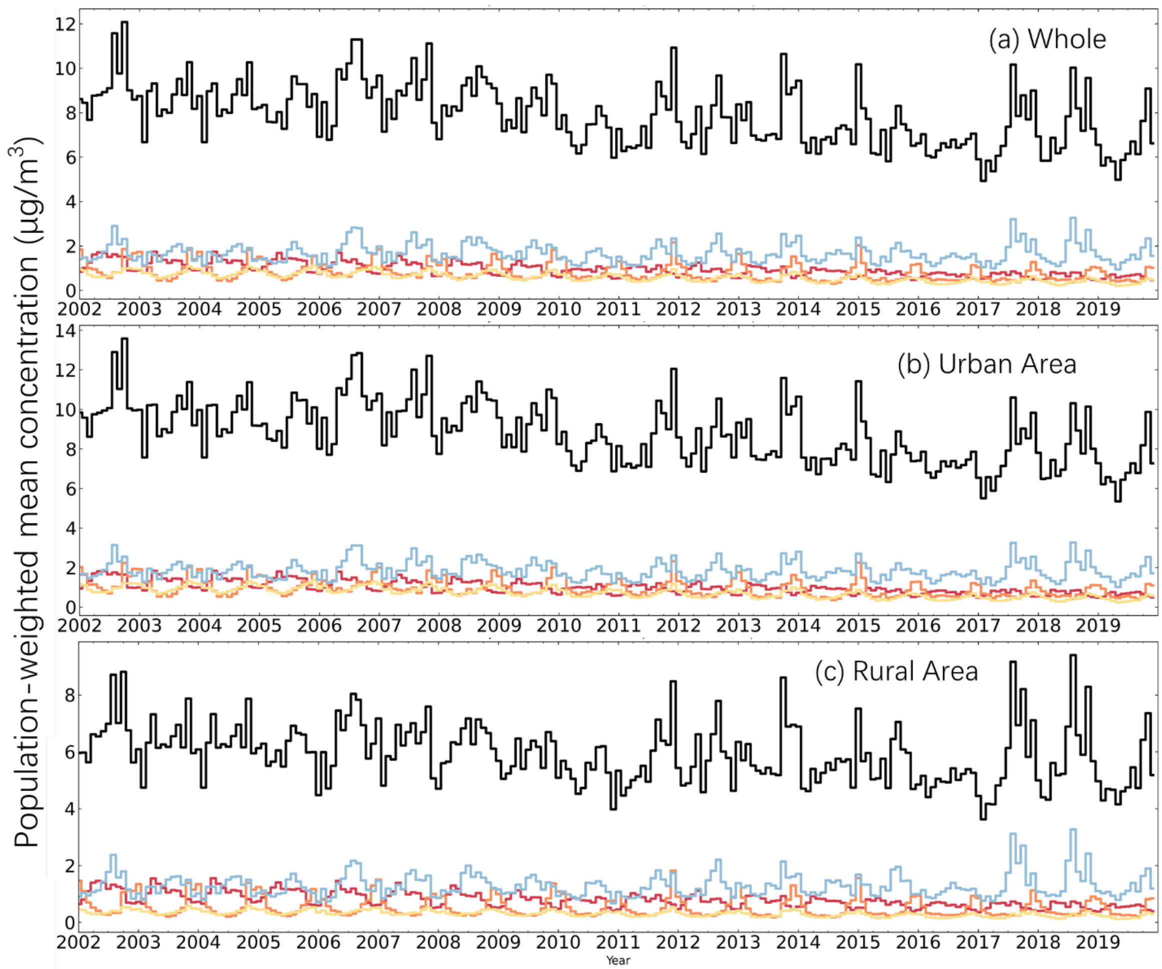

Figure 7.10. Monthly population-weighted mean concentrations from 2002 to 2019 based on CMAQ-speciated data.

7.6 Summary

This section observed the spatiotemporal patterns of $PM_{2.5}$ components using the high-resolution estimates generated in Section 6. Spatiotemporal analysis of the modeling results revealed that urban populations were exposed to 1.5–2 times higher concentrations of SO_4^{2-}, NO_3^-, EC, and OC compared to rural populations, with similar exposure levels for DUST. These trends exhibited consistent declines over time, accompanied by notable interannual variability, seasonal cycles, and regional differences. The influence of wildfires was particularly significant, with daily OC and NO_3^- concentrations increasing by approximately sixfold during the Camp Fire relative to non-fire periods. The estimated spatial patterns and long-term trends showed strong agreement with ground-based observations and more realistic spatiotemporal variability than CMAQ-speciated simulations. The $PM_{2.5}$ component modeling approach presented here demonstrates strong potential for application in other regions with ground-level $PM_{2.5}$ chemical speciation monitoring networks. These results provide the foundation for subsequent analyses of California-specific patterns (Section 8) and decomposition of meteorological, wildfire, and anthropogenic drivers (Section 9).

8. Exploring $PM_{2.5}$ Component Exposure Hotspots in California and Their Trends

8.1 Introduction

Understanding the spatial and temporal behavior of $PM_{2.5}$ and its major chemical components is critical for air quality management in California. The state's diverse emission sources—from dense urban traffic corridors to agricultural ammonia in the Central Valley—combine with complex topography and meteorology to produce pronounced regional variability in both total $PM_{2.5}$ mass and its chemical composition. These variations have important implications for public health, particularly for vulnerable populations such as children, the elderly, and individuals with pre-existing respiratory or cardiovascular conditions.

This section first extends the $PM_{2.5}$ component modeling framework introduced in Section 6 to 2000, 2001, and 2020 by substituting CMAQ predictors with MERRA-2 and MERRA-2 GMI reanalysis variables, ensuring coverage for the entire 2000–2020 period. With this extended dataset, we examine statewide spatiotemporal patterns for total $PM_{2.5}$ and its five components. We identify long-term exposure hotspots (areas with persistently high multi-year mean concentrations) and quantify linear trends over two decades. Finally, we integrate these results with Census tract boundaries and CalEnviroScreen data to compare concentration distributions between high- and low-vulnerability communities, highlighting environmental disparity considerations. The remainder of this section is organized as follows: Section 8.2 describes the modeling for 2000, 2001, and 2020 and the integrated performance assessment for 2000–2020. Section 8.3 describes the statistical analysis methods for exploring the spatiotemporal trends in $PM_{2.5}$ and its component exposure. Section 8.4 presents spatial hotspot maps and linear trend analyses for total $PM_{2.5}$ and each component. Section 8.5 examines population exposure patterns, contrasting high- and low-vulnerability communities.

8.2 Modeling $PM_{2.5}$ Component Estimates over California for 2000, 2001, and 2020

8.2.1 Data and modeling

To generate daily, 1-km gap-free estimates of SO_4^{2-}, NO_3^-, EC, OC, and DUST for 2000, 2001, and 2020—years without available CMAQ-speciated simulations—we applied the deep-forest modeling framework described in Section 6, with a key modification: CMAQ predictors were replaced by species mass concentrations from the MERRA-2 and MERRA-2 GMI reanalysis products. The selected reanalysis predictors were:

- **BCSMASS** – Black carbon surface mass concentration
- **DMSSMASS** – Dry matter surface mass concentration from biomass burning aerosol
- **DUSMASS25** – Dust surface mass concentration for particles with diameter < 2.5 μm
- **OCSMASS** – Organic carbon surface mass concentration
- **SO2SMASS** – Sulfur dioxide surface mass concentration

- **SO4SMASS** – Sulfate aerosol surface mass concentration
- **SSSMASS25** – Sea salt surface mass concentration for particles with diameter < 2.5 μm

These reanalysis fields were bilinearly resampled to the 1-km grid. Following resampling, we performed component-specific screening to decide which reanalysis variables to keep in each model: for every component, we computed correlations with observations and removed predictors that were weakly associated with the target or highly collinear with other candidates. The ultimately retained MERRA-2 / MERRA-2 GMI variables therefore differ by component; Table 8.1 lists the retained fields for each component together with their correlation coefficients against the corresponding observations. All other predictors and modeling settings mirrored those in the main framework: AOD-derived total $PM_{2.5}$ (the primary predictor), meteorological and land-use/population covariates, and explicit spatiotemporal features. We evaluated performance with 10-fold random (sample-based), spatial (site-based), and temporal (day-based) cross-validation.

Table 8.1. The final set of reanalysis predictors used for each component modeling and their correlations with ground-level observations (obs) for 2000, 2001, and 2020.

	SO_4^{2-} obs		NO_3^- obs		EC obs
SO2SMASS	0.214	SO2SMASS	0.269	SO2SMASS	0.375
SO4SMASS	0.397	SO4SMASS	0.172	SO4SMASS	0.221
DMSSMASS	0.080	SSSMASS25	0.040	SSSMASS25	0.021
SSSMASS25	0.061	DUSMASS25	-0.024	DMSSMASS	-0.013
DUSMASS25	0.060	DMSSMASS	0.008	DUSMASS25	-0.019
	OC obs		DUST obs		
BCSMASS	0.469	DUSMASS25	0.352		
SO4SMASS	0.226	BCSMASS	0.145		
DMSSMASS	-0.026	SSSMASS25	-0.088		
DUSMASS25	-0.011	SO4SMASS	0.081		
SSSMASS25	-0.001	DMSSMASS	-0.065		

8.2.2 Model performance

Model performance for the three modeled years was evaluated using sample-based (random), spatial, and temporal 10-fold cross-validation (CV). The CV R^2 values for 2000, 2001, and 2020 (Table 8.2) ranged from 0.70–0.84 for the random CV, 0.65–0.74 for the spatial CV, and 0.53–0.81 for the temporal CV, depending on species. NO_3^- consistently achieved the highest R^2 across all CV types, while EC and DUST showed relatively lower spatial predictive performance. These results indicate that, even without CMAQ-speciated inputs, the models retained strong predictive skill across species, comparable to that achieved for 2002–2019 using CMAQ data (Section 6.4.2).

Table 8.2. CV R^2 of $PM_{2.5}$ component models across California for 2000, 2001, and 2020.

CV type	SO_4^{2-}	NO_3^-	EC	OC	DUST
Random	0.71	0.84	0.70	0.77	0.74
Spatial	0.66	0.74	0.68	0.73	0.65
Temporal	0.66	0.81	0.53	0.72	0.64

We then combined the three-year CV results with the 2002–2019 evaluation results summarized in Section 6. As shown in Table 8.3, the statewide day-level 2000–2020 CV ranges span 0.66–0.89 (random), 0.60–0.85 (spatial), and 0.53–0.88 (temporal) across SO_4^{2-}, NO_3^-, EC, OC, and DUST. As in Section 6, aggregation further improves accuracy at monthly and annual scales. Taken together, the integrated 2000–2020 validation supports the use of these California estimates for subsequent spatiotemporal trend analyses and population-weighted exposure assessments.

Table 8.3. Validation results of component modeling across CA for 2000-2020

Species	10CV method	R^2	RMSE (μg/m³)
DUST	sample-based	0.6643	0.5540
DUST	spatial	0.5281	0.6707
DUST	temporal	0.6051	0.6002
EC	sample-based	0.7499	0.3638
EC	spatial	0.6595	0.4260
EC	temporal	0.7415	0.3699
NO_3^-	sample-based	0.8859	1.0394
NO_3^-	spatial	0.8386	1.2503
NO_3^-	temporal	0.8688	1.1141
OC	sample-based	0.7158	1.8352
OC	spatial	0.6387	2.0788
OC	temporal	0.6989	1.8890
SO_4^{2-}	sample-based	0.8086	0.4452
SO_4^{2-}	spatial	0.7491	0.5112
SO_4^{2-}	temporal	0.7816	0.4755

8.3 Statistical analysis methods

To investigate long-term spatial patterns and temporal changes in $PM_{2.5}$ and its major components across California, we performed three complementary statistical analyses: hotspot identification, trend estimation, and population exposure assessment, including a focus on vulnerable communities. All analyses were conducted using the daily, 1-km resolution dataset covering 2000–2020, which integrates the modeled estimates for 2000, 2001, and 2020 (Section 8.2) with the previously developed 2002–2019 fields from Section 6. Beyond the pixel-level analysis, we also performed a regional analysis distinguishing between urban and rural areas, with urban areas defined according to the 2010 U.S. Census Bureau urbanized area boundaries (https://www.census.gov/cgi-bin/geo/shapefiles/index.php). This analysis further examined five representative megacity regions—San Francisco Bay (SF Bay), Sacramento, Fresno, Los Angeles (LA metro), and San Diego (SD) megaregions—as shown in Fig. 8.1.

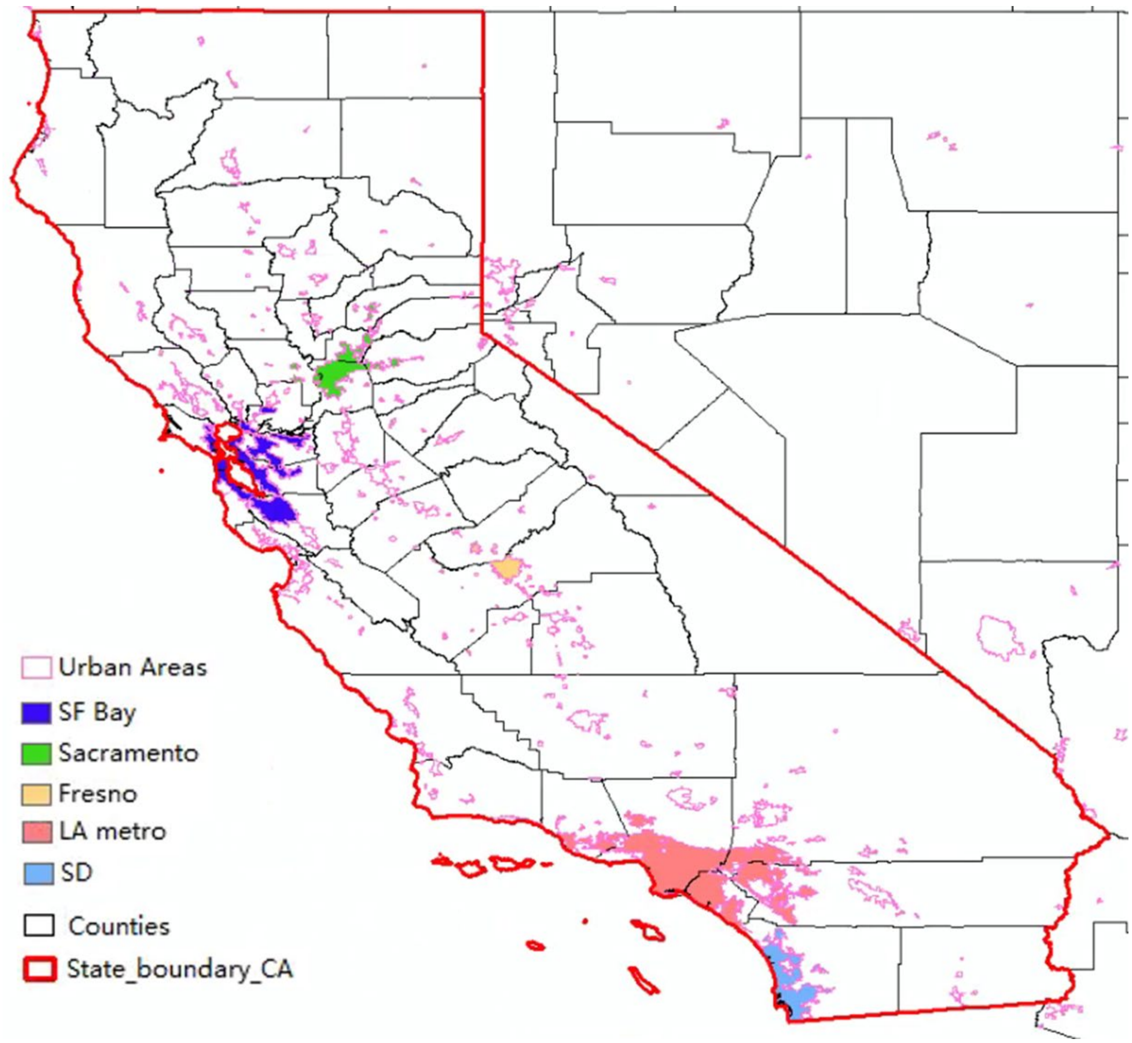

Figure 8.1. Spatial distributions of urban areas with the five major metropolitan areas highlighted.

8.3.1 Hotspot identification

We first calculated the spatial distribution of the multi-year mean daily concentrations for total $PM_{2.5}$ and each of the five components over the 21-year period. The multi-year mean maps for each pollutant were examined to identify long-term hotspots, defined here as areas with persistently elevated concentrations that stand out from their surroundings in spatial distribution.

8.3.2 Trend analysis

The trend estimation method follows that was detailed in Section 7.1.2. Overall, to evaluate long-term changes while removing seasonality, we formed monthly anomalies at each grid cell by subtracting the corresponding 18-year monthly mean. We then applied ordinary least squares to the anomaly series to estimate linear trends and their statistical significance ($p < 0.01$, $p < 0.05$, and $p<0.1$), following Weatherhead et al. (1998). Trends are reported as $\mu g/m^3$ per year.

8.3.3 Population exposure assessment

We quantified population exposure to total $PM_{2.5}$ and each component at the statewide and regional scales. Monthly total population exposure was calculated following the approach described in Section 7.1.3, using annual LandScan 1-km population data to match the resolution of our concentration fields.

To assess environmental exposure risks among vulnerable populations, we conducted an additional exposure analysis at the census tract level. Total $PM_{2.5}$ and component estimates were overlaid with California census tract boundaries, and mean concentrations for each tract were calculated using zonal statistics. We then joined these tract-level concentrations to CalEnviroScreen 4.0 data, which provides indicators of population vulnerability, including rates of children under age 10, elderly over age 65, asthma emergency department visits, heart attack emergency department visits, and low birth-weight births. This allowed us to compare exposures between communities in the highest vulnerability quartile (>75%) and those in the lowest quartile (<25%) of CalEnviroScreen scores. The comparison focused on differences in average exposures, interquartile ranges, and the spatial distribution of elevated exposure burdens.

8.4 Spatial hotspots and their changes of $PM_{2.5}$ and its components

Across California, multi-year mean maps (Fig. 8.2) for 2000–2020 reveal that the highest concentrations of both total $PM_{2.5}$ and several components are concentrated in the Los Angeles (LA) megaregion and the San Joaquin Valley (SJV). These areas are affected by dense emission sources, frequent stagnation episodes, and meteorological conditions that inhibit dispersion. Outside these hotspots, concentrations are generally lower but still elevated in other urban centers, such as the San Francisco Bay Area, and along certain inland transport corridors. Linear trend analysis (Fig. 8.3) reveals substantial spatial and compositional variations in $PM_{2.5}$ and its components over the past two decades, with the magnitude—and in some cases, even the direction—of change differing markedly by location and species, although an overall decline in $PM_{2.5}$ and most components over the two decades. The component-to-$PM_{2.5}$ ratio analysis (%) for the five species (Fig. 8.4) further indicates that OC was the dominant component across California, particularly in the Sacramento and Los Angeles metropolitan regions. The analysis also highlights pronounced seasonal variations in species composition.

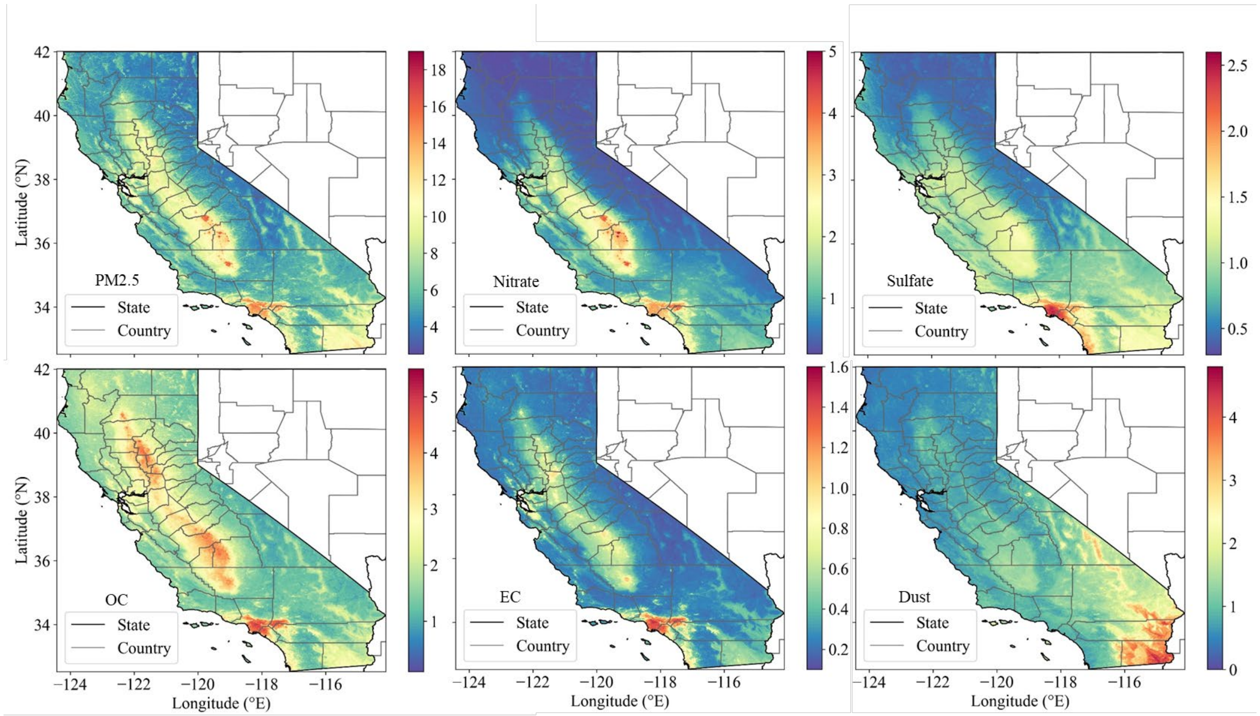

Figure 8.2. Spatial distributions of multi-year mean $PM_{2.5}$ and its five component concentrations (µg/m³) across California from 2000 to 2020.

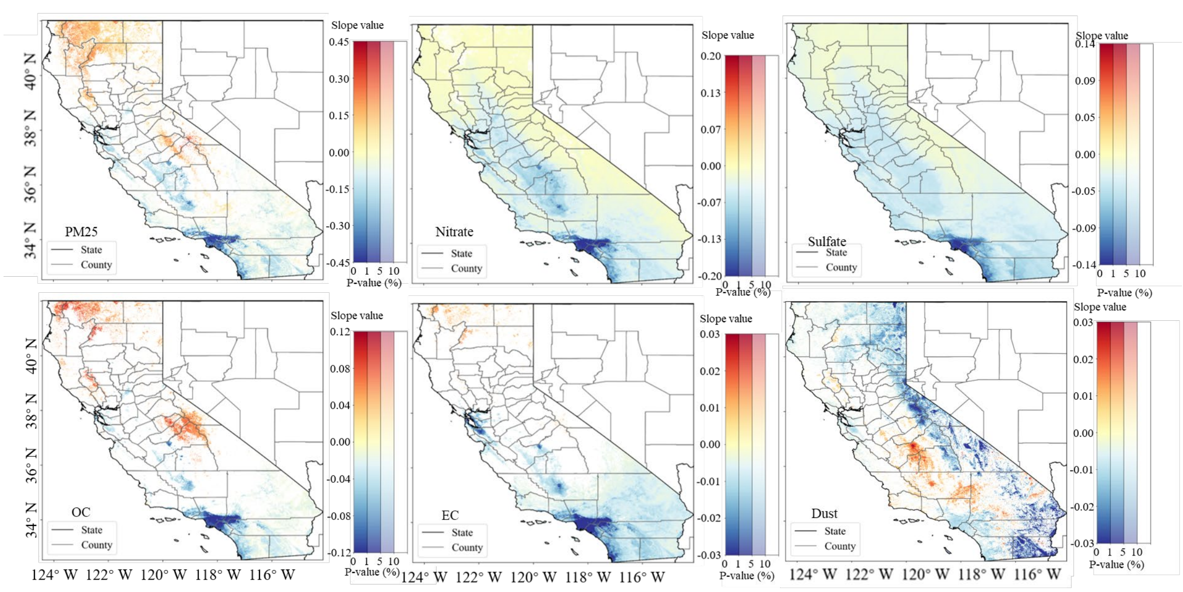

Figure 8.3. Spatial distributions of linear trends of $PM_{2.5}$ and its five components across California from 2000 to 2020.

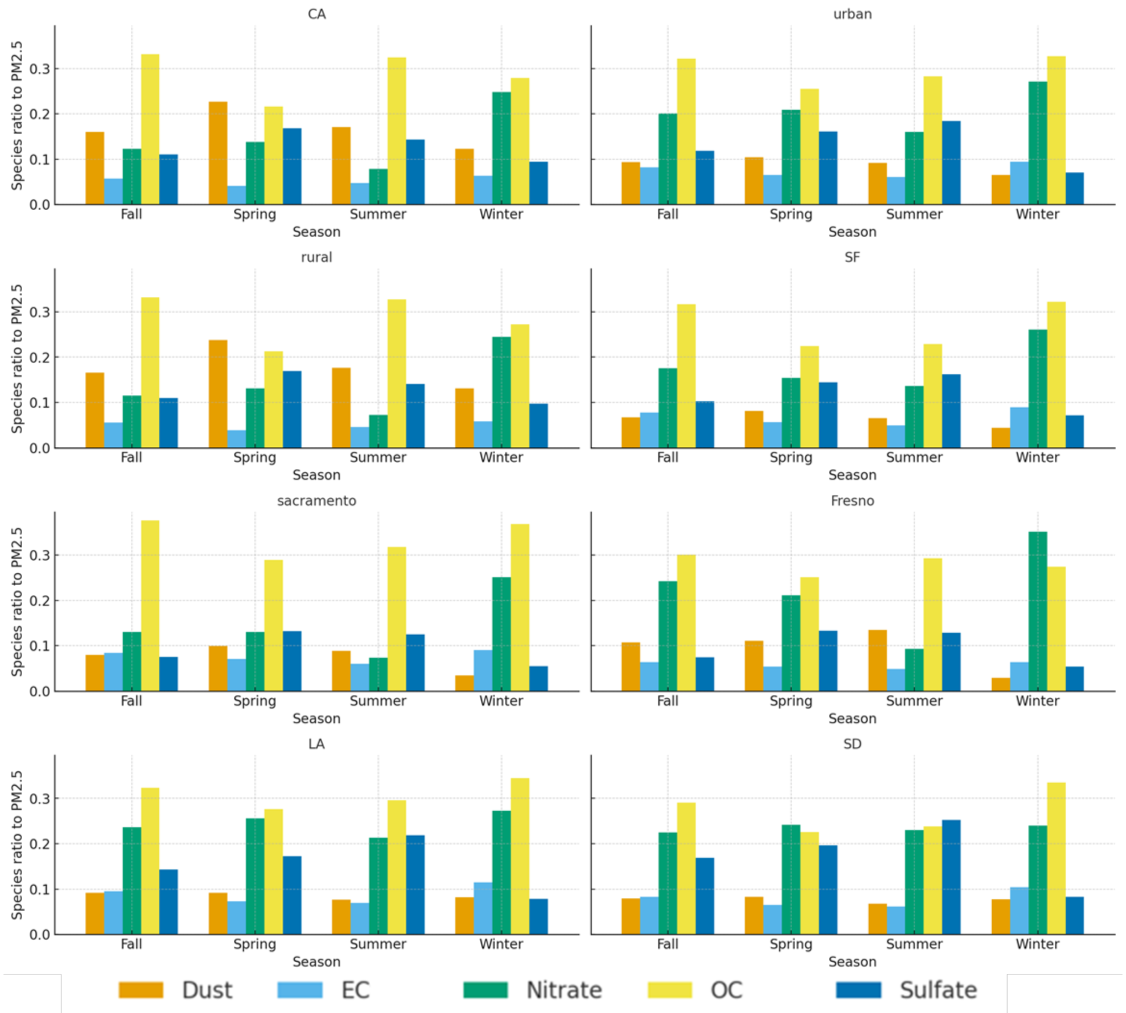

Figure 8.4. Bar plots of component-to-$PM_{2.5}$ ratios (%) of $PM_{2.5}$ and its five components across California and typical regions from 2000 to 2020.

8.4.1 $PM_{2.5}$ total mass

The spatial distribution of multi-year mean total $PM_{2.5}$ shows extremely high concentrations in the LA megaregion, where most grid cells record annual means between 12 and 17 μg/m³. These levels place the entire metropolitan area well above the last U.S. National Ambient Air Quality Standard (NAAQS) of 12 μg/m³ for annual $PM_{2.5}$. The central SJV emerges as another major hotspot, with extensive areas also exceeding 12 μg/m³, reflecting the combined effects of heavy agricultural activity, wintertime ammonium NO_3^- formation, and limited atmospheric mixing due to valley topography. When the latest NAAQS annual standard of 9 μg/m³ is applied—compared to the previous 12 μg/m³ threshold—the exceedance area expands considerably. In addition to the LA megaregion and central SJV, large portions of the broader Central Valley, the San Francisco Bay Area, and the southern California–western Arizona boundary surpass this level.

Pixel-level long-term linear trends reveal pronounced decreases in many of these high-concentration areas. In the LA megaregion, the majority of grid cells exhibit slopes between −0.3 and −0.7 μg/m³ per year ($p < 0.05$), indicative of substantial reductions over two decades. The SJV also shows notable declines, though generally smaller in magnitude, with slopes between −0.1 and −0.3 μg/m³ per year ($p < 0.05$). In contrast, parts of northern California and along the central Sierra Nevada Mountain range display increasing trends exceeding 0.1 μg/m³ per year ($p < 0.05$), suggesting localized influences or source shifts that offset statewide improvements.

8.4.2 Sulfate

SO_4^{2-} concentrations are highest in the LA megaregion, where multi-year means typically range from 1.8 to 2.5 μg/m³. Elevated SO_4^{2-} also extends into the Central Valley, the San Francisco Bay Area, and the southern California–western Arizona boundary, with most concentrations in these areas between 1.2 and 1.6 μg/m³.

Trend analysis shows a consistent downward trajectory across the state. The most substantial declines are concentrated in the LA megaregion, where most slopes are less than −0.10 μg/m³ per year ($p < 0.05$). The Central Valley and the southern California–western Arizona corridor also show widespread decreases, with typical slopes between −0.03 and −0.06 μg/m³ per year ($p < 0.05$).

The sulfate-to-$PM_{2.5}$ ratio is relatively low, averaging 12.97% statewide and ranging from 9.45% in winter to 16.84% in spring. The ratios are nearly identical between urban (12.92%) and rural (12.98%) areas. However, SO_4^{2-} become the dominant contributor to total $PM_{2.5}$ in summer within the San Diego region.

8.4.3 Nitrate

NO_3^- exhibits a spatial footprint similar to total $PM_{2.5}$, with very high concentrations in the LA megaregion (generally 3–5 μg/m³) and extensive elevated areas in the SJV, particularly in its southern portion where many grid cells exceeded 4 μg/m³. The San Francisco Bay Area also showes moderately high NO_3^- concentrations of (2–3 μg/m³).

Linear trends show marked decreases statewide, with the steepest declines in the LA megaregion, where most slopes fall between −0.20 and −0.30 μg/m³ per year ($p < 0.05$). The Central Valley also records broad decreases, typically −0.05 to −0.10 μg/m³ per year ($p < 0.05$), reflecting the combined effect of NO_x emission controls and changes in atmospheric chemistry.

The component ratio analysis further indicates that NO_3^- contributions are substantially higher in urban areas than in rural regions (21.12% vs. 13.00% statewide), with particularly elevated ratios in the Fresno (24.77%), Los Angeles (24.05%), and San Diego (23.01%) megaregions. In addition, wintertime NO_3^- fractions (24.84%) were markedly higher than those in other seasons (7.80–13.81%).

8.4.4 Elemental carbon

EC hotspots are concentrated in the LA megaregion, where multi-year means typically range from 1.2 to 1.5 μg/m³. Other elevated areas include the Central Valley and San Francisco Bay Area, generally around 0.8 μg/m³.

Trend maps reveal only modest decreases in EC, with the LA megaregion showing slopes between −0.01 and −0.05 μg/m³ per year ($p < 0.05$). Notably, increases are observed in parts of northern California and along the central Sierra Nevada, with slopes of 0.02 to 0.05 μg/m³ per year ($p < 0.05$).

Among the five major $PM_{2.5}$ species, EC contributes the least to total $PM_{2.5}$ mass, with a statewide mean EC-to-$PM_{2.5}$ ratio of 5.20%. Over rural areas, the ratio varies only slightly across seasons (3.86–5.87%). However, over urban areas, EC fractions are substantially higher in autumn (8.24%) and winter (9.42%) than in spring (6.49%) and summer (6.07%).

8.4.5 Organic carbon

OC shows the highest spatial extent among the carbonaceous species. The LA megaregion records multi-year means above 4 μg/m³, while the South Sacramento Valley and southern SJV air basins also show pronounced hotspots with similar concentration levels. The San Francisco Bay Area and the southern California–western Arizona boundary have moderate OC levels, around 2.5 μg/m³. The larger footprint compared to EC reflects both primary emissions and secondary organic aerosol formation from VOC precursors.

Trend patterns for OC resemble those of EC but with stronger magnitudes. In the LA megaregion, slopes typically range from −0.10 to −0.20 μg/m³ per year ($p < 0.05$), indicating meaningful reductions. However, in northern California and the central Sierra Nevada, increases of 0.04 to 0.12 μg/m³ per year ($p < 0.05$) are observed, suggesting region-specific influences such as biomass burning or changing precursor availability.

OC remains the dominant contributor to total $PM_{2.5}$, with a statewide mean OC-to-$PM_{2.5}$ ratio of 29.60%. Seasonally, autumn (33.14%) and summer (32.44%) exhibit notably higher OC fractions than winter (27.92%) and spring (21.62%). OC ranks as the leading component across nearly all regions and seasons, except during winter in the Fresno region and spring–summer in the San Diego megaregion, where other species (e.g., NO_3^-) become more influential.

8.4.6 Mineral dust

The DUST component shows the highest concentrations in the southwestern areas of California, particularly along the border with Arizona, where most grid cells exceed 3 μg/m³. DUST levels in southern California, especially across the southern SJV and San Bernardino County (mostly 1.2–2 μg/m³), are clearly higher than those in the northern part of the state, where most values remain below 1.2 μg/m³.

Linear trend analysis shows the largest decreasing trends (below −0.03 μg/m³ per year, $p < 0.05$) along the southern state border, where the highest DUST concentrations are observed, and eastern border areas. In contrast, increasing trends of 0.003–0.05 μg/m³ per year ($p < 0.05$) are observed in the southern SJV region.

The DUST component is the second-largest contributor to total $PM_{2.5}$ statewide, with a dust-to-$PM_{2.5}$ ratio of 17.01%. Its contribution is larger in rural areas (17.79%) than in urban areas (8.70%), and highest in spring (22.64%), compared with 12.30–17.12% in other seasons.

8.4.7 Discussion

The analysis of $PM_{2.5}$ speciation across California reveals distinct spatial patterns, temporal trends, and compositional characteristics that collectively inform understanding of the dominant contributors to fine particulate pollution and their evolution over the past two decades. Total $PM_{2.5}$ concentrations remain highest in the LA megaregion and the SJV, both consistently exceeding national air quality standards, although significant long-term declines are evident due to regulatory and technological improvements.

Among the five major components, OC is the dominant contributor statewide, accounting for nearly 30% of total $PM_{2.5}$, while EC is the smallest contributor statewide (5%). The carbonaceous fraction (OC + EC) collectively represents a substantial portion of $PM_{2.5}$ mass, particularly in the LA megaregion and urbanized areas, reflecting the influence of combustion sources and secondary organic aerosol formation. In contrast, NO_3^- is the second-largest contributor in urban areas (21%) and during winter (25%), especially in the SJV and southern California, driven by low temperatures and stagnant meteorology that promote ammonium NO_3^- formation. SO_4^{2-} contributions have steadily decreased, consistent with long-term SO_2 emission reductions, though localized dominance in SD megaregion during summer highlights the continued influence of photochemical oxidation and marine air. DUST remains an important contributor in rural areas (18%), particularly in the southwestern border areas and southern region, where resuspension and agricultural activities play key roles.

Spatially, $PM_{2.5}$ composition exhibits strong regional differentiation—OC and NO_3^- species dominate in densely populated basins, while DUST is more pronounced in rural regions. Pixel-level linear trend analysis (Fig. 8.3) reveals pronounced spatial heterogeneity in the magnitude and direction of $PM_{2.5}$ and species changes across California. Most grid cells in the LA metro megaregion and the SJV show statistically significant ($p < 0.05$) decreases in total $PM_{2.5}$, generally between −0.3 and −0.7 μg/m³/yr and −0.1 to −0.3 μg/m³/yr, respectively, reflecting sustained emission reductions. SO_4^{2-} and NO_3^- exhibit the strongest and most spatially coherent declines statewide, consistent with long-term control of sulfur and nitrogen precursors. OC and EC also decrease over most urban basins but with smaller magnitudes and localized reversals in parts of northern California and the Sierra Nevada foothills, suggesting region-specific influence, potentially linked to biomass burning, wildfire smoke, and changing precursor availability. Dust shows weak or mixed trends, with slight decreases along the southern border and minor increases in the southern SJV.

Overall, the compositional and trend analyses highlight that, despite substantial progress in reducing $PM_{2.5}$ mass, species-specific dynamics vary across regions and seasons. OC and NO_3^- remain key targets for further mitigation, while the persistence of elevated levels or local increases in EC and DUST emphasize the need for regionally tailored strategies that consider source profiles, chemical regimes, and meteorological constraints.

8.5 Spatiotemporal patterns of population exposure to $PM_{2.5}$ and its components

8.5.1 Long-term variation of total population exposures

From 2000 to 2020, the population-weighted mean $PM_{2.5}$ concentration in California was 11.87 ug/m³. Annual exposure declined from approximately 14.8 ug/m³ in 2000 to 11.9 ug/m³ in 2020, a reduction of about 20% (Fig. 8.5). Despite this long-term decrease, seasonal and interannual variability remained substantial: exposures were highest in winter (13.52 ug/m³) and lowest in spring (9.46 ug/m³).

These seasonal patterns are largely driven by variations in component concentrations, particularly OC and NO_3^-. NO_3^- exhibited the most pronounced long-term reduction, declining by about 52% from 3.9 μg/m³ in 2000 to 1.9 μg/m³ in 2020. On average, NO_3^- contributed ~2.6 μg/m³ to total $PM_{2.5}$ but displayed sharp seasonality, with winter peaks often exceeding 6–7 μg/m³ during stagnant episodes. These peaks make NO_3^- the single most important driver of $PM_{2.5}$ seasonality in California. SO_4^{2-} also showed substantial reductions, falling by nearly 50% from 2.2 μg/m³ in 2000 to 1.1 μg/m³ in 2020. Its average contribution (~1.6 μg/m³) is smaller than that of NO_3^- or OC. Seasonal variation is less pronounced, though concentrations tend to be slightly higher in summer due to enhanced photochemical production. Importantly, sulfate's decline is steady and spatially consistent, reflecting effective controls on sulfur emissions. OC contributed the largest share of any single species, averaging ~3.5 μg/m³ over the study period. Its long-term decline was modest (~6%), but year-to-year variability was high. Seasonal patterns showed clear winter (4.48 μg/m³) and fall (4.16 μg/m³) peaks, moderate summer levels (3.04 μg/m³), and pronounced summer–fall surges from 2017 to 2020 associated with major wildfire episodes. These episodic events have interrupted the otherwise gradual downward trend and highlight the growing role of wildfire emissions in recent years. EC decreased modestly, from 1.23 μg/m³ in 2000 to 1.08 μg/m³ in 2020 (a 13% reduction). With an average contribution of ~1.0 μg/m³, EC plays a smaller role compared with NO_3^- or OC. Seasonal variation is weaker, though somewhat elevated levels occur in fall and winter, consistent with combustion-related sources such as residential burning and on-road traffic. DUST remained relatively stable, averaging ~0.9 μg/m³ with little change over the two decades. It consistently contributes the least among the five components, with limited seasonal variability except for slightly higher concentrations in spring and fall, likely due to resuspension and natural surface processes.

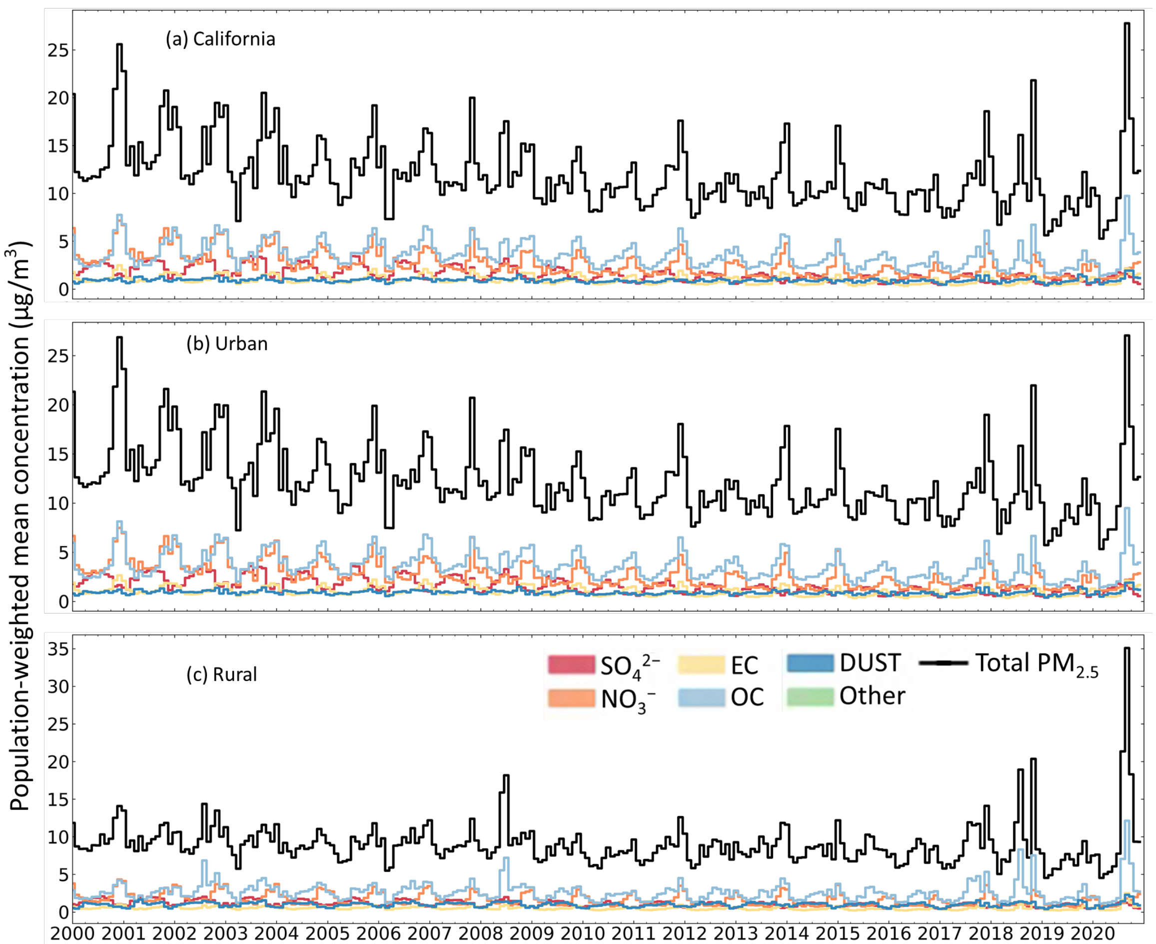

Figure 8.5. Monthly population-weighted mean concentrations from 2000 to 2020.

Overall, population-weighted $PM_{2.5}$ exposure in California decreased by about 20% from 2000 to 2020. As illustrated in the bar chart comparing 2000 and 2020 (Fig. 8.6), the relative reduction in total $PM_{2.5}$ exposure is driven most strongly by NO_3^- and SO_4^{2-}, both of which declined by roughly half. NO_3^- continues to dominate the pronounced wintertime peaks, while SO_4^{2-} shows a steady decline during summer. OC and EC exhibit relatively modest long-term decreases of approximately

12% and 6%, respectively, with OC trends further obscured by episodic wildfire-driven spikes in recent years. In contrast, dust concentrations remain essentially unchanged.

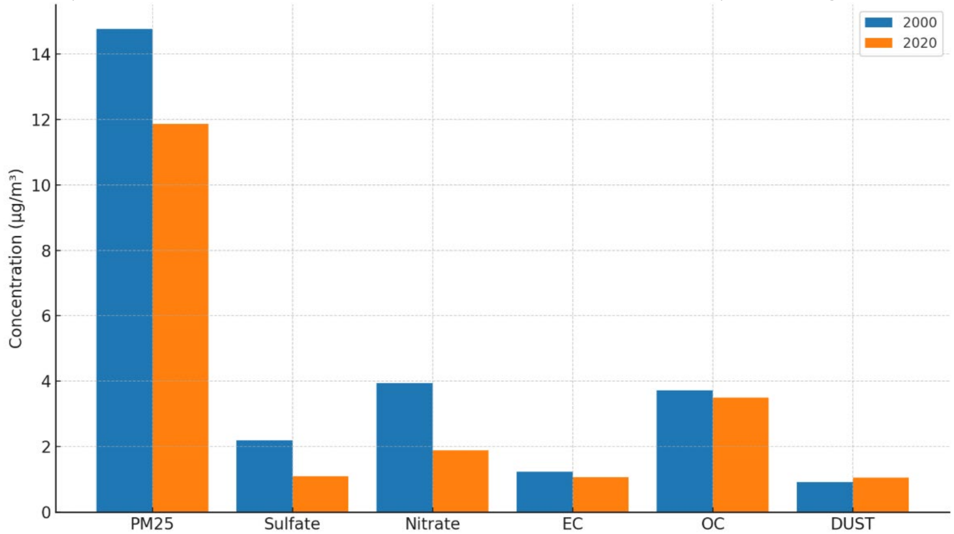

Figure 8.6. Average population-weighted mean concentrations of $PM_{2.5}$ and its five components in 2000 vs. 2020 over California.

These spatiotemporal variations in exposure reflect the combined influence of emission controls, changing atmospheric conditions, and episodic wildfire events. However, the patterns are complex and not fully explained by emissions alone, underscoring the need for further decomposition of the roles of meteorology, wildfires, and human emissions. Such decomposition analysis will be presented in Section 9.

8.5.2 Exposure risk patterns of vulnerable populations

8.5.2.1 Spatial patterns of vulnerable populations by CalEnviroScreen

Figure 8.7 presents six vulnerability parameters: total population (TotPop19), children under 10 (Child_10), elderly adults 65+ (Elderly65), and three health outcome indicators—asthma emergency department visits (Asthma), cardiovascular emergency department visits (Cardiovas), and low birth-weight births (LowBirthWt). TotPop19 indicates that densely populated tracts cluster in the major metropolitan areas—most prominently the LA megaregion, the SF Bay Area, and portions of the SJV. Tracts with higher counts of children under 10 are likewise concentrated in metropolitan corridors and the SJV, whereas elderly populations show an inverse pattern, with higher densities in parts of the western/northern mountain and foothill regions.

The three health indicators display distinct but partially overlapping spatial patterns that are consistent with the maps shown. Asthma emergency department (ED) visits are elevated in portions of the LA basin, along the SJV corridor, and in pockets of the Inland Empire, reflecting known respiratory burdens in these regions. Cardiovascular ED visits show a broader inland footprint, with higher values across much of the SJV, the LA megaregion, and scattered rural tracts elsewhere. Low birth-weight prevalence is more heterogeneous, with elevated rates in selected tracts of the SF

Bay Area, across segments of the Central Valley, and in southern California, including parts of greater LA.

Taken together, these patterns indicate that vulnerable populations are not confined to a single geography: different indicators highlight different at-risk groups across both urban and rural California. This spatial diversity underscores the importance of overlaying these vulnerability layers with $PM_{2.5}$ and species-specific exposures to pinpoint tracts where pollution burden and population sensitivity coincide most strongly, thereby guiding targeted mitigation and health-protection strategies.

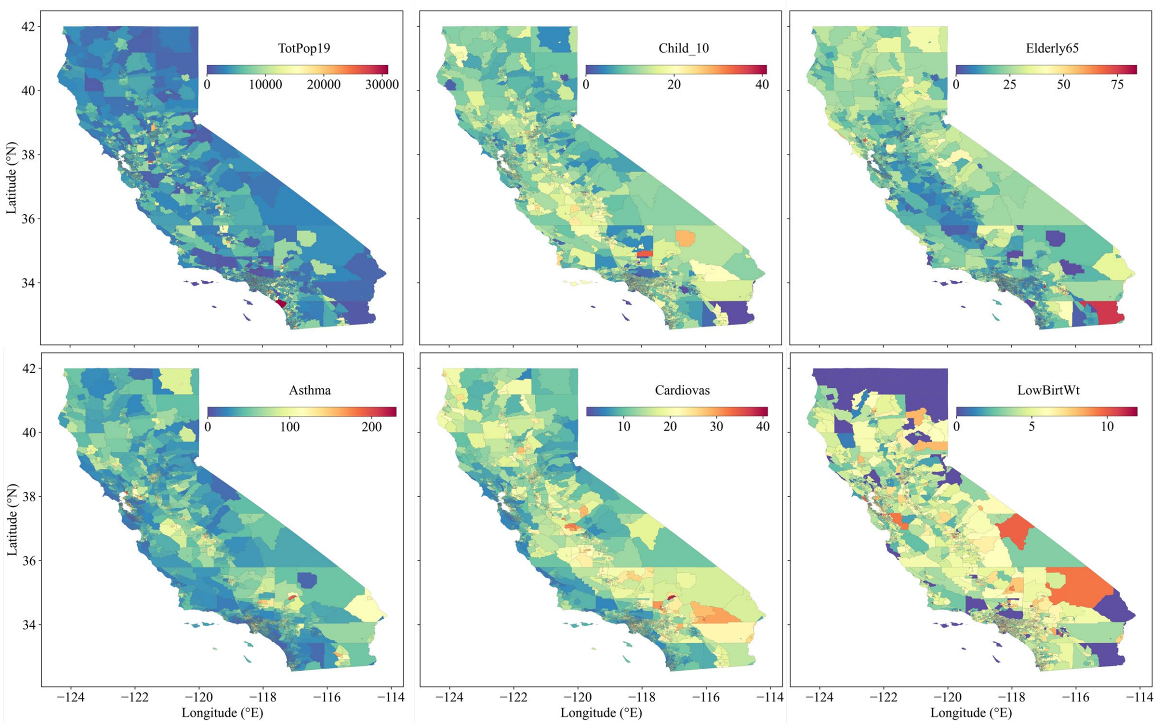

Figure 8.7. Spatial distributions of CalEnviroScreen scores for Total population (TotPop19), populations with ages below 10 (Child_10), populations older than 65 (Elderly65), Asthma, Cardiovas, and low birth weight (LowBirtWt).

8.5.2.2 Exposure patterns of vulnerable populations to $PM_{2.5}$ components

To assess population vulnerability to PM2.5 component exposures, we examined spatial variation in total $PM_{2.5}$ and its five major species relative to six CalEnviroScreen 4.0 indicators (TotPop19, Child_10, Elderly65, Asthma, Cardiovas, LowBirthWt). Figures 8.8 identifies tracts where high pollution burdens coincide with high vulnerability. Because tracts with high $PM_{2.5}$ (total and by component ≥75th percentile) are concentrated in southern California, overlaid hotspots occur primarily in southern California such as the LA metro and Fresno urban clusters.

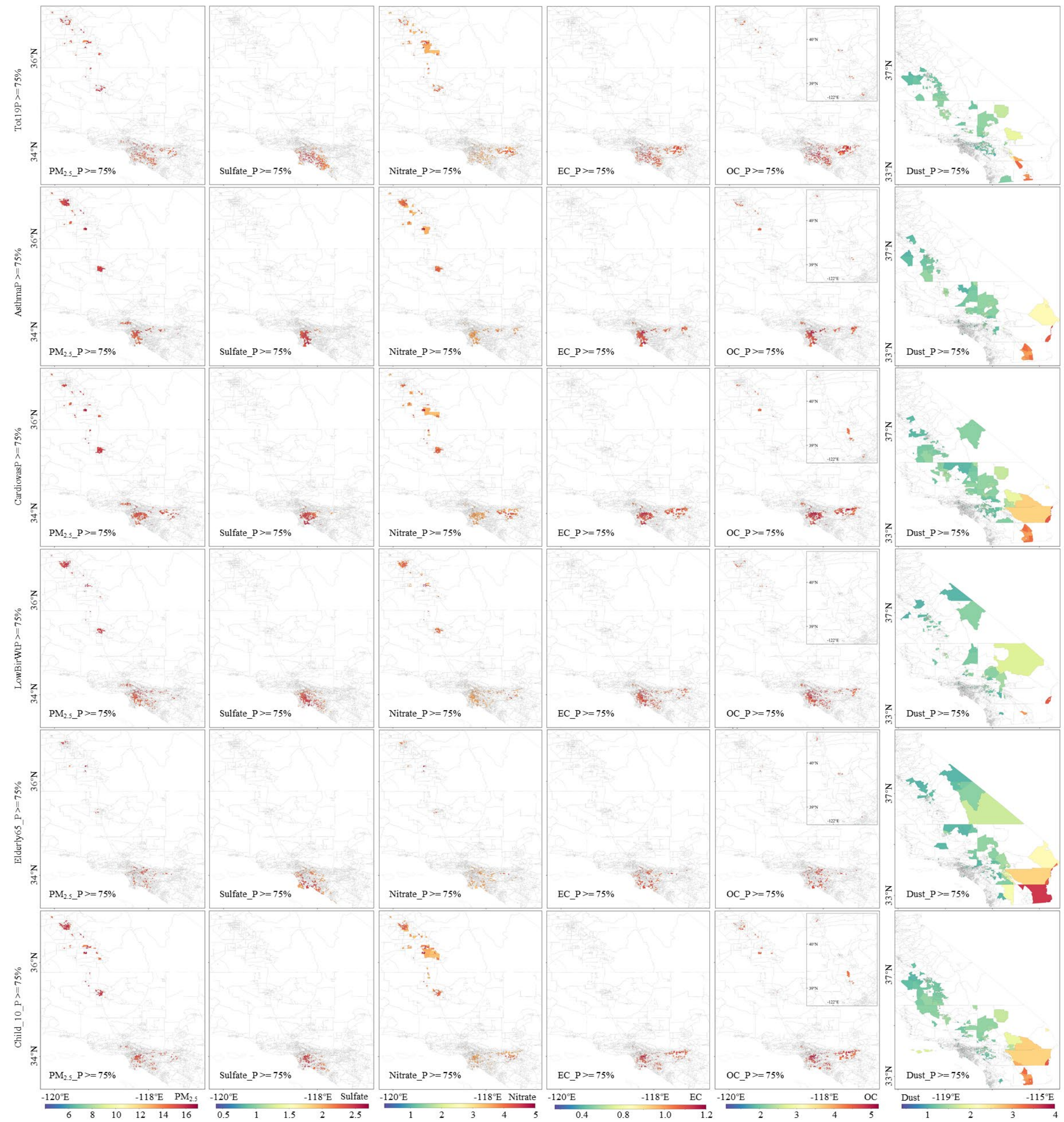

Figure 8.8. Spatial distributions of $PM_{2.5}$ total mass and component concentrations over census tracts where both the selected CalEnviroScreen parameter scores and pollutant concentrations are at or above the 75th percentile during the 2000–2020 period.

Figures 8.9–8.10, together with the 2000–2020 time series shown in Figure 8.11, demonstrate pronounced and systematic disparities in $PM_{2.5}$ component exposures between census tracts in the highest vulnerability quartile (≥75%) and those in the lowest quartile (≤25%). The magnitude and direction of these disparities vary across vulnerability indicators and $PM_{2.5}$ species. For total population (TotPop19), exposure differences between high- and low-vulnerability tracts are generally small, with slightly higher NO_3^- (+0.03 μg/m³) and DUST (+0.01 μg/m³) but marginally lower SO_4^{2-} (−0.03 μg/m³), EC (EC; −0.03 μg/m³), and OC (OC; −0.05 μg/m³) concentrations in ≥75% tracts. In contrast, disparities are much larger for tracts with elevated asthma emergency department (ED) visit rates, which are concentrated in the San Joaquin Valley (SJV), Los Angeles

Basin, and Inland Empire. These high-asthma areas experience substantially higher exposures to NO_3^- (+0.30 μg/m³), OC (+0.45 μg/m³), EC (+0.10 μg/m³), and DUST (+0.14 μg/m³), with SO_4^{2-} slightly lower (−0.04 μg/m³). Similar patterns are observed for cardiovascular ED visits, with systematically elevated exposures in the ≥75% quartile across all major components: NO_3^- (+0.41 μg/m³), OC (+0.59 μg/m³), EC (+0.13 μg/m³), DUST (+0.27 μg/m³), and SO_4^{2-} (+0.07 μg/m³). Tracts with high prevalence of low birth weight likewise face consistently higher exposures to all components, including NO_3^- (+0.47 μg/m³), OC (+0.42 μg/m³), EC (+0.16 μg/m³), SO_4^{2-} (+0.18 μg/m³), and DUST (+0.07 μg/m³). Children-dense areas (≥75% for population under age 10) show elevated NO_3^- (+0.31 μg/m³), OC (+0.28 μg/m³), EC (+0.05 μg/m³), and DUST (+0.15 μg/m³) exposures, while SO_4^{2-} concentrations are virtually identical between groups (1.54 μg/m³). In contrast to all other indicators, tracts with high proportions of older adults (≥65 years) — often located in more rural or northern California regions — experience significantly lower exposures than their ≤25% counterparts across all components, with NO_3^- (−0.75 μg/m³), OC (−0.68 μg/m³), SO_4^{2-} (−0.29 μg/m³), EC (−0.24 μg/m³), and DUST (−0.12 μg/m³) all markedly reduced.

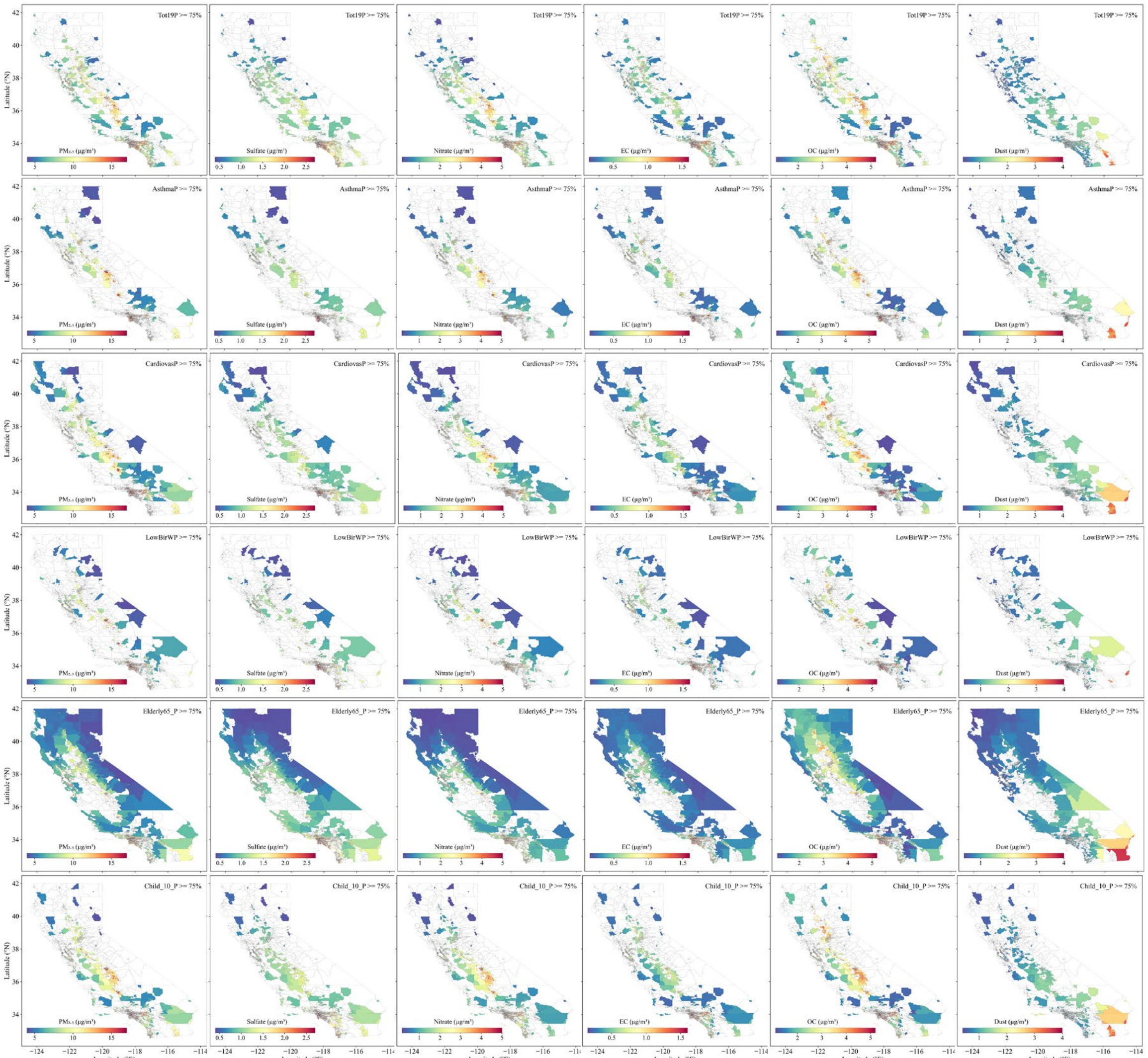

Figure 8.9. Spatial distributions of $PM_{2.5}$ total mass and component concentrations over tracts with each selected parameter's CalEnvironSreen score percentile>=75% for the 2000-2020 period.

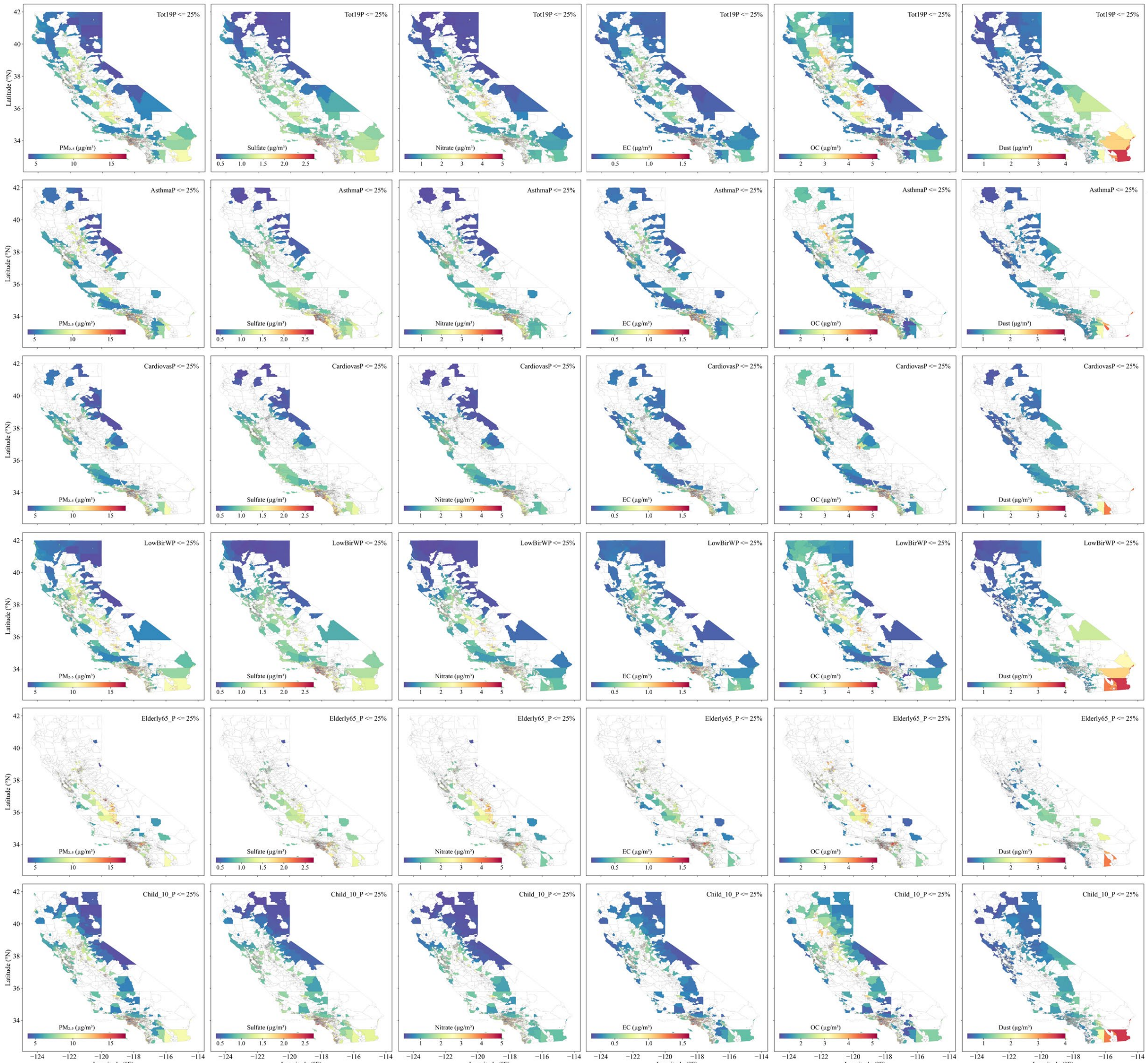

Figure 8.10. Spatial distributions of $PM_{2.5}$ total mass and component concentrations over tracts with each selected parameter's CalEnvironSreen score percentile<=25% for the 2000-2020 period.

Figure 8.11. Time series of $PM_{2.5}$ component exposures (2000–2020) for census tracts in the highest (≥75%) and lowest (≤25%) quartiles of CalEnviroScreen indicators, including asthma, cardiovascular disease, low birth weight, children under 10, elderly population, and total population density. Colors indicate $PM_{2.5}$ components: Colors represent $PM_{2.5}$ components: NO_3^- (N, red), OC (O, green), EC (E, black), SO_4^{2-} (S, blue), and DUST (D, orange).

Taken together, the results reveal that NO_3^- and OC drive the largest disparities for most vulnerable groups, particularly those defined by elderly, cardiovascular disease, and low birth weight indicators. EC and DUST also show elevated exposures in these groups, though to a lesser degree. In contrast, SO_4^{2-} exhibits weaker disparities, and for asthma, children and total population indicators, differences are negligible or reversed. Across CalEnviroScreen indicators, communities defined by cardiovascular disease, child, and low birth weight vulnerabilities consistently experienced higher $PM_{2.5}$ component exposures in ≥75% tracts compared with ≤25% tracts, whereas elderly-dense tracts showed a reversed pattern with lower exposures. Asthma-vulnerable tracts displayed mixed disparities, with higher exposures for NO_3^-, OC, EC, and DUST but slightly lower SO_4^{2-} concentrations. In contrast, total population indicators exhibited only negligible differences. These findings underscore that exposure inequalities in California are primarily tied to local combustion sources and wildfire episodes, rather than regional background pollutants.

8.6 Summary

Using long-term, high-resolution estimates spanning 2000–2020, this section examined the spatial hotspots and temporal trends of $PM_{2.5}$ and its five major components across California. Spatial analyses revealed persistent hotspots in the SJV, the LA metropolitan area, and parts of the SF Bay Area, where elevated concentrations of OC and NO_3^- contributed strongly to the overall $PM_{2.5}$ burden. Pixel-level linear trend analysis showed substantial statewide declines in SO_4^{2-} and NO_3^-, typically −0.05 to −0.10 µg/m³/yr ($p < 0.05$), while OC and EC decreased more modestly—about −0.01 to −0.05 µg/m³/yr —with localized reversals in northern California and the Sierra Nevada foothills. DUST exhibited weak or mixed trends, showing slight decreases along the southern border and minor increases in the southern SJV.

Population-weighted exposure analyses revealed somewhat different spatial and temporal patterns when pollutant estimates were overlaid with population distribution. Statewide $PM_{2.5}$ exposure declined by roughly 20 %, from 14.8 µg/m³ in 2000 to 11.9 µg/m³ in 2020, with about half of this reduction driven by decreases in NO_3^- and SO_4^{2-}. Seasonal cycles were evident, with NO_3^- driving winter peaks that often exceeded 6–7 µg/m³ and OC dominating in summer and fall, frequently amplified during wildfire years. When exposures were stratified by CalEnviroScreen vulnerability indicators, clear disparities emerged: communities with higher burdens of asthma, cardiovascular disease, child populations, or low birth weight experienced 0.28–0.59 µg/m³ higher NO_3^- and OC exposures, whereas tracts with higher proportions of elderly residents showed lower exposures, and differences by total population were minimal. These results highlight the uneven distribution of pollution burdens among sensitive populations.

Overall, the analysis demonstrates that although air-quality improvements have substantially reduced $PM_{2.5}$ and its major components statewide over the past two decades, exposure inequalities persist and remain most pronounced for nitrate- and carbon-related species in vulnerable communities. The complexity of these spatiotemporal patterns underscores the need to further decompose the drivers of variability—including the respective roles of meteorology, wildfires, and anthropogenic emissions—which are examined in Section 9.

9. Unraveling Contributions of Meteorology and Wildfire Smoke to $PM_{2.5}$ Components in California and Their Influence on Recent Trends

9.1 Introduction

Understanding the drivers of temporal and spatial variability in $PM_{2.5}$ and its major chemical components is fundamental for developing targeted mitigation strategies in California, where complex interactions among anthropogenic emissions, meteorology, and wildfires lead to substantial fluctuations in air quality. Unlike many other regions of the United States, California is characterized by a unique combination of persistent urban and industrial emissions, intensive agricultural activity, and frequent large-scale wildfires. These diverse sources contribute to both chronic background pollution and acute episodic events, resulting in significant health and environmental burdens. A particular challenge for assessing the effectiveness of air quality controls in California is thus that both meteorological variability and wildfire smoke can obscure underlying emission-driven trends. Meteorological conditions influence pollutant dispersion, chemical transformation, and deposition, generating substantial interannual and seasonal variability independent of emission changes. Wildfires, in contrast, produce abrupt but increasingly frequent episodes of extreme particulate matter pollution—especially enriched in OC and EC—that can mask or even offset improvements achieved through anthropogenic emission reductions. Without explicitly accounting for these two influences, evaluations of long-term air quality trends risk underestimating the benefits of emission controls or misattributing the sources of observed changes.

In Section 8, we identified the major spatial hotspots of $PM_{2.5}$ and its components across California and documented long-term declines in population-weighted exposures, particularly in SO_4^{2-}, NO_3^-, EC, and OC. We also highlighted the growing role of wildfire events in shaping exposure extremes, often offsetting some of the gains achieved through emission reductions. Building upon those findings, the present section seeks to disentangle the underlying drivers of these observed patterns. Specifically, we aim to quantify how much of the observed variability and trend in $PM_{2.5}$ species can be attributed to meteorological conditions versus wildfire smoke, and how much reflects long-term changes in anthropogenic emissions. Therefore, separating the relative contributions of meteorology, wildfire smoke, and human activities is critical for accurate attribution.

In this section, we employ the extended daily, high-resolution $PM_{2.5}$ component dataset developed in Sections 6 and 8, covering the 2000–2020 period, to conduct decomposition analyses that partition observed variability into contributions from meteorology, wildfire smoke, and anthropogenic emissions. By removing the confounding effects of weather and episodic fire events, we obtain a clearer picture of long-term emission-driven trajectories in $PM_{2.5}$ and its species. Finally, we discuss the implications of these refined trends for emission-control policy and provide recommendations for strategies that simultaneously address chronic and episodic particulate matter sources in California.

9.2 Decomposing Meteorological and Wildfire Smoke Contributions to $PM_{2.5}$ and Its Five Components

To better interpret the drivers of $PM_{2.5}$ variability and identify components most responsive to emission control policies, we decomposed daily $PM_{2.5}$ (and its major chemical species) into three additive parts: meteorologically driven, wildfire-related, and anthropogenic residual components, as below abstracted equation:

$$\mathrm{PM}_{g,t} = \underbrace{\mathrm{Meteo}_{g,t}}_{(1)} + \underbrace{\mathrm{Smoke}_{g,t}}_{(2)} + \underbrace{\mathrm{Residual}_{g,t}}_{(3)}$$

(1) Meteorological contribution: Meteorological effects were removed using the normalization method described in Section 9.2.1, which predicts $PM_{2.5}$ concentrations under a fixed reference meteorological condition. The difference between observed and meteorology-normalized values thus represents the contribution of short-term meteorological variability.

(2) HMS-based smoke on the met-normalized field: The HMS-based five-step procedure (Section 9.2.2-9.2.3) was applied to the meteorology-normalized $PM_{2.5}$ and component fields to identify and remove days and locations affected by wildfire smoke. This yielded estimates of wildfire-attributed $PM_{2.5}$, as well as "clean" non-fire concentrations.

(3) Anthropogenic/policy-sensitive residual: Finally, we attributed the remaining portion after accounting for meteorological and wildfire influences as the anthropogenic or policy-sensitive residual, reflecting long-term changes primarily driven by emission sources and regulatory actions. This residual component provides an empirical foundation for assessing emission-driven trends, tracking sector-specific mitigation progress, and supporting policy considerations for sustained and equitable air-quality improvements across California.

9.2.1 Meteorological effect decomposition

To isolate the effect of meteorological variability, we replaced daily meteorological inputs in the modeling framework with "long-term stable" meteorology generated through a seasonal resampling procedure. For each grid cell and day, 50 values were randomly drawn for each meteorological variable (Table 6.1) from the same calendar month across the baseline period (2000–2020). This approach preserves the seasonal cycle but removes interannual anomalies. For example, meteorological inputs for 1 July 2018 were replaced with values randomly sampled from all July records (1–31) across the 2000–2020 period. All other predictors, including emissions, land use, and wildfire indicators, were held constant.

The model was rerun with the resampled meteorology to produce meteorology-normalized concentrations. The contribution of meteorology was then calculated as:

$$PM_{MET}(g,t) = PM_{ori}(g,t) - PM_{DEMET}(g,t)$$

where $PM_{ori}(g,t)$ is the original estimate at location g and day t, and $PM_{DEMET}(g,t)$ is the estimate under resampled meteorology. Positive values of $PM_{MET}(g,t)$ indicate meteorological conditions that enhanced pollution, whereas negative values indicate conditions that favored dispersion or removal.

9.2.2 Separating wildfire-smoke contribution

After meteorological normalization, we applied a five-step algorithm (Steps A–E) to separate wildfire-related ("fire") and background ("non-fire") $PM_{2.5}$ concentrations. This approach integrates spatial smoke plume data with a robust statistical filtering framework to generate spatially and temporally consistent estimates of both smoke-related and background $PM_{2.5}$ across California. This design quarantines smoke-driven extremes from the background, so that the background (i.e., the anthropogenic-attributed residual after removing meteorology and wildfire) is not inflated by smoke spikes, while the full wildfire signal is retained in dedicated smoke metrics. Daily wildfire smoke plume information was obtained from the NOAA Hazard Mapping System (HMS) smoke product, which provides daily polygons delineating smoke plume extents based on satellite observations. As the HMS record is available only from mid-2005 onward, the wildfire-related analysis period was therefore set to 2005–2020.

Step A. HMS-based candidate gate (source plausibility)

All datasets were reprojected to a standardized 1-km grid consistent with the high-resolution $PM_{2.5}$ component data spanning the land area of California. For each grid cell g and day t, a candidate flag was assigned if the pixel intersected any HMS smoke plume polygon on that day. This step defines all locations potentially influenced by wildfire emissions, serving as the initial plausibility screen.

Step B. Baseline construction

For each grid cell g and day t, a baseline distribution of $PM_{2.5}$ concentrations was established using all non-fire days for the same day of year (DOY) during the 2005-2020 study period. To ensure a sufficient number of valid values for calculating baseline statistics, we applied a small DOY window: all days s satisfying $|DOY(s)-DOY(t)|\leq W$ were included, where W=1 by default and expanded as needed.

Days already flagged as fire candidates in Step A were excluded for the baseline calculation. From this baseline sample, we computed the upper quartile Q3(g,t) and the interquartile range IQR(g,t)=Q3−Q1. If the baseline sample size was smaller than a minimum threshold (Nmin=10), the window width W was expanded incrementally (e.g., 3-7 days) until sufficient samples were obtained.

Step C. Robust exceedance test

Each grid cell and day were classified as smoke-impacted if:

$$PM_{g,t} > Q3(g,t) + k \cdot IQR(g,t),$$

where k=1.5 defines a robust-outlier threshold, following previous study (Wei et al. 2023).
To reduce false positives from isolated local anomalies, spatial coherence was imposed: a cell remained flagged only if at least one neighboring pixel within a 3×3 window was also identified as smoke-impacted.

Step D. Baseline refinement

After smoke-impacted pixels were identified, baseline statistics (i.e., median) were recomputed using only non-fire days. This refinement step effectively "decontaminates" the baseline by removing smoke influence, ensuring that the background levels represent true non-fire conditions. For reproducibility, Step C was not rerun with the refined baselines.

Step E. Derivation of fire and non-fire $PM_{2.5}$

To obtain daily non-fire concentrations, grid cells identified as fire-affected outliers (Step C) were replaced with their corresponding background values. The background was defined as the DOY-specific median from the meteorology-normalized dataset in Step D, thereby preserving seasonal cycles while excluding fire anomalies. The resulting dataset represents daily spatial distributions of non-fire concentrations ($PM_{nonfire}(g,t)$).

$$PM_{nonfire(g,t)} = \begin{cases} DOY\_median_{PM\,DEMET}(g,t), & if\ fire\ pixel, \\ PM_{DEMET}(g,t), & if\ not\ fire\ pixel, \end{cases}$$

Fire-attributed concentrations were then calculated as the residual between the meteorology-normalized dataset and the background dataset:

$$PM_{fire}(g,t) = PM_{DEMET}(g,t) - PM_{nonfire}(g,t)$$

This procedure ensures that wildfire impacts are quantified relative to the expected seasonal background under meteorology-normalized conditions.

9.2.3 Application to $PM_{2.5}$ components

The smoke-filtering framework described above (Section 9.2) was further extended to quantify the wildfire and non-wildfire contributions of five major $PM_{2.5}$ components— SO_4^{2-}, NO_3^-, OC, EC, and DUST—each available as daily 1-km estimates for the same domain and period (2005–2020). Because total $PM_{2.5}$ represents the combined signal of these species and exhibits the highest signal-to-noise ratio, the Step C smoke mask derived from total $PM_{2.5}$ was adopted as a common indicator of smoke exposure for all components. This approach ensures consistency in the spatial and temporal definition of fire events and facilitates cross-species comparisons.

For each species, we recomputed baseline statistics following the procedure in Step D but using the species-specific concentration fields. All grid cells and days flagged as smoke-impacted in Step C were excluded from these calculations, yielding "clean" DOY baselines that represent typical non-fire conditions for each component. These species-specific baselines (median, Q1, Q3, and IQR) were subsequently used to separate smoke and non-fire concentrations.

In analogy with Step D, smoke and non-fire concentrations for each species were derived. This design maintains a coherent smoke definition across all $PM_{2.5}$ components—ensuring that species are evaluated under a shared exposure framework—while allowing each component to preserve its own statistically independent baseline for replacement and analysis. The resulting species-specific smoke and non-fire fields provide a consistent basis for inter-species comparison of wildfire impacts, emission policies, and long-term compositional changes in $PM_{2.5}$ across California.

9.2.4 Statistical analyses of the decomposed results

After generating the daily 1-km fields with meteorological and wildfire contributions decomposed, we conducted spatiotemporal statistical analyses at both the pixel level and for region-aggregated units. Our focus areas included the Southern California Air Basin (SoCAB) and the San Joaquin Valley (SJV), as well as urban areas (UA) in California defined by the 2010 U.S. Census Bureau urbanized-area boundaries; rural areas (RA) were defined as the portions of California outside those urban boundaries. To capture heterogeneity in sources and impacts, we further examined five major metropolitan areas—San Francisco Bay Area (SF Bay), Los Angeles Metropolitan Area (LA), Sacramento, Fresno, and San Diego (SD) (Fig. 9.1).

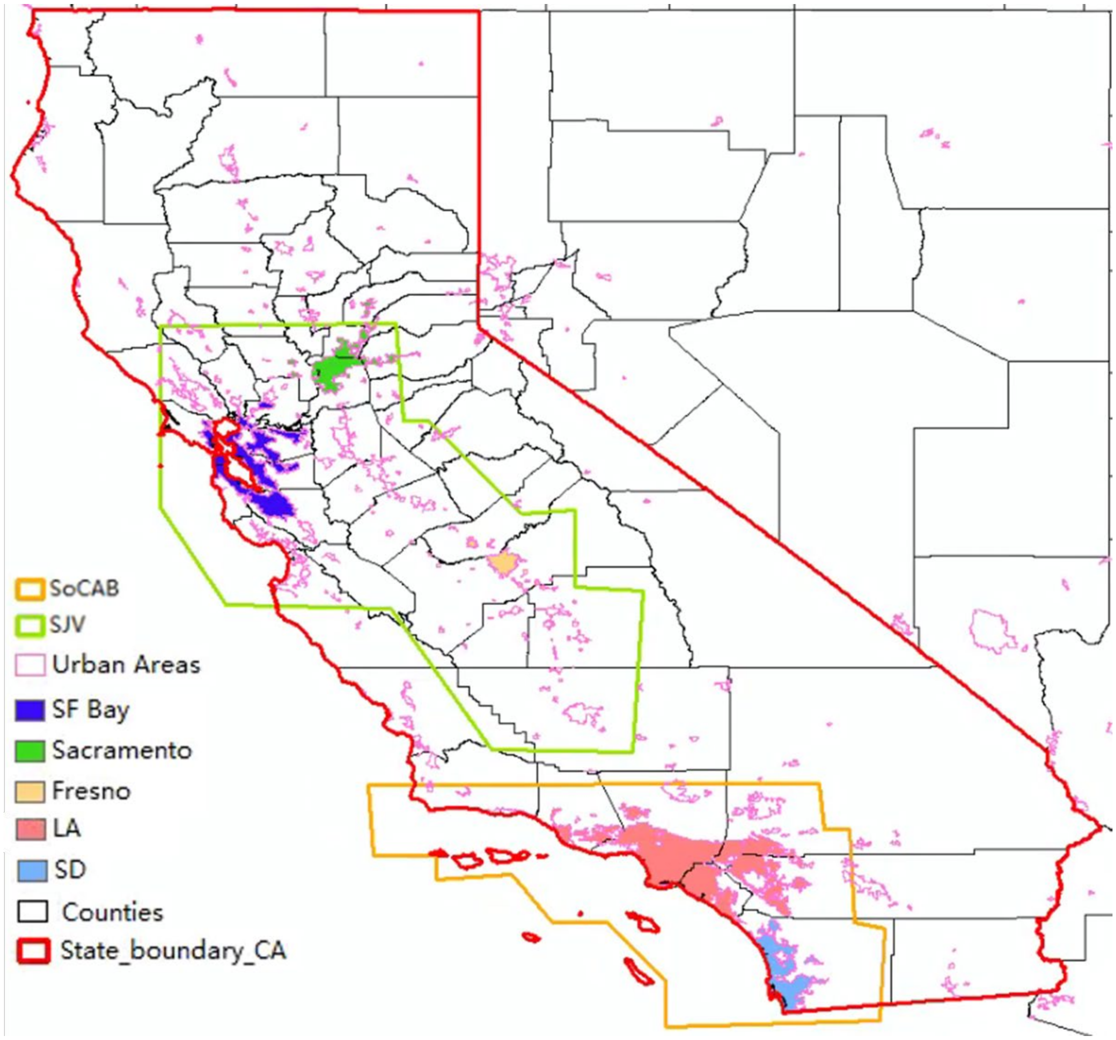

Figure 9.1. Spatial distributions of the selected decomposition regions with the five major metropolitan areas highlighted.

(1) Statistical analysis for meteorological decomposed data

We quantified the role of meteorology in explaining year-to-year variation using both the original estimates and the meteorology-normalized results. For total $PM_{2.5}$ and each component, we constructed annual time series (2000–2020) of the ground-level concentrations without decomposition (referring to $PM_{obs}(g,t)$ hereafter), the meteorology-normalized concentrations $PM_{DEMET}(g,t)$, and the meteorological contribution $PM_{MET}(g,t)$. We calculated the sample variance of $PM_{obs}(g,t)$ and $PM_{MET}(g,t)$ across years, and defined the meteorological share of variance ($VarShare$) was then calculated as:

$$VarShare\% = \frac{Var(PM_{MET}(g,t)}{Var(PM_{obs}(g,t)} \times 100$$

This ratio expresses the fraction of observed interannual variability attributable to meteorology. Values close to 0% indicate negligible meteorological influence, whereas values approaching or exceeding 100% suggest that meteorology alone accounts for most or all of the observed variability. To aid interpretation, we also reported the standard deviation (SD) of PM_{obs} (Obs_SD), PM_{DEMET}

(DEmet_SD), and PM_{MET} (Met_SD) to characterize the typical magnitude of variability, the correlation coefficient between $PM_{MET}(g,t)$ and $PM_{Obs}(g,t)$ to show how closely meteorological fluctuations track observed concentrations, and long-term means of both originally estimated (Mean_Obs) and meteorological (Mean_Met) contributions.

(2) Statistical analysis for smoke and non-fire data

To evaluate spatial hotspots following the two-step decomposition, we calculated multi-year mean concentrations of fire and non-fire $PM_{2.5}$ and its components. To assess the interannual variability across selected subregions and further characterize the contribution of emission policies, we decomposed the non-fire monthly time series into three components—long-term trend, seasonal cycle, and residuals—using a moving average (Kendall & Stuart, 1983). method. This approach enables a clearer distinction between emission-driven changes, seasonal patterns, and short-term fluctuations. Representative examples of this decomposition are shown in Figure 9.2.

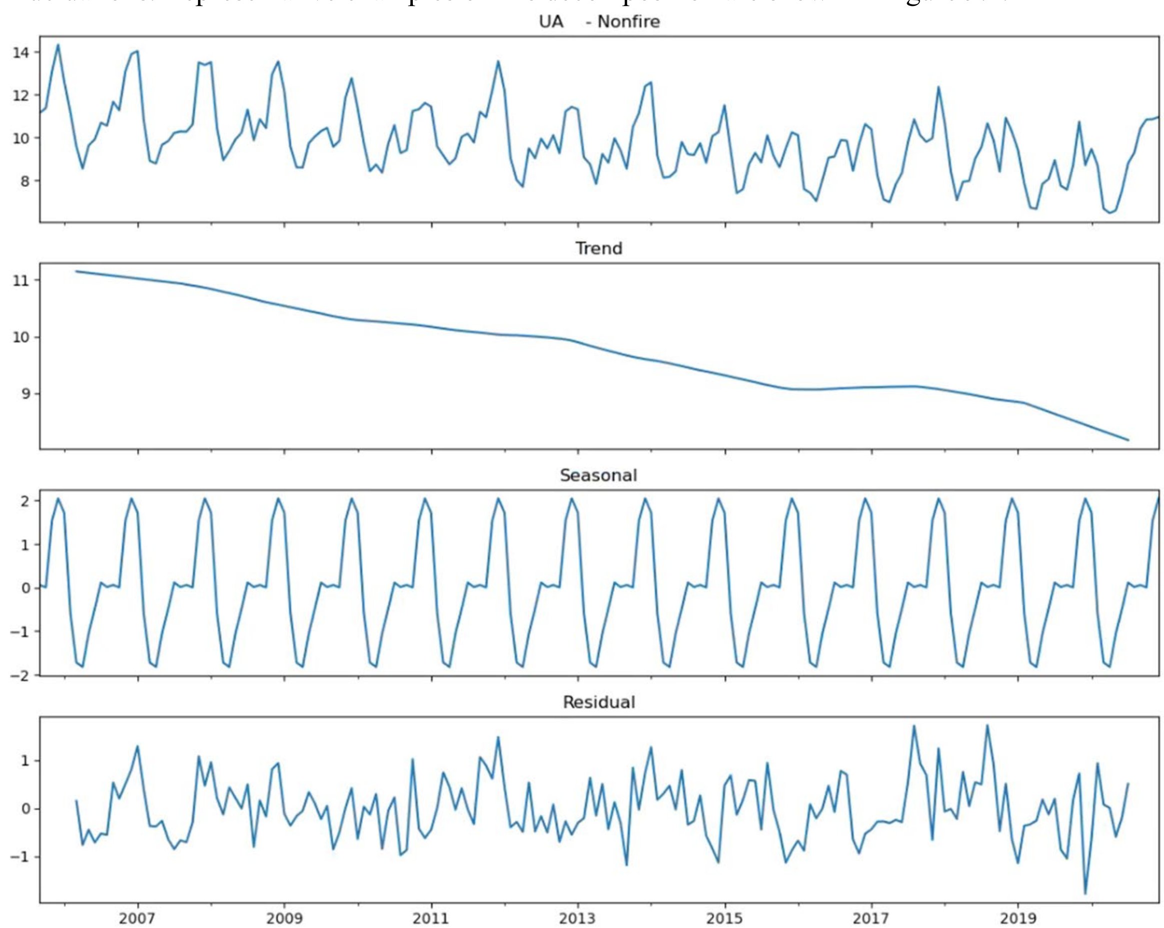

Figure 9.2. Decomposition of nitrate $PM_{2.5}$ for urban area of California.

9.3 Changing roles of meteorology on $PM_{2.5}$ and its five components trend

Table 9.1-9.12 summarized the predicted net $PM_{2.5}$ total and the five component concentrations, those with effects of meteorological changes removed and the differences in $PM_{2.5}$ and component concentrations associated with these meteorological changes (termed PM_{ori}, PM_{DEMET}, and PM_{MET}) by region and year. It is clear that the meteorological parameters show changing roles in $PM_{2.5}$ and its component long-term variation with significant regional difference.

9.3.1 $PM_{2.5}$ (total mass)

Across regions, meteorology explains a modest share of interannual $PM_{2.5}$ variability, with VarShare averaging ~8.4% and ranging from ~3% (RA, CA statewide) to ~20% (SoCAB). Regions with larger meteorological influence also show higher Met_SD (e.g., SoCAB Met_SD ≈ 0.40 μg/m³; Fresno ≈ 0.72 μg/m³), while statewide values remain small (Met_SD ≈ 0.21 μg/m³). The Obs–Met correlation is weak in magnitude and often slightly negative (e.g., SD ≈ −0.29, UA ≈ −0.22, SF Bay ≈ −0.12). Notably, the de-meteorologized variability (DEmet_SD) is equal to or modestly larger than Obs_SD in several areas (e.g., SoCAB 1.05 vs 0.91; SD 1.83 vs 1.62; LA metro 2.17 vs 2.12). Mean levels differ in intuitive ways—Mean_Obs is highest in UA and Fresno, moderate in SJV and Sacramento, and lowest statewide and in RA—while Mean_Met tends to be slightly negative in sign for many regions.

Table 9.1. The predicted original total $PM_{2.5}$, $PM_{2.5}$ with meteorological changes removed ($PM25_{DEMET}$), and the changes in $PM_{2.5}$ associated with the meteorological changes ($PM25_{MET}$) for different regions from 2000 to 2020 (unit: μg/m³). A positive $PM25_{MET}$ value (orange) indicates that the meteorological change is favorable to $PM_{2.5}$ formation, while a negative value (blue) indicates unfavorable.

Region	Indicator	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
CA	PM25	7.116	6.649	7.398	6.531	6.297	5.887	6.355	6.475	7.848	5.901	5.040	5.734	6.051	6.335	6.039	6.192	5.261	6.701	7.872	4.698	10.093
	$PM25_{DEMET}$	7.622	7.080	7.487	6.962	6.553	6.352	6.767	6.728	8.048	6.620	5.672	6.259	6.403	6.567	6.220	6.333	5.514	7.031	7.751	5.087	10.792
	$PM25_{MET}$	-0.506	-0.431	-0.089	-0.431	-0.257	-0.465	-0.412	-0.253	-0.200	-0.719	-0.632	-0.525	-0.352	-0.233	-0.181	-0.141	-0.252	-0.330	0.120	-0.389	-0.699
SJV	PM25	9.768	8.645	9.981	8.444	8.746	7.996	8.799	8.659	9.986	7.875	6.597	7.690	7.253	8.159	7.551	8.088	6.855	8.201	9.852	5.995	12.380
	$PM25_{DEMET}$	9.593	8.989	9.777	8.894	8.518	8.382	9.155	8.839	10.108	7.743	7.249	7.997	7.597	7.847	7.456	7.851	7.062	8.237	9.285	6.357	12.866
	$PM25_{MET}$	0.174	-0.344	0.204	-0.450	0.228	-0.386	-0.356	-0.180	-0.122	0.132	-0.651	-0.307	-0.344	0.312	0.095	0.236	-0.207	-0.037	0.566	-0.361	-0.486
SoCAB	PM25	8.667	9.481	8.552	9.120	7.750	7.846	7.864	8.368	8.223	7.018	6.617	7.324	7.387	7.069	7.332	6.555	6.669	7.389	7.464	5.684	7.763
	$PM25_{DEMET}$	10.284	9.869	9.946	9.385	8.694	8.322	8.350	8.592	8.656	8.148	7.323	7.898	7.749	7.511	7.412	7.056	6.848	7.709	7.668	6.275	8.579
	$PM25_{MET}$	-1.617	-0.388	-1.394	-0.265	-0.944	-0.476	-0.486	-0.223	-0.433	-1.130	-0.706	-0.574	-0.362	-0.442	-0.080	-0.501	-0.179	-0.320	-0.204	-0.591	-0.815
UA	PM25	13.104	13.755	13.186	12.470	11.632	11.193	11.001	11.432	11.740	10.286	9.325	10.371	9.582	10.063	9.705	9.929	8.828	9.852	10.804	7.746	11.484
	$PM25_{DEMET}$	13.903	13.830	13.833	12.581	12.280	11.469	11.389	11.511	11.883	10.520	10.054	10.591	9.819	9.887	9.529	9.362	8.962	9.841	10.419	8.262	12.277
	$PM25_{MET}$	-0.798	-0.075	-0.648	-0.111	-0.649	-0.276	-0.388	-0.078	-0.143	-0.234	-0.729	-0.220	-0.237	0.176	0.176	0.567	-0.135	0.010	0.385	-0.516	-0.794
RA	PM25	6.792	6.264	7.085	6.209	6.008	5.599	6.103	6.207	7.638	5.664	4.808	5.482	5.860	6.133	5.840	5.990	5.068	6.530	7.713	4.533	10.018
	$PM25_{DEMET}$	7.282	6.715	7.143	6.657	6.243	6.074	6.517	6.469	7.840	6.409	5.434	6.024	6.218	6.388	6.040	6.169	5.327	6.879	7.607	4.915	10.712
	$PM25_{MET}$	-0.490	-0.451	-0.059	-0.448	-0.236	-0.475	-0.414	-0.262	-0.203	-0.745	-0.627	-0.542	-0.358	-0.255	-0.200	-0.179	-0.259	-0.349	0.106	-0.382	-0.694
SF Bay	PM25	11.889	11.539	11.901	10.256	10.555	9.882	9.717	9.604	10.463	9.227	8.068	9.115	7.995	9.398	8.280	9.087	7.455	9.218	10.884	7.092	11.116
	$PM25_{DEMET}$	12.564	12.239	12.349	10.792	11.226	10.371	10.236	9.898	10.693	9.242	8.816	9.392	8.471	8.871	8.333	8.320	7.936	9.318	10.291	7.596	11.892
	$PM25_{MET}$	-0.675	-0.700	-0.448	-0.536	-0.672	-0.489	-0.519	-0.294	-0.231	-0.015	-0.748	-0.277	-0.476	0.527	-0.053	0.767	-0.481	-0.100	0.593	-0.503	-0.776
Sacramento	PM25	11.873	11.546	12.262	10.686	10.844	10.106	10.394	10.743	12.412	9.986	8.627	9.897	8.709	9.972	8.810	9.967	7.891	9.333	12.167	7.400	13.756
	$PM25_{DEMET}$	12.666	11.985	12.372	11.153	11.157	10.556	10.809	10.798	12.228	9.563	9.537	10.019	9.135	9.377	8.849	8.996	8.214	9.512	11.371	7.791	14.226
	$PM25_{MET}$	-0.793	-0.439	-0.110	-0.467	-0.313	-0.450	-0.414	-0.055	0.184	0.423	-0.910	-0.122	-0.426	0.595	-0.038	0.971	-0.324	-0.179	0.796	-0.391	-0.470
Fresno	PM25	17.170	16.527	18.180	15.070	14.997	14.088	14.585	16.943	16.711	14.787	13.055	15.130	12.846	14.431	13.038	14.180	11.562	13.130	14.621	9.588	17.730
	$PM25_{DEMET}$	16.479	16.148	17.103	15.128	14.254	14.024	14.090	15.742	16.401	13.126	13.302	14.041	12.490	12.618	11.473	12.102	11.076	11.900	13.031	9.862	17.902
	$PM25_{MET}$	0.691	0.379	1.077	-0.057	0.743	0.064	0.494	1.200	0.309	1.661	-0.246	1.089	0.357	1.813	1.565	2.077	0.486	1.230	1.591	-0.274	-0.173
LA	PM25	15.163	17.110	15.142	15.280	13.362	13.096	12.473	12.995	12.511	11.288	10.620	11.602	10.948	10.583	11.177	10.627	9.969	10.622	10.755	8.520	10.987
	$PM25_{DEMET}$	16.179	16.564	16.291	14.874	14.347	13.151	12.866	12.988	12.723	12.178	11.337	11.873	11.031	10.757	10.838	10.274	9.917	10.576	10.602	9.140	11.983
	$PM25_{MET}$	-1.017	0.546	-1.149	0.405	-0.985	-0.055	-0.393	0.007	-0.213	-0.890	-0.717	-0.271	-0.084	-0.174	0.339	0.352	0.052	0.046	0.154	-0.620	-0.996
SD	PM25	12.710	14.092	12.060	13.064	11.128	11.359	11.151	11.631	11.146	9.645	9.197	9.885	9.672	9.362	9.372	9.112	8.938	9.160	9.451	7.761	9.111
	$PM25_{DEMET}$	14.204	14.539	13.769	13.172	12.546	11.889	11.684	12.194	11.562	10.726	10.197	10.489	9.791	9.666	9.456	9.266	9.085	9.380	9.451	8.385	9.891
	$PM25_{MET}$	-1.493	-0.447	-1.710	-0.108	-1.418	-0.530	-0.533	-0.564	-0.416	-1.082	-1.000	-0.604	-0.119	-0.304	-0.084	-0.155	-0.147	-0.221	0.000	-0.624	-0.779

Table 9.2. PM2.5 total mass concentrations metrics for the 2000-2020 period by region.

Region	Obs_SD	Met_SD	DEmet_SD	VarShare%	Corr_Obs_Met	Mean_Obs	Mean_Met
CA	1.16	0.21	1.16	3.19	0.07	6.50	-0.35
SJV	1.41	0.32	1.38	5.20	0.22	8.45	-0.11
SoCAB	0.91	0.40	1.05	19.63	-0.15	7.63	-0.58
UA	1.51	0.38	1.64	6.16	-0.22	10.83	-0.22
RA	1.17	0.21	1.17	3.08	0.09	6.26	-0.36

SF Bay	1.39	0.44	1.51	10.16	-0.12	9.65	-0.29
Sacramento	1.61	0.49	1.62	9.30	0.12	10.35	-0.14
Fresno	2.10	0.72	2.15	11.76	0.11	14.68	0.77
LA	2.12	0.53	2.17	6.15	0.02	12.13	-0.27
SD	1.62	0.50	1.83	9.34	-0.29	10.43	-0.59

9.3.2 Sulfate

Interannual variability in SO_4^{2-} is only weakly explained by meteorology. VarShare values are uniformly small (≈0.15–0.44%), with the highest shares in Fresno (~0.44%) and SF Bay (~0.35%), and the lowest in UA (~0.15%) and SD (~0.16%). Consistent with this, Met_SD magnitudes are an order of magnitude smaller than Obs_SD in every region (e.g., CA: Met_SD ≈ 0.008 vs Obs_SD ≈ 0.177; LA: 0.029 vs 0.674). Obs–Met correlations are weakly negative across most regions (≈ −0.08 to −0.48). De-meteorologized variability (DEmet_SD) is similar to or slightly larger than Obs_SD.

Table 9.3. The predicted original sulfate $PM_{2.5}$, sulfate $PM_{2.5}$ with meteorological changes removed ($Sulfate_{DEMET}$), and the changes in sulfate $PM_{2.5}$ associated with the meteorological changes ($Sulfate_{MET}$) for different regions from 2000 to 2020 (unit: µg/m³).

Region	Indicator	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
CA	Sulfate	1.128	1.120	1.079	1.055	1.000	0.989	0.955	0.946	0.959	0.822	0.729	0.794	0.767	0.770	0.731	0.695	0.623	0.650	0.664	0.580	0.654
	$Sulfate_{DEMET}$	1.153	1.127	1.107	1.065	1.016	0.992	0.972	0.968	0.987	0.840	0.743	0.810	0.784	0.793	0.738	0.695	0.634	0.655	0.674	0.589	0.673
	$Sulfate_{MET}$	-0.024	-0.008	-0.028	-0.009	-0.016	-0.003	-0.016	-0.022	-0.028	-0.018	-0.014	-0.016	-0.017	-0.023	-0.007	0.001	-0.011	-0.006	-0.010	-0.009	-0.019
SJV	Sulfate	1.345	1.289	1.366	1.314	1.278	1.252	1.247	1.160	1.192	1.023	0.905	0.994	0.906	0.924	0.884	0.876	0.765	0.792	0.822	0.714	0.810
	$Sulfate_{DEMET}$	1.371	1.298	1.397	1.333	1.305	1.258	1.267	1.198	1.229	1.047	0.911	1.007	0.934	0.963	0.899	0.878	0.779	0.805	0.845	0.726	0.829
	$Sulfate_{MET}$	-0.026	-0.009	-0.032	-0.019	-0.027	-0.006	-0.020	-0.038	-0.038	-0.023	-0.006	-0.013	-0.028	-0.040	-0.015	-0.002	-0.014	-0.013	-0.023	-0.012	-0.019
SoCAB	Sulfate	1.851	2.005	1.658	1.725	1.529	1.582	1.479	1.481	1.401	1.198	1.054	1.115	1.062	1.013	0.990	0.876	0.871	0.897	0.899	0.787	0.934
	$Sulfate_{DEMET}$	1.903	2.019	1.689	1.737	1.539	1.580	1.502	1.503	1.448	1.221	1.085	1.150	1.087	1.043	0.995	0.871	0.879	0.896	0.896	0.791	0.979
	$Sulfate_{MET}$	-0.053	-0.014	-0.031	-0.013	-0.010	0.002	-0.024	-0.022	-0.048	-0.023	-0.031	-0.035	-0.025	-0.030	-0.005	0.005	-0.008	0.002	0.003	-0.004	-0.045
UA	Sulfate	1.985	2.097	2.003	1.976	1.834	1.826	1.706	1.675	1.599	1.397	1.192	1.240	1.121	1.095	1.061	1.008	0.934	0.950	0.968	0.861	1.031
	$Sulfate_{DEMET}$	2.041	2.124	2.043	2.001	1.848	1.833	1.743	1.705	1.657	1.430	1.229	1.287	1.159	1.146	1.075	1.012	0.948	0.965	0.978	0.872	1.071
	$Sulfate_{MET}$	-0.056	-0.028	-0.040	-0.025	-0.014	-0.007	-0.037	-0.031	-0.058	-0.032	-0.037	-0.047	-0.038	-0.051	-0.014	-0.004	-0.013	-0.015	-0.010	-0.011	-0.040
RA	Sulfate	1.082	1.067	1.029	1.006	0.955	0.944	0.915	0.906	0.925	0.791	0.704	0.770	0.748	0.752	0.713	0.678	0.607	0.634	0.647	0.565	0.634
	$Sulfate_{DEMET}$	1.104	1.073	1.056	1.014	0.971	0.947	0.930	0.928	0.951	0.808	0.717	0.784	0.764	0.774	0.720	0.677	0.617	0.639	0.658	0.574	0.652
	$Sulfate_{MET}$	-0.023	-0.007	-0.028	-0.008	-0.016	-0.003	-0.015	-0.022	-0.026	-0.017	-0.012	-0.014	-0.016	-0.022	-0.006	0.001	-0.010	-0.005	-0.011	-0.009	-0.018
SF Bay	Sulfate	1.397	1.386	1.525	1.428	1.396	1.383	1.321	1.245	1.323	1.169	1.002	1.045	0.926	0.920	0.907	0.899	0.781	0.796	0.838	0.736	0.816
	$Sulfate_{DEMET}$	1.446	1.414	1.568	1.474	1.447	1.408	1.366	1.302	1.387	1.211	1.024	1.094	0.964	0.988	0.930	0.914	0.797	0.834	0.866	0.760	0.837
	$Sulfate_{MET}$	-0.050	-0.028	-0.043	-0.045	-0.052	-0.025	-0.046	-0.056	-0.064	-0.042	-0.023	-0.050	-0.038	-0.068	-0.023	-0.015	-0.017	-0.037	-0.029	-0.023	-0.021
Sacramento	Sulfate	1.253	1.249	1.249	1.220	1.181	1.143	1.088	1.068	1.061	0.961	0.862	0.918	0.817	0.816	0.793	0.785	0.653	0.665	0.688	0.614	0.744
	$Sulfate_{DEMET}$	1.287	1.265	1.301	1.248	1.221	1.178	1.114	1.107	1.102	0.994	0.877	0.948	0.849	0.867	0.805	0.790	0.672	0.689	0.722	0.635	0.765
	$Sulfate_{MET}$	-0.034	-0.016	-0.051	-0.028	-0.040	-0.035	-0.026	-0.039	-0.041	-0.033	-0.015	-0.030	-0.032	-0.051	-0.012	-0.006	-0.018	-0.024	-0.034	-0.021	-0.021
Fresno	Sulfate	1.632	1.600	1.734	1.654	1.573	1.542	1.486	1.445	1.438	1.274	1.158	1.280	1.092	1.089	1.032	1.113	0.917	0.971	1.007	0.868	1.029
	$Sulfate_{DEMET}$	1.664	1.619	1.772	1.695	1.612	1.554	1.528	1.508	1.473	1.300	1.167	1.295	1.146	1.161	1.052	1.102	0.950	0.996	1.037	0.892	1.055
	$Sulfate_{MET}$	-0.032	-0.020	-0.038	-0.041	-0.038	-0.012	-0.042	-0.063	-0.034	-0.026	-0.009	-0.015	-0.054	-0.072	-0.021	0.011	-0.032	-0.025	-0.030	-0.024	-0.026
LA	Sulfate	2.708	2.959	2.813	2.798	2.542	2.543	2.320	2.303	2.093	1.821	1.499	1.536	1.384	1.329	1.297	1.179	1.142	1.165	1.175	1.051	1.289
	$Sulfate_{DEMET}$	2.783	2.997	2.854	2.822	2.525	2.533	2.362	2.316	2.172	1.857	1.560	1.604	1.434	1.386	1.313	1.186	1.155	1.174	1.169	1.053	1.350
	$Sulfate_{MET}$	-0.075	-0.038	-0.041	-0.023	0.017	0.009	-0.042	-0.012	-0.079	-0.036	-0.062	-0.068	-0.051	-0.057	-0.015	-0.007	-0.013	-0.008	0.005	-0.003	-0.062
SD	Sulfate	2.838	3.104	2.497	2.548	2.304	2.332	2.244	2.252	2.103	1.730	1.478	1.510	1.388	1.330	1.262	1.136	1.131	1.122	1.141	1.032	1.346
	$Sulfate_{DEMET}$	2.926	3.141	2.526	2.555	2.290	2.338	2.276	2.257	2.148	1.748	1.537	1.567	1.414	1.353	1.272	1.137	1.138	1.127	1.132	1.031	1.400
	$Sulfate_{MET}$	-0.088	-0.038	-0.029	-0.007	0.014	-0.006	-0.032	-0.005	-0.045	-0.019	-0.058	-0.057	-0.027	-0.023	-0.010	-0.001	-0.007	-0.006	0.010	0.002	-0.053

Color coding follows the scheme described in Table 9.1.

Table 9.4. Sulfate component concentrations metrics for the 2000-2020 period by region.

Region	Obs_SD	Met_SD	DEmet_SD	VarShare%	Corr_Obs_Met	Mean_Obs	Mean_Met
CA	0.18	0.01	0.18	0.20	-0.37	0.84	-0.01
Fresno	0.28	0.02	0.28	0.44	-0.14	1.28	-0.03
LA	0.67	0.03	0.68	0.19	-0.08	1.85	-0.03
RA	0.16	0.01	0.17	0.22	-0.36	0.81	-0.01
SD	0.64	0.03	0.65	0.16	-0.30	1.80	-0.02
SF Bay	0.26	0.02	0.27	0.35	-0.48	1.11	-0.04
SJV	0.22	0.01	0.22	0.25	-0.30	1.04	-0.02
Sacramento	0.22	0.01	0.23	0.30	-0.42	0.94	-0.03
SoCAB	0.37	0.02	0.37	0.22	-0.29	1.26	-0.02

UA	0.43	0.02	0.43	0.15	-0.29	1.41	-0.03

9.3.3 Nitrate

For NO_3^-, meteorology plays a modest but more noticeable role than for SO_4^{2-}, with VarShare generally between ~0.5% and ~2.3%. The largest meteorological shares occur in Fresno (~2.34%), Sacramento (~2.01%), and SJV (~1.58%). Obs–Met correlations are consistently positive and moderate (≈0.37–0.61 across regions). Despite this, Met_SD remains substantially smaller than Obs_SD (e.g., Fresno: 0.12 vs 0.77 µg/m³; LA: 0.09 vs 1.17 µg/m³), and DEmet_SD stays close to or below Obs_SD. Mean NO_3^- levels are highest in Fresno and UA, moderate in LA/SJV/SD, and lowest statewide and in RA; mean meteorological terms (Mean_Met) are slightly negative relative to the chosen reference climate.

Table 9.5. The predicted original nitrate $PM_{2.5}$, nitrate $PM_{2.5}$ with meteorological changes removed ($Nitrate_{DEMET}$), and the changes in nitrate $PM_{2.5}$ associated with the meteorological changes ($Nitrate_{MET}$) for different regions from 2000 to 2020 (unit: µg/m³).

Region	Indicator	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
CA	Nitrate	1.303	1.201	1.309	1.098	1.075	1.016	0.970	1.006	0.937	0.819	0.734	0.818	0.734	0.793	0.701	0.744	0.632	0.698	0.766	0.569	0.787
	$Nitrate_{DEMET}$	1.309	1.218	1.307	1.108	1.092	1.007	0.986	1.032	0.968	0.835	0.736	0.819	0.765	0.821	0.735	0.771	0.660	0.733	0.798	0.598	0.900
	$Nitrate_{MET}$	-0.006	-0.018	0.002	-0.009	-0.017	0.009	-0.016	-0.026	-0.031	-0.016	-0.002	-0.001	-0.031	-0.028	-0.034	-0.026	-0.028	-0.035	-0.032	-0.029	-0.113
SJV	Nitrate	2.218	1.881	2.418	1.909	2.020	1.860	1.873	1.882	1.720	1.568	1.350	1.591	1.322	1.579	1.296	1.500	1.181	1.332	1.487	1.016	1.354
	$Nitrate_{DEMET}$	2.186	1.904	2.406	1.919	2.044	1.854	1.873	1.903	1.748	1.576	1.347	1.572	1.367	1.606	1.352	1.518	1.219	1.377	1.529	1.079	1.543
	$Nitrate_{MET}$	0.032	-0.023	0.012	-0.010	-0.025	0.005	-0.001	-0.021	-0.028	-0.008	0.003	0.019	-0.045	-0.027	-0.056	-0.018	-0.038	-0.045	-0.042	-0.063	-0.189
SoCAB	Nitrate	2.608	2.863	2.396	2.301	1.911	1.961	1.765	1.855	1.592	1.385	1.367	1.406	1.303	1.207	1.201	1.104	1.072	1.134	1.189	0.953	1.265
	$Nitrate_{DEMET}$	2.631	2.839	2.345	2.298	1.926	1.902	1.799	1.903	1.684	1.423	1.362	1.424	1.359	1.275	1.258	1.164	1.122	1.190	1.225	0.974	1.452
	$Nitrate_{MET}$	-0.022	0.024	0.050	0.003	-0.015	0.060	-0.033	-0.047	-0.091	-0.038	0.005	-0.018	-0.055	-0.068	-0.057	-0.061	-0.050	-0.055	-0.036	-0.021	-0.186
UA	Nitrate	3.531	3.737	3.558	3.152	2.834	2.779	2.543	2.656	2.297	2.103	1.937	2.114	1.783	1.876	1.703	1.777	1.504	1.660	1.784	1.351	1.756
	$Nitrate_{DEMET}$	3.527	3.700	3.496	3.131	2.848	2.712	2.563	2.690	2.371	2.125	1.923	2.119	1.835	1.940	1.756	1.830	1.552	1.701	1.793	1.381	1.982
	$Nitrate_{MET}$	0.004	0.037	0.062	0.021	-0.014	0.067	-0.020	-0.034	-0.074	-0.022	0.014	-0.006	-0.053	-0.065	-0.053	-0.052	-0.048	-0.042	-0.009	-0.030	-0.227
RA	Nitrate	1.183	1.063	1.188	0.987	0.980	0.920	0.885	0.916	0.863	0.749	0.669	0.748	0.677	0.734	0.646	0.688	0.585	0.646	0.711	0.526	0.735
	$Nitrate_{DEMET}$	1.189	1.084	1.188	0.998	0.997	0.915	0.900	0.942	0.892	0.765	0.672	0.748	0.707	0.760	0.679	0.713	0.612	0.681	0.744	0.555	0.842
	$Nitrate_{MET}$	-0.007	-0.021	-0.001	-0.011	-0.017	0.006	-0.016	-0.026	-0.029	-0.016	-0.003	-0.001	-0.030	-0.026	-0.033	-0.025	-0.027	-0.035	-0.033	-0.029	-0.107
SF Bay	Nitrate	2.532	2.400	2.771	2.261	2.303	2.187	1.992	1.946	1.927	1.767	1.506	1.708	1.346	1.741	1.371	1.600	1.189	1.424	1.643	1.149	1.447
	$Nitrate_{DEMET}$	2.545	2.406	2.745	2.263	2.342	2.191	2.005	1.970	1.949	1.762	1.497	1.719	1.374	1.766	1.396	1.607	1.221	1.468	1.622	1.199	1.601
	$Nitrate_{MET}$	-0.013	-0.005	0.026	-0.002	-0.040	-0.004	-0.013	-0.024	-0.022	0.005	0.009	-0.010	-0.028	-0.025	-0.024	-0.007	-0.032	-0.043	0.022	-0.050	-0.154
Sacramento	Nitrate	2.138	1.947	2.406	1.953	1.974	1.813	1.756	1.795	1.619	1.522	1.361	1.622	1.226	1.488	1.208	1.446	1.078	1.202	1.396	1.031	1.264
	$Nitrate_{DEMET}$	2.097	1.973	2.407	1.944	2.006	1.821	1.767	1.871	1.678	1.565	1.342	1.653	1.299	1.616	1.273	1.500	1.128	1.298	1.453	1.100	1.460
	$Nitrate_{MET}$	0.041	-0.026	-0.002	0.009	-0.033	-0.007	-0.011	-0.075	-0.059	-0.043	0.019	-0.031	-0.073	-0.128	-0.064	-0.055	-0.050	-0.095	-0.056	-0.069	-0.196
Fresno	Nitrate	4.682	4.349	5.283	3.990	4.050	3.987	3.904	4.590	3.655	3.573	3.159	3.910	2.967	3.616	3.212	3.369	2.565	3.079	3.157	1.996	2.718
	$Nitrate_{DEMET}$	4.569	4.356	5.219	4.001	4.089	3.927	3.894	4.584	3.682	3.564	3.135	3.877	3.044	3.630	3.173	3.383	2.659	3.121	3.178	2.162	3.194
	$Nitrate_{MET}$	0.112	-0.007	0.064	-0.011	-0.039	0.060	0.009	0.006	-0.027	0.009	0.024	0.033	-0.077	-0.014	0.039	-0.014	-0.094	-0.042	-0.022	-0.166	-0.476
LA	Nitrate	4.897	5.586	4.841	4.499	3.745	3.748	3.320	3.473	2.847	2.572	2.463	2.580	2.244	2.061	2.054	1.991	1.792	1.927	2.045	1.625	2.101
	$Nitrate_{DEMET}$	4.892	5.473	4.729	4.453	3.741	3.606	3.353	3.519	2.990	2.617	2.445	2.590	2.322	2.170	2.139	2.084	1.851	1.966	2.046	1.614	2.386
	$Nitrate_{MET}$	0.005	0.112	0.112	0.046	0.004	0.142	-0.034	-0.046	-0.143	-0.045	0.018	-0.010	-0.079	-0.109	-0.085	-0.093	-0.059	-0.039	-0.001	0.011	-0.286
SD	Nitrate	3.972	4.327	3.426	3.471	2.875	2.865	2.746	2.889	2.511	2.131	2.062	2.096	2.003	1.855	1.749	1.720	1.635	1.744	1.838	1.507	1.955
	$Nitrate_{DEMET}$	4.028	4.324	3.366	3.423	2.873	2.810	2.820	2.936	2.580	2.147	2.054	2.123	2.013	1.892	1.814	1.799	1.708	1.769	1.853	1.526	2.167
	$Nitrate_{MET}$	-0.055	0.004	0.060	0.048	0.002	0.055	-0.073	-0.047	-0.069	-0.016	0.008	-0.028	-0.010	-0.037	-0.065	-0.079	-0.073	-0.025	-0.014	-0.019	-0.212

Color coding follows the scheme described in Table 9.1.

Table 9.6. Nitrate component concentrations metrics for the 2000-2020 period by region.

Region	Obs_SD	Met_SD	DEmet_SD	VarShare%	Corr_Obs_Met	Mean_Obs	Mean_Met
CA	0.21	0.02	0.21	1.30	0.40	0.89	-0.02
Fresno	0.77	0.12	0.71	2.34	0.60	3.61	-0.03
LA	1.17	0.09	1.12	0.65	0.61	2.97	-0.03
RA	0.19	0.02	0.18	1.46	0.37	0.81	-0.02
SD	0.81	0.06	0.79	0.53	0.37	2.45	-0.03
SF Bay	0.45	0.04	0.44	0.64	0.40	1.82	-0.02
SJV	0.35	0.04	0.33	1.58	0.53	1.64	-0.03
Sacramento	0.37	0.05	0.34	2.01	0.57	1.58	-0.05

SoCAB	0.55	0.05	0.52	0.88	0.54	1.61	-0.03
UA	0.72	0.06	0.69	0.70	0.58	2.31	-0.03

9.3.4 Organic carbon

For OC, meteorology explains a small but non-negligible fraction of interannual variability. VarShare ranges from ~0.21% (RA) to about 2.16% (SD) and 1.94% (SF Bay), with most regions below ~1.5%. Obs–Met correlations are positive (generally 0.25–0.55; statewide ≈ 0.52), consistent with OC's sensitivity to temperature/oxidant regimes and stagnation that promote secondary organic aerosol formation. Even so, Met_SD remains much smaller than Obs_SD (e.g., CA: 0.023 vs 0.484 µg/m³; LA: 0.070 vs 0.769), and DEmet_SD is slightly below Obs_SD in most regions. Mean OC is highest in Fresno and LA/UA, intermediate in SJV/Sacramento/SD/SF Bay, and lower statewide and in RA; Mean_Met is near zero to slightly negative.

Table 9.7. The predicted original OC $PM_{2.5}$, OC $PM_{2.5}$ with meteorological changes removed (OC_{DEMET}), and the changes in OC $PM_{2.5}$ associated with the meteorological changes (OC_{MET}) for different regions from 2000 to 2020 (unit: µg/m³).

Region	Indicator	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
CA	OC	1.817	1.655	2.481	2.061	1.886	1.704	1.950	1.902	2.603	1.651	1.312	1.544	1.682	1.767	1.652	1.819	1.407	2.203	2.786	1.272	3.146
	OC_{DEMET}	1.877	1.702	2.481	2.109	1.921	1.738	1.984	1.907	2.606	1.664	1.356	1.573	1.692	1.755	1.658	1.832	1.432	2.207	2.748	1.286	3.143
	OC_{MET}	-0.060	-0.047	0.001	-0.047	-0.036	-0.033	-0.034	-0.005	-0.002	-0.013	-0.044	-0.029	-0.010	0.012	-0.005	-0.013	-0.025	-0.005	0.037	-0.014	0.002
SJV	OC	2.556	2.321	3.404	2.786	2.707	2.386	2.748	2.629	3.233	2.303	1.824	2.211	2.046	2.444	2.149	2.391	1.983	2.797	3.488	1.768	4.200
	OC_{DEMET}	2.612	2.346	3.420	2.834	2.746	2.436	2.781	2.614	3.225	2.308	1.886	2.244	2.056	2.378	2.122	2.394	2.001	2.755	3.348	1.757	4.156
	OC_{MET}	-0.056	-0.025	-0.016	-0.048	-0.039	-0.050	-0.034	0.015	0.008	-0.005	-0.062	-0.033	-0.009	0.066	0.026	-0.003	-0.019	0.042	0.140	0.010	0.044
SoCAB	OC	2.035	2.122	2.536	2.676	2.236	2.209	2.391	2.570	2.396	2.035	1.768	1.946	1.920	1.806	1.783	1.613	1.624	1.909	1.894	1.367	2.287
	OC_{DEMET}	2.099	2.199	2.576	2.722	2.278	2.257	2.412	2.557	2.375	2.039	1.810	1.966	1.929	1.806	1.800	1.638	1.657	1.911	1.914	1.413	2.326
	OC_{MET}	-0.063	-0.077	-0.040	-0.046	-0.042	-0.048	-0.022	0.013	0.021	-0.004	-0.042	-0.020	-0.009	0.000	-0.017	-0.025	-0.033	-0.002	-0.020	-0.047	-0.039
UA	OC	3.346	3.407	4.241	3.905	3.674	3.478	3.634	3.760	3.847	3.302	2.829	3.158	2.900	3.042	2.768	2.807	2.437	2.906	3.194	2.186	3.421
	OC_{DEMET}	3.370	3.462	4.270	3.973	3.731	3.560	3.647	3.708	3.785	3.283	2.907	3.152	2.892	2.965	2.771	2.839	2.501	2.903	3.132	2.241	3.472
	OC_{MET}	-0.025	-0.055	-0.029	-0.068	-0.057	-0.082	-0.013	0.052	0.062	0.020	-0.077	0.006	0.008	0.076	-0.004	-0.032	-0.064	0.003	0.063	-0.055	-0.051
RA	OC	1.734	1.560	2.386	1.961	1.789	1.608	1.859	1.801	2.536	1.561	1.230	1.457	1.616	1.698	1.592	1.765	1.351	2.165	2.763	1.223	3.131
	OC_{DEMET}	1.796	1.607	2.384	2.008	1.823	1.639	1.894	1.809	2.542	1.576	1.272	1.487	1.627	1.690	1.597	1.777	1.374	2.170	2.727	1.234	3.126
	OC_{MET}	-0.062	-0.047	0.002	-0.046	-0.035	-0.031	-0.035	-0.008	-0.006	-0.015	-0.042	-0.031	-0.011	0.008	-0.006	-0.012	-0.023	-0.005	0.036	-0.011	0.005
SF Bay	OC	3.081	2.961	3.618	3.053	3.030	2.758	2.792	2.769	3.038	2.583	2.165	2.515	2.138	2.673	2.208	2.427	1.953	2.749	3.073	1.953	3.500
	OC_{DEMET}	3.095	3.002	3.605	3.098	3.057	2.812	2.814	2.735	2.979	2.572	2.245	2.490	2.162	2.539	2.200	2.440	1.997	2.671	2.880	1.972	3.480
	OC_{MET}	-0.014	-0.041	0.013	-0.045	-0.028	-0.054	-0.022	0.033	0.059	0.012	-0.079	0.025	-0.023	0.134	0.008	-0.013	-0.044	0.079	0.193	-0.020	0.020
Sacramento	OC	3.621	3.536	4.814	3.955	3.862	3.475	3.652	3.735	4.375	3.440	2.803	3.385	2.926	3.579	3.051	3.371	2.615	3.384	4.465	2.589	4.342
	OC_{DEMET}	3.602	3.523	4.812	4.092	3.950	3.594	3.731	3.711	4.372	3.453	2.945	3.407	2.954	3.447	3.025	3.404	2.693	3.401	4.220	2.615	4.382
	OC_{MET}	0.019	0.013	0.003	-0.136	-0.087	-0.119	-0.078	0.024	0.004	-0.013	-0.142	-0.022	-0.028	0.133	0.026	-0.033	-0.077	-0.017	0.245	-0.026	-0.039
Fresno	OC	4.966	4.943	5.594	4.602	4.297	3.881	4.268	4.606	4.741	3.994	3.395	3.912	3.385	3.871	3.296	3.704	3.154	3.834	4.413	2.685	5.554
	OC_{DEMET}	4.905	4.802	5.529	4.646	4.366	4.036	4.295	4.516	4.778	4.012	3.540	3.960	3.421	3.738	3.328	3.743	3.168	3.764	4.196	2.699	5.610
	OC_{MET}	0.061	0.140	0.065	-0.044	-0.069	-0.155	-0.027	0.090	-0.037	-0.017	-0.145	-0.049	-0.035	0.133	-0.032	-0.039	-0.015	0.070	0.217	-0.014	-0.056
LA	OC	3.650	3.927	4.842	4.846	4.516	4.420	4.632	4.862	4.553	4.094	3.572	3.837	3.604	3.371	3.284	3.014	2.754	2.967	2.953	2.336	3.132
	OC_{DEMET}	3.695	4.033	4.920	4.919	4.570	4.511	4.600	4.759	4.423	4.044	3.640	3.818	3.565	3.311	3.299	3.065	2.848	3.004	3.000	2.438	3.230
	OC_{MET}	-0.045	-0.106	-0.077	-0.073	-0.054	-0.091	0.032	0.102	0.129	0.050	-0.068	0.018	0.039	0.060	-0.015	-0.051	-0.094	-0.036	-0.047	-0.101	-0.097
SD	OC	3.056	3.205	3.335	3.573	3.092	3.131	3.218	3.386	3.276	2.963	2.726	2.923	2.804	2.692	2.582	2.496	2.306	2.453	2.488	2.067	2.441
	OC_{DEMET}	3.066	3.301	3.420	3.619	3.197	3.225	3.258	3.368	3.203	2.920	2.761	2.882	2.769	2.659	2.585	2.534	2.385	2.505	2.555	2.171	2.580
	OC_{MET}	-0.011	-0.095	-0.085	-0.046	-0.105	-0.095	-0.040	0.018	0.073	0.043	-0.035	0.041	0.035	0.033	-0.003	-0.038	-0.080	-0.052	-0.066	-0.105	-0.139

Color coding follows the scheme described in Table 9.1.

Table 9.8. OC component concentrations metrics for the 2000-2020 period by region.

Region	Obs_SD	Met_SD	DEmet_SD	VarShare%	Corr_Obs_Met	Mean_Obs	Mean_Met
CA	0.48	0.02	0.47	0.23	0.52	1.92	-0.02
Fresno	0.77	0.09	0.74	1.44	0.31	4.15	0.00
LA	0.77	0.07	0.75	0.83	0.36	3.77	-0.03
RA	0.50	0.02	0.48	0.21	0.55	1.85	-0.02
SD	0.41	0.06	0.39	2.16	0.25	2.87	-0.04
SF Bay	0.46	0.06	0.45	1.94	0.28	2.72	0.01
SJV	0.60	0.05	0.58	0.64	0.43	2.59	0.00
Sacramento	0.60	0.09	0.57	2.12	0.34	3.57	-0.02

SoCAB	0.34	0.02	0.34	0.51	0.06	2.05	-0.03
UA	0.51	0.05	0.50	0.93	0.12	3.25	-0.02

9.3.5 Elemental carbon

Similar to OC, interannual EC variability is also weakly influenced by meteorology. VarShare is uniformly small—typically ≤ ~2.25%, with the upper end in Fresno (~2.25%) and Sacramento (~2.06%), and ≤1% elsewhere (e.g., CA ≈ 0.10%, SJV ≈ 0.25%). Consistent with this, Met_SD is much smaller than Obs_SD across regions (e.g., LA: 0.021 vs 0.231 μg/m³; SF Bay: 0.013 vs 0.174), and DEmet_SD is essentially similar to or slightly below Obs_SD. Obs–Met correlations are near zero to modestly positive (statewide ≈ 0.02; many regions 0.19–0.61). Mean EC levels are highest in LA, UA, and Fresno, moderate in SD/SF Bay/Sacramento/SJV, and lower statewide and in RA; Mean_Met hovers near zero.

Table 9.9. The predicted original EC $PM_{2.5}$, EC $PM_{2.5}$ with meteorological changes removed (EC_{DEMET}), and the changes in EC $PM_{2.5}$ associated with the meteorological changes (EC_{MET}) for different regions from 2000 to 2020 (unit: μg/m³).

Region	Indicator	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
CA	EC	0.423	0.370	0.415	0.374	0.352	0.348	0.371	0.353	0.402	0.281	0.233	0.264	0.262	0.276	0.250	0.266	0.219	0.315	0.409	0.223	0.678
	EC_{DEMET}	0.431	0.379	0.416	0.379	0.358	0.348	0.372	0.355	0.401	0.282	0.238	0.264	0.266	0.274	0.255	0.273	0.224	0.318	0.407	0.226	0.682
	EC_{MET}	-0.008	-0.010	-0.001	-0.004	-0.007	0.000	-0.001	-0.001	0.001	-0.001	-0.005	0.001	-0.004	0.002	-0.005	-0.007	-0.005	-0.004	0.002	-0.003	-0.004
SJV	EC	0.607	0.522	0.560	0.498	0.500	0.484	0.524	0.491	0.530	0.407	0.335	0.396	0.344	0.413	0.346	0.375	0.318	0.424	0.528	0.327	0.888
	EC_{DEMET}	0.616	0.531	0.564	0.505	0.509	0.483	0.520	0.488	0.528	0.405	0.341	0.389	0.349	0.398	0.350	0.382	0.323	0.419	0.519	0.325	0.889
	EC_{MET}	-0.009	-0.009	-0.003	-0.006	-0.008	0.001	0.003	0.003	0.002	0.001	-0.006	0.007	-0.005	0.015	-0.004	-0.007	-0.005	0.004	0.009	0.002	-0.001
SoCAB	EC	0.628	0.629	0.624	0.657	0.564	0.596	0.629	0.641	0.574	0.498	0.438	0.464	0.435	0.406	0.393	0.360	0.351	0.403	0.415	0.333	0.544
	EC_{DEMET}	0.632	0.640	0.628	0.657	0.579	0.595	0.624	0.637	0.566	0.493	0.444	0.462	0.437	0.405	0.399	0.364	0.360	0.404	0.417	0.340	0.561
	EC_{MET}	-0.004	-0.011	-0.003	0.000	-0.014	0.000	0.006	0.004	0.008	0.006	-0.006	0.002	-0.003	0.001	-0.005	-0.004	-0.009	-0.002	-0.002	-0.007	-0.017
UA	EC	1.047	1.037	0.996	0.973	0.915	0.940	0.982	0.982	0.949	0.849	0.733	0.796	0.698	0.730	0.654	0.655	0.576	0.658	0.719	0.609	0.985
	EC_{DEMET}	1.041	1.045	0.996	0.972	0.933	0.934	0.961	0.966	0.925	0.832	0.742	0.778	0.699	0.705	0.660	0.659	0.593	0.654	0.709	0.611	0.998
	EC_{MET}	0.006	-0.008	0.000	0.001	-0.018	0.007	0.021	0.016	0.024	0.017	-0.009	0.018	-0.001	0.025	-0.006	-0.004	-0.017	0.004	0.010	-0.002	-0.013
RA	EC	0.389	0.333	0.383	0.342	0.321	0.316	0.338	0.319	0.372	0.251	0.206	0.235	0.238	0.251	0.228	0.245	0.199	0.296	0.392	0.202	0.661
	EC_{DEMET}	0.398	0.343	0.384	0.347	0.327	0.316	0.340	0.321	0.372	0.253	0.210	0.236	0.243	0.251	0.233	0.252	0.204	0.300	0.391	0.205	0.665
	EC_{MET}	-0.009	-0.010	-0.001	-0.005	-0.006	0.000	-0.002	-0.002	0.000	-0.002	-0.004	0.000	-0.004	0.001	-0.005	-0.007	-0.005	-0.004	0.001	-0.003	-0.003
SF Bay	EC	1.025	0.966	0.845	0.765	0.788	0.779	0.774	0.714	0.725	0.641	0.539	0.617	0.490	0.626	0.485	0.531	0.427	0.547	0.615	0.511	0.998
	EC_{DEMET}	1.013	0.964	0.831	0.762	0.782	0.756	0.750	0.703	0.702	0.630	0.544	0.593	0.500	0.591	0.495	0.537	0.441	0.536	0.593	0.507	0.990
	EC_{MET}	0.012	0.002	0.014	0.003	0.006	0.023	0.024	0.012	0.022	0.011	-0.005	0.024	-0.010	0.035	-0.010	-0.007	-0.015	0.011	0.022	0.004	0.007
Sacramento	EC	0.954	0.931	0.897	0.815	0.818	0.818	0.837	0.829	0.853	0.763	0.639	0.755	0.631	0.789	0.647	0.689	0.566	0.675	0.876	0.800	1.361
	EC_{DEMET}	0.971	0.954	0.891	0.835	0.835	0.811	0.828	0.813	0.850	0.754	0.660	0.730	0.638	0.729	0.645	0.694	0.580	0.672	0.825	0.758	1.335
	EC_{MET}	-0.017	-0.023	0.006	-0.020	-0.018	0.007	0.010	0.017	0.003	0.009	-0.021	0.025	-0.007	0.060	0.001	-0.005	-0.014	0.003	0.051	0.041	0.026
Fresno	EC	1.097	1.061	1.006	0.943	0.899	0.922	1.019	1.028	0.959	0.862	0.731	0.846	0.718	0.845	0.684	0.723	0.631	0.763	0.821	0.640	1.163
	EC_{DEMET}	1.117	1.068	1.020	0.951	0.936	0.929	0.988	0.998	0.971	0.863	0.762	0.834	0.723	0.794	0.704	0.754	0.652	0.744	0.820	0.643	1.200
	EC_{MET}	-0.020	-0.007	-0.014	-0.008	-0.037	-0.007	0.031	0.030	-0.011	-0.001	-0.031	0.012	-0.005	0.051	-0.020	-0.031	-0.021	0.019	0.001	-0.003	-0.036
LA	EC	1.264	1.301	1.319	1.343	1.210	1.275	1.359	1.390	1.301	1.167	1.013	1.073	0.962	0.904	0.873	0.818	0.734	0.793	0.817	0.713	0.987
	EC_{DEMET}	1.252	1.317	1.323	1.330	1.242	1.268	1.321	1.365	1.252	1.137	1.020	1.051	0.954	0.884	0.876	0.817	0.758	0.791	0.816	0.723	1.017
	EC_{MET}	0.011	-0.016	-0.004	0.013	-0.033	0.008	0.037	0.025	0.049	0.029	-0.007	0.022	0.008	0.020	-0.003	0.001	-0.024	0.002	0.000	-0.009	-0.030
SD	EC	1.041	1.059	0.955	1.008	0.901	0.951	0.996	1.014	0.972	0.891	0.806	0.811	0.731	0.691	0.648	0.638	0.596	0.651	0.678	0.608	0.746
	EC_{DEMET}	1.010	1.040	0.959	1.006	0.925	0.950	0.978	1.000	0.937	0.856	0.804	0.795	0.732	0.686	0.660	0.646	0.617	0.655	0.684	0.627	0.787
	EC_{MET}	0.031	0.018	-0.004	0.002	-0.024	0.002	0.018	0.014	0.036	0.035	0.002	0.016	-0.001	0.005	-0.011	-0.009	-0.022	-0.005	-0.006	-0.019	-0.041

Color coding follows the scheme described in Table 9.1.

Table 9.10. EC component concentrations metrics for the 2000-2020 period by region.

Region	Obs_SD	Met_SD	DEmet_SD	VarShare%	Corr_Obs_Met	Mean_Obs	Mean_Met
CA	0.10	0.00	0.10	0.10	0.02	0.34	0.00
Fresno	0.15	0.02	0.16	2.25	0.05	0.87	-0.01
LA	0.23	0.02	0.22	0.85	0.41	1.08	0.00
RA	0.10	0.00	0.10	0.08	0.04	0.31	0.00
SD	0.16	0.02	0.15	1.52	0.61	0.83	0.00
SF Bay	0.17	0.01	0.17	0.58	0.37	0.69	0.01
SJV	0.13	0.01	0.13	0.25	-0.03	0.47	0.00
Sacramento	0.16	0.02	0.16	2.06	0.24	0.81	0.01
SoCAB	0.11	0.01	0.11	0.34	0.19	0.50	0.00

UA	0.16	0.01	0.16	0.70	0.20	0.83	0.00

9.3.6 Mineral dust

Among the species, DUST shows the strongest meteorological imprint. VarShare is substantially higher than for SO_4^{2-}/OC/EC—reaching ~40% in Fresno, ~33% in LA, ~31% in UA, and ~29% in Sacramento/SD; even statewide (CA) and in RA the shares are ~7–8%. This aligns with dust's sensitivity to wind, precipitation, and soil moisture regimes. Obs–Met correlations are generally positive and moderate-to-strong (e.g., Fresno ≈ 0.90; SJV ≈ 0.68; SoCAB ≈ 0.66), and Met_SD can approach or exceed half of Obs_SD in high- DUST regions (e.g., Fresno 0.113 vs 0.180 µg/m³). Notably, DEmet_SD is often lower than Obs_SD. Mean DUST levels are highest in Fresno/LA/SoCAB, moderate in UA/SJV, and lower in SF Bay—a spatial pattern consistent with local resuspension, aridity, and basin meteorology.

Table 9.11. The predicted original DUST $PM_{2.5}$, DUST $PM_{2.5}$ with meteorological changes removed (DUST $_{DEMET}$), and the changes in DUST $PM_{2.5}$ associated with the meteorological changes (DUST $_{MET}$) for different regions from 2000 to 2020 (unit: µg/m³).

Region	Indicator	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
CA	DUST	1.018	1.106	1.362	1.142	1.089	0.986	1.119	1.180	1.324	1.084	0.913	1.008	1.156	1.203	1.143	1.050	0.967	1.110	1.177	0.913	1.127
	$DUST_{DEMET}$	1.093	1.147	1.389	1.194	1.144	1.069	1.136	1.182	1.299	1.099	0.977	1.068	1.145	1.180	1.125	1.081	1.002	1.118	1.189	0.953	1.120
	$DUST_{MET}$	-0.075	-0.041	-0.028	-0.052	-0.055	-0.083	-0.017	-0.002	0.025	-0.015	-0.064	-0.060	0.011	0.023	0.018	-0.031	-0.035	-0.008	-0.012	-0.040	0.006
SJV	DUST	0.752	0.859	1.071	0.956	0.957	0.883	0.970	0.983	1.067	0.910	0.752	0.810	0.908	1.010	0.971	0.899	0.816	0.901	0.987	0.858	1.118
	$DUST_{DEMET}$	0.843	0.899	1.120	1.008	1.007	0.950	1.010	0.988	1.041	0.934	0.833	0.908	0.904	0.954	0.932	0.923	0.851	0.902	0.959	0.842	1.046
	$DUST_{MET}$	-0.091	-0.040	-0.049	-0.052	-0.050	-0.067	-0.040	-0.004	0.027	-0.024	-0.081	-0.098	0.005	0.056	0.038	-0.024	-0.036	-0.002	0.028	0.016	0.072
SoCAB	DUST	0.988	1.066	1.202	1.167	1.056	1.016	1.117	1.223	1.235	1.056	0.921	1.015	1.117	1.117	1.131	0.991	1.012	1.115	1.130	0.912	1.047
	$DUST_{DEMET}$	1.064	1.159	1.254	1.243	1.126	1.131	1.139	1.207	1.208	1.071	1.021	1.081	1.122	1.102	1.130	1.030	1.057	1.119	1.176	1.002	1.086
	$DUST_{MET}$	-0.076	-0.093	-0.052	-0.076	-0.070	-0.115	-0.023	0.016	0.028	-0.015	-0.101	-0.066	-0.006	0.015	0.000	-0.039	-0.045	-0.004	-0.046	-0.089	-0.039
UA	DUST	0.919	0.990	1.076	1.019	0.966	0.909	0.984	1.008	1.029	0.939	0.801	0.858	0.911	0.967	0.936	0.896	0.850	0.923	0.946	0.847	1.077
	$DUST_{DEMET}$	1.000	1.085	1.135	1.080	1.033	0.997	1.004	1.006	1.006	0.955	0.899	0.930	0.916	0.931	0.918	0.921	0.891	0.921	0.954	0.890	1.071
	$DUST_{MET}$	-0.081	-0.096	-0.059	-0.061	-0.067	-0.088	-0.020	0.002	0.022	-0.016	-0.098	-0.072	-0.004	0.035	0.018	-0.025	-0.041	0.002	-0.008	-0.043	0.006
RA	DUST	1.023	1.112	1.377	1.149	1.096	0.990	1.126	1.189	1.340	1.092	0.919	1.016	1.169	1.215	1.154	1.058	0.973	1.120	1.190	0.917	1.130
	$DUST_{DEMET}$	1.098	1.150	1.403	1.201	1.150	1.073	1.143	1.192	1.315	1.106	0.981	1.075	1.157	1.193	1.136	1.089	1.008	1.129	1.202	0.957	1.123
	$DUST_{MET}$	-0.075	-0.038	-0.026	-0.052	-0.055	-0.083	-0.017	-0.002	0.025	-0.015	-0.062	-0.059	0.012	0.022	0.018	-0.031	-0.035	-0.009	-0.012	-0.040	0.006
SF Bay	DUST	0.623	0.703	0.766	0.695	0.669	0.585	0.604	0.588	0.637	0.565	0.503	0.534	0.535	0.620	0.575	0.553	0.494	0.581	0.593	0.555	0.703
	$DUST_{DEMET}$	0.662	0.720	0.776	0.704	0.676	0.622	0.617	0.597	0.618	0.576	0.539	0.569	0.540	0.579	0.550	0.565	0.518	0.560	0.579	0.535	0.679
	$DUST_{MET}$	-0.039	-0.017	-0.010	-0.009	-0.007	-0.036	-0.013	-0.009	0.019	-0.011	-0.036	-0.035	-0.005	0.041	0.026	-0.013	-0.024	0.021	0.014	0.021	0.024
Sacramento	DUST	0.660	0.809	0.858	0.760	0.778	0.705	0.747	0.739	0.798	0.712	0.619	0.641	0.708	0.784	0.732	0.706	0.626	0.695	0.777	0.721	1.034
	$DUST_{DEMET}$	0.770	0.858	0.886	0.820	0.809	0.756	0.774	0.755	0.773	0.739	0.690	0.719	0.700	0.720	0.703	0.724	0.666	0.704	0.739	0.688	0.950
	$DUST_{MET}$	-0.110	-0.050	-0.028	-0.060	-0.031	-0.051	-0.027	-0.016	0.025	-0.026	-0.072	-0.077	0.008	0.064	0.028	-0.018	-0.040	-0.008	0.038	0.033	0.083
Fresno	DUST	0.904	1.117	1.253	1.189	1.165	1.116	1.263	1.309	1.436	1.268	1.070	1.073	1.274	1.445	1.359	1.273	1.151	1.267	1.466	1.358	1.760
	$DUST_{DEMET}$	1.054	1.174	1.348	1.278	1.272	1.225	1.316	1.332	1.405	1.315	1.239	1.292	1.279	1.301	1.258	1.306	1.215	1.261	1.328	1.237	1.539
	$DUST_{MET}$	-0.150	-0.058	-0.095	-0.089	-0.107	-0.108	-0.053	-0.022	0.031	-0.047	-0.169	-0.218	-0.005	0.145	0.101	-0.033	-0.064	0.006	0.138	0.121	0.221
LA	DUST	1.068	1.092	1.148	1.142	1.070	1.029	1.120	1.149	1.132	1.049	0.872	0.945	0.983	0.999	1.001	0.953	0.920	0.987	0.960	0.855	1.161
	$DUST_{DEMET}$	1.158	1.262	1.240	1.212	1.153	1.136	1.132	1.135	1.102	1.058	0.992	1.011	0.991	0.977	0.997	0.977	0.963	0.986	1.005	0.960	1.203
	$DUST_{MET}$	-0.089	-0.171	-0.092	-0.070	-0.083	-0.106	-0.013	0.014	0.030	-0.009	-0.120	-0.066	-0.008	0.022	0.003	-0.024	-0.043	0.001	-0.045	-0.105	-0.042
SD	DUST	0.873	0.880	0.892	0.902	0.787	0.804	0.840	0.858	0.845	0.787	0.690	0.740	0.738	0.752	0.768	0.760	0.746	0.781	0.765	0.695	0.772
	$DUST_{DEMET}$	0.923	0.988	0.968	0.955	0.886	0.879	0.864	0.864	0.849	0.805	0.776	0.789	0.777	0.774	0.780	0.779	0.779	0.795	0.811	0.784	0.861
	$DUST_{MET}$	-0.050	-0.109	-0.076	-0.053	-0.098	-0.075	-0.024	-0.006	-0.004	-0.018	-0.086	-0.050	-0.039	-0.022	-0.012	-0.020	-0.033	-0.014	-0.045	-0.089	-0.089

Color coding follows the scheme described in Table 9.1.

Table 9.12. EC component concentrations metrics for the 2000-2020 period by region.

Region	Obs_SD	Met_SD	Adj_SD	VarShare%	Corr_Obs_Met	Mean_Obs	Mean_Met
CA	0.12	0.03	0.10	7.79	0.64	1.10	-0.03
Fresno	0.18	0.11	0.09	39.93	0.90	1.26	-0.02
LA	0.09	0.05	0.10	32.82	0.14	1.03	-0.05
RA	0.12	0.03	0.10	7.25	0.65	1.11	-0.03
SD	0.06	0.03	0.07	28.24	0.04	0.79	-0.05
SF Bay	0.07	0.02	0.07	10.91	0.21	0.60	0.00
SJV	0.10	0.05	0.08	23.14	0.68	0.93	-0.02
Sacramento	0.09	0.05	0.07	28.79	0.62	0.74	-0.02
SoCAB	0.09	0.04	0.07	21.10	0.66	1.08	-0.04

UA	0.07	0.04	0.07	31.43	0.33	0.95	-0.03

9.3.7 Summary and discussion

The meteorological decomposition alone results show clear regional contrasts in the influence of meteorology on pollutant variability (Table 9.1-9.12). Across regions, meteorological influence is highly uneven. Fresno exhibits the strongest weather imprint overall, driven by a very large VarShare for dust, indicating that winds, aridity, and resuspension account for a substantial share of its year-to-year swings; NO_3^- shows a smaller but noticeable meteorological share consistent with cool-season partitioning sensitivity. SoCAB presents the highest VarShare in total $PM_{2.5}$, with LA likewise showing elevated DUST VarShare; in both basins, SO_4^{2-}, OC, and EC have small meteorological shares, and NO_3^- is modest, implying that most interannual variation in those species is not primarily weather-driven. In the SJV as a whole and Sacramento, total $PM_{2.5}$ has a moderate meteorological imprint, DUST is again prominent, and NO_3^- is modestly weather-sensitive; SF Bay and San Diego show nontrivial meteorological contributions to total $PM_{2.5}$ and elevated DUST VarShare, while other species remain small to modest. Comparing spatial strata, urban areas display larger DUST VarShare than rural areas, whereas statewide and rural VarShare for total $PM_{2.5}$ are small, suggesting that most statewide interannual variability, once meteorology is removed, arises from non-meteorological factors. Observed–meteorology correlations reinforce these patterns: they are typically positive for NO_3^-, OC, and dust, indicating that years with meteorology conducive to secondary formation or resuspension tend to coincide with higher observations.

Although the meteorology-only decomposition indicates that weather exerts its strongest influence in Fresno and SoCAB, across pollutants meteorology explains only a small share of interannual variability in $PM_{2.5}$ total, SO_4^{2-}, NO_3^-, OC, and EC—generally on the order of ~2–10%, depending on species and region. These modest variance shares reflect that meteorological contributions often alternate in sign across years, with positive anomalies in some years offset by negative anomalies in others, yielding near-zero mean effects. In sharp contrast, DUST exhibits substantial meteorological dependence, with variance shares frequently exceeding ~30%, and especially high in Fresno and LA Metro. This pattern underscores that long-term changes in most $PM_{2.5}$ components are governed primarily by emissions and atmospheric chemistry, whereas meteorology remains a major modulator of interannual DUST variability. Because wildfire influences remain in the deweathered series at this stage, we treat these results as an intermediate diagnostic: they isolate the meteorological contribution and prepare the ground for subsequent smoke separation, after which the non-fire residual can be used to interrogate longer-term, non-meteorological drivers. These findings have important implications for trend attribution: the small meteorological variance shares (except for DUST) confirm that observed declines in SO_4^{2-}, NO_3^-, OC, and EC primarily reflect emission-control effectiveness rather than favorable weather patterns, lending confidence to policy impact assessments presented in Section 9.6.

9.4 Spatial patterns of fire-attributed $PM_{2.5}$ and its components

Figure 9.4 maps multiyear mean fire-attributed $PM_{2.5}$ and its components at 1-km resolution after meteorology and wildfire decomposition. The total smoke panel (upper left) shows two distinct hotspots: North Coast–Klamath hotspot and Central Sierra Nevada hotspot. These areas coincide with frequent large wildfires and prevailing downwind transport pathways, resulting in annual mean fire contributions exceeding 2.0 $\mu g/m^3$ in some locations. A secondary elevated belt is evident along the northern Sierra/Feather River–southern Cascades shoulder (Plumas–Butte), adjacent to the North Coast–Klamath hotspot but at comparatively lower levels (commonly between 1-5 $\mu g/m^3$). In contrast, coastal southern California and much of the immediate urban coast appear relatively less affected in the multiyear mean (typically lower than 1 $\mu g/m^3$).

Across components, OC most closely mirrors the total smoke pattern, with pronounced enhancements over those hotspots with fire-attributed concentrations larger than 1 $\mu g/m^3$. EC is generally weak but elevated spatially in the smoke signal, particularly in the northwestern wildfire hotspot. SO_4^{2-}, NO_3^-, and DUST remain low statewide in the fire-attributed means, with only faint, localized features. This compositional picture indicates that the multiyear average wildfire contribution is OC-dominated, with minor inorganic and DUST components.

Taken together, these spatial patterns highlight the concentration of wildfire smoke influence in forested northern and interior mountain regions, with downwind enhancement onto adjacent valleys, while coastal and southern urban corridors exhibit comparatively lower multiyear smoke means. This contrast is consistent with the geography of large-fire occurrence and transport pathways over the study period.

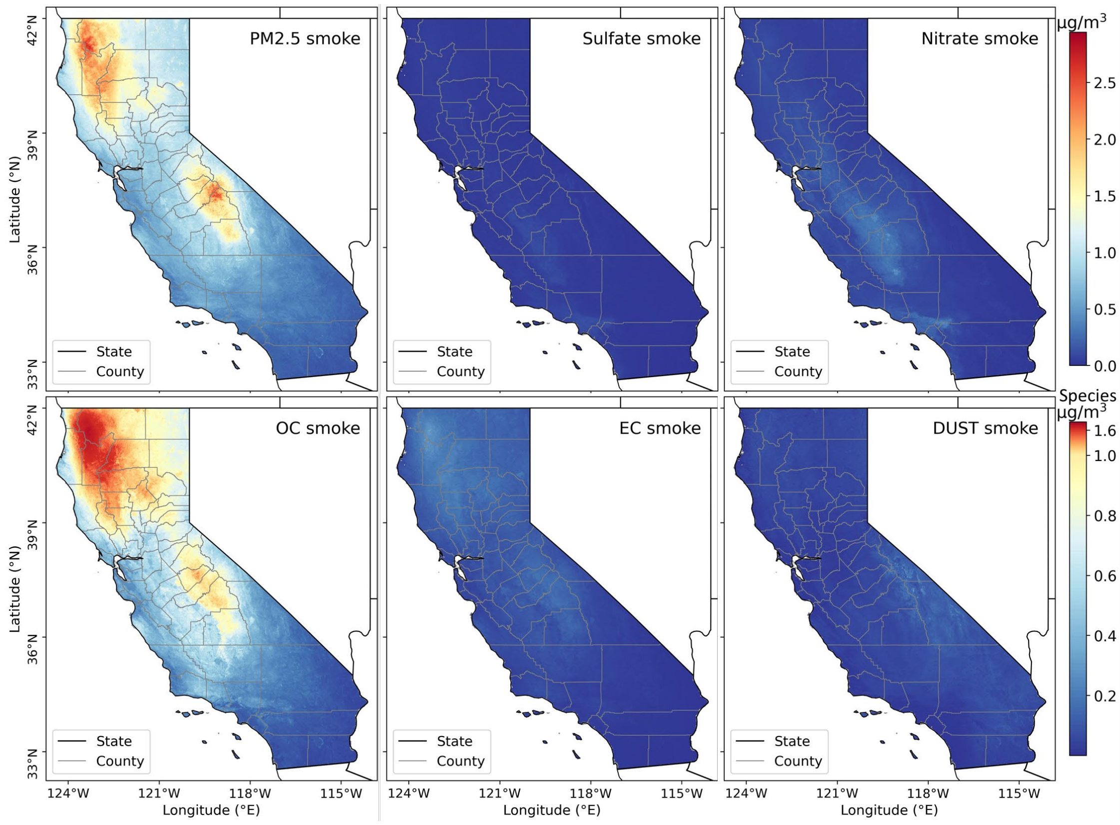

Figure 9.3. Spatial distribution of multiyear mean fire-attributed $PM_{2.5}$ and its major components (2006–2020) across California, derived from meteorology- and wildfire-decomposed daily 1-km estimates.

9.5 Spatial and temporal patterns of nonfire $PM_{2.5}$ and its components

This section examines the spatiotemporal patterns of $PM_{2.5}$ total mass and its five major components after meteorological and wildfire decomposition during 2005–2020. Figure 9.5 illustrates the spatial distribution of non-fire concentrations, revealing persistent hotspots for both total $PM_{2.5}$ and individual species. Throughout the study period, the SJV and Southern California urban regions — dominated by NO_3^- and OC — remain the primary centers of non-fire $PM_{2.5}$, while coastal, northern, and mountainous areas are comparatively clean. Figure 9.6 shows their long-term interannual evolution, characterized by substantial declines since 2005, particularly during the 2000s, though notable differences across components and regions persist. Detailed component-specific spatial contrasts and temporal trends are presented in the following subsections.

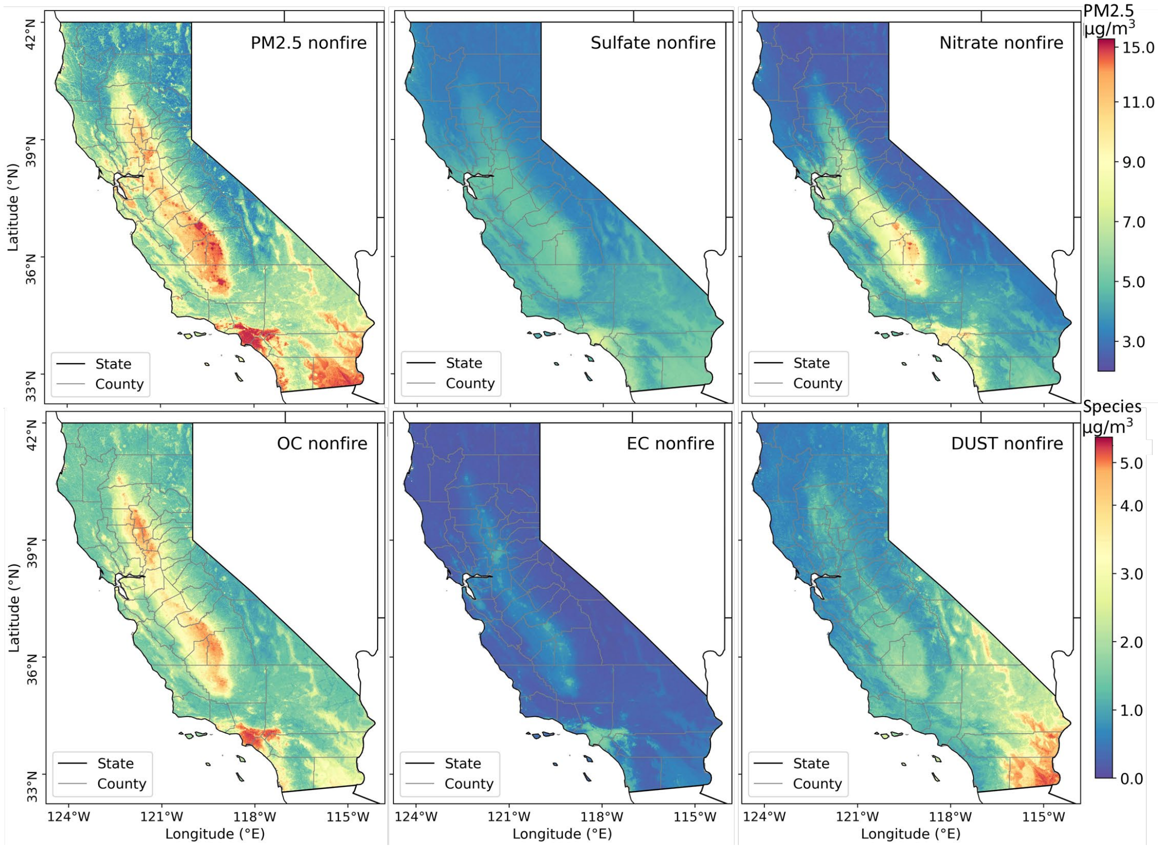

Figure 9.4. Spatial distribution of multiyear mean nonfire $PM_{2.5}$ and its major components (2006–2020) across California, derived from meteorology- and wildfire-decomposed daily estimates.

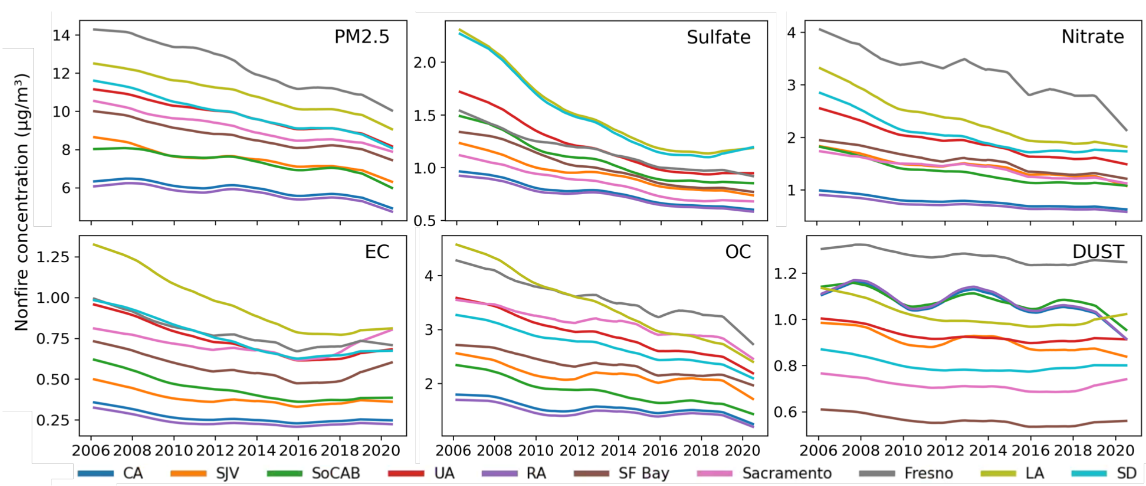

Figure 9.5. Interannual trends (2005–2020) of total $PM_{2.5}$ and its major components (SO_4^{2-}, NO_3^-, EC, OC, and DUST) across ten California regions. Trends are derived from meteorology- and wildfire-decomposed (nonfire) estimates using a moving-average method to isolate long-term changes.

9.5.1 $PM_{2.5}$ (Total)

The long-term spatial pattern (2005–2020) shows three dominant hotspots: the SoCAB urban belt, the SJV corridor, the Imperial Valley region. Values along the Fresno–Bakersfield axis and

LA/Imperial commonly fall in the ~9–14 μg/m³ range, with localized maxima across the LA urban cluster where many areas are ≳12 μg/m³. In contrast, the North/Central Coast, rural uplands, and higher elevations are typically ~3–6 μg/m³. Overall, the multi-year mean concentrations across California is 5.9 μg/m³.

Complementing this, region-aggregated interannual time series (moving-average) reveal a broad statewide decline in non-fire $PM_{2.5}$ over 2005–2020, with a sharper drop through ~2014 followed by a more gradual decrease thereafter. The broader UA also show elevated burdens, ~11.15 → ~8.17 μg/m³, while RA remain consistently lower, ~6.07 → 4.76 μg/m³. Fresno and LA Metro begin the period with the highest concentrations (near 14.28 and 12.50 μg/m³, respectively) and, despite strong decreases to ~9–10 μg/m³ by 2020, remain the most polluted regions. The SF Bay starts near ~10 μg/m³ and declines to ~7.45 μg/m³, the lowest among the five major metropolitan regions.

9.5.2 Sulfate

SO_4^{2-} shows a clear south-basin emphasis, with higher burdens in SoCAB coastal corridors. By contrast, the SJV has only modest SO_4^{2-} compared with its NO_3^- and OC burdens, and northern California remains generally clean. The LA metro and San Diego megaregions contain the broadest elevated areas, with much of each region near 1.5–2.0 μg/m³.

The interannual series indicates substantial statewide reductions from 2005 to 2020. LA metro and SD show the steepest declines (>1 μg/m³), while Sacramento decreases more gradually (≈0.4 μg/m³) and falls below ~0.7 μg/m³ by 2020—the cleanest among the five urban clusters. SoCAB exhibits modest leveling in the late 2010s, with small rebounds in LA metro and SD. By 2020, SO_4^{2-} concentrations across regions converge to ~0.5–1.0 μg/m³.

9.5.3 Nitrate

NO_3^- hotspot is predominantly concentrated in the SJV, forming the largest continuous region of NO_3^- pollution in California, with concentrations commonly exceeding 2.0 μg/m³. Among the five major urban clusters, Fresno is the most polluted, with multi-year averages around 3.2 μg/m³, followed by LA Metro (2.3 μg/m³) and San Diego (2.0 μg/m³), while Sacramento remains the cleanest urban center (1.4 μg/m³). By contrast, northern and mountainous California are comparatively clean, generally below 1 μg/m³.

The interannual trends further underscore NO_3^-'s central role in the Valley. Fresno begins the period with the highest NO_3^- concentrations in the state—around 4.1 μg/m³—and, despite a substantial decrease to ~2.1 μg/m³ by 2020, it remains the most polluted region. LA Metro also starts high (near 3.3 μg/m³) and exhibits a gradual decline with signs of plateauing in recent years, while San Diego follows a similar trajectory (~2.8 → 1.7 μg/m³). Both the twp typical basins, SJV and SoCAB, show comparable downward trends, decreasing from approximately 1.8 μg/m³ to 1.1 μg/m³ over the study period.

9.5.4 Elemental carbon

EC concentrations are generally the lowest among the five $PM_{2.5}$ components (the statewide multi-year mean = 0.3 μg/m³), with the highest levels observed in Southern California — particularly within the Los Angeles Basin, where heavy traffic, diesel emissions, and industrial activities are dominant sources. Secondary hotspots are evident in major urban centers such as SD, Fresno, Sacramento, and Bakersfield, while northern, mountainous, and coastal regions of the state remain comparatively clean.

Long-term trends reveal a pronounced statewide decline in EC between 2005 and 2020. In the LA megaregion, concentrations decreased from approximately 1.32 μg/m³ in 2005 to about 0.8 μg/m³ by 2020, while San Diego experienced a reduction from ~0.9 μg/m³ to ~0.7 μg/m³ over the same period. All five major megaregions, however, exhibit signs of a rebound in recent years, particularly in the Sacramento and San Francisco Bay areas. On average, urban EC levels fell substantially from around 1.0 μg/m³ to 0.7 μg/m³, whereas rural areas started lower (~0.3 μg/m³) and declined more modestly to ~0.2 μg/m³.

9.5.5 Organic carbon

OC emerges as the dominant $PM_{2.5}$ component in California (statewide multi-year mean = 1.5 μg/m³), particular for the SJV and Southern California, with the highest levels observed in the LA Basin, where concentrations commonly exceed 4 μg/m³. In contrast, northern, mountainous, and coastal California remain comparatively clean.

The interannual trends reveal substantial statewide declines in OC from 2005 to 2020, although with greater variability than observed for EC. The LA urban area begins the period with the highest concentrations, near 4.6 μg/m³, and despite decreasing to ~2.4 μg/m³ by 2020, it ranks as the third most polluted among the urban clusters. Fresno and Sacramento exhibit similar declining trajectories, starting from ~4.3 μg/m³ and 3.5 μg/m³, respectively, with smaller decreases observed between 2012 and 2018. San Diego and the SF Bay Area begin between 2.7 and 3.3 μg/m³ and fall to around 2.0 μg/m³ by 2020. Rural areas remain consistently lower, decreasing slightly from ~1.7 to 1.2 μg/m³ over the study period.

9.5.6 Mineral dust

Spatial hotspots are largely confined to the southeastern desert regions near the Arizona border, where concentrations exceed 3 μg/m³, while SJV is only moderately elevated, with values commonly around 1.2 μg/m³. Urban centers such as SD, Sacramento, and the SF Bay Area remain comparatively clean, with DUST contributing only a minor share of total $PM_{2.5}$ (multi-year mean

<1 μg/m³). Notably, DUST is the only component for which concentrations in rural areas are higher than in urban areas (1.1 vs. 0.9 μg/m³ on a multi-year mean basis).

The interannual trends reinforce this picture, showing consistently low baseline levels statewide, typically between 0.9 and 1.1 μg/m³. Unlike other components that exhibit clear decreasing trajectories, DUST concentrations show no significant long-term increase or decrease across any of the 10 focused regions. Among the five urban clusters, Fresno records the highest DUST levels, fluctuating from about 1.3 μg/m³ in 2005 to around 1.2 μg/m³ in recent years.

9.5.7 Summary and discussion

Overall, the long-term meteorology- and wildfire-decomposed results reaffirm well-established $PM_{2.5}$ patterns in California while offering new insights into component-specific differences. The identification of the SJV and Southern California as persistent $PM_{2.5}$ hotspots is consistent with decades of monitoring and prior studies, yet the decomposition clarifies the relative roles of individual species. SO_4^{2-} concentrations remain slightly higher in the SoCAB than in the SJV (1.07 vs 0.93 μg/m³), reflecting stronger influences from port activity and industrial emissions. EC is most prominent in the Los Angeles megaregion, consistent with its combustion-heavy source mix, but both SO_4^{2-} and EC exhibit plateaus or rebounds after 2016–2018 in Sacramento and the San Francisco Bay Area. NO_3^- continues to dominate in Fresno, underscoring persistent secondary formation in the Valley; despite substantial declines during the study period, the anthropogenic residuals indicate that NO_3^- remains a major contributor to total $PM_{2.5}$ in most urban areas as well as in the SJV and SoCAB basins. Although OC concentrations peak in the Los Angeles metropolitan area, the San Joaquin Valley also shows pronounced and persistent OC hotspots—an under-recognized feature given that the region's $PM_{2.5}$ pollution has traditionally been attributed mainly to agricultural and meteorological influences. Finally, DUST, while often highlighted in regulatory discussions, emerges as a relatively minor contributor across most populated regions. Together, these findings provide a more refined understanding of regional $PM_{2.5}$ composition, revealing distinct chemical "fingerprints" associated with different source environments and reaffirming the dominant role of anthropogenic emissions in shaping long-term air-quality burdens across California. Importantly, because meteorological and wildfire influences have been removed, the observed challenges—such as the SO_4^{2-} and EC plateaus or rebounds after 2016–2018 in Sacramento and the SF Bay Area, and the persistent OC hotspot in the SJV—directly reflect anthropogenic emission dynamics and warrant closer examination of emerging or insufficiently regulated sources.

9.6 Implications for Emission-Control Policy

The "non-fire, de-meteorologized" concentrations represent the anthropogenic or policy-sensitive residual, so they can indeed serve as the empirical basis for discussing emission-control implications. In general, those results clarify that the sustained long-term declines in $PM_{2.5}$ and its five major components across California are primarily attributable to reductions in anthropogenic emissions rather than favorable weather or changes in wildfire frequency. These findings offer several policy-relevant insights into source attribution and future control priorities.

9.6.1 Reinforcing the effectiveness—and emerging limits—of past emission controls

The decomposition results confirm that the sharp declines in SO_4^{2-} and EC through the mid-2010s primarily reflect the success of long-standing combustion and fuel-quality regulations. SO_4^{2-} reductions trace directly to the statewide transition to ultra-low-sulfur fuels, refinery desulfurization, and tighter industrial sulfur dioxide controls, while EC decreases follow diesel-engine retrofits, fleet turnover, and cleaner fuels in urban basins.

However, the non-fire interannual trends reveal a recent plateau and mild rebound of both species in several urban clusters—particularly Los Angeles, San Diego, Sacramento, and the San Francisco Bay Area—after years of steady decline. These reversals suggest that early emission gains are nearing saturation and that new or unregulated source sectors (e.g., port freight, construction machinery, off-road diesel, localized industrial activity) may now dominate the remaining EC and SO_4^{2-} burden.

Maintaining progress therefore requires next-generation control strategies: continued diesel-fleet electrification, expanded zero-emission freight and drayage operations, refinery process optimization, and port-area sulfur and EC surveillance. Reinforcing these measures would prevent the observed rebounds from eroding two decades of air-quality gains and ensure sustained improvements in California's most densely populated corridors.

9.6.2 Nitrate as a persistent hotspot in the San Joaquin Valley

After removing meteorological and wildfire influences, NO_3^- remains the most spatially concentrated $PM_{2.5}$ component in California, forming the largest and most continuous hotspot along the Fresno–Bakersfield corridor of the SJV, with typical non-fire concentrations exceeding 2 $\mu g/m^3$ and localized maxima above 3 $\mu g/m^3$. This pattern reflects strong secondary NO_3^- formation in a basin characterized by high ammonia emissions from agriculture and confined livestock operations, coupled with stagnant winter meteorology that favors ammonium NO_3^- partitioning.

While long-term NO_3^- levels have declined since 2005—particularly during the 2000s—the persistence of these hotspots indicates that current ammonia mitigation remains insufficient to complement ongoing NO_x reductions from mobile-source and industrial controls. Further progress therefore depends on coordinated precursor management, including precision fertilizer use, improved manure handling, and emission-reduction incentives for dairies, alongside continued NO_x control through transportation electrification and industrial combustion efficiency. Strengthening ammonia monitoring and incorporating agricultural NH_3 inventories into regional attainment planning would directly address the Valley's remaining NO_3^- burden.

9.6.3 Organic carbon as the dominant contributor in urban and valley regions

OC emerges as the dominant $PM_{2.5}$ component in both the SoCAB and the SJV, with concentrations commonly > 4 μg/m³ in the Los Angeles Basin and > 3–4 μg/m³ across the central Valley. These levels exceed those of NO_3^-, SO_4^{2-}, or EC, underscoring the substantial contribution of both primary combustion emissions and secondary organic aerosol formation from volatile organic compound (VOC) precursors.

Statewide OC has declined markedly since 2005, but its sustained prominence in SoCAB and the SJV points to the need for integrated control of combustion and VOC sources. Strengthened measures targeting residential wood combustion, small-scale industrial processes, and evaporative VOC emissions from consumer products, coatings, and oil-and-gas activities could further reduce OC levels. Policies that jointly mitigate VOCs and NO_x would deliver co-benefits for both $PM_{2.5}$ and ozone, helping to lower chronic exposure in California's most populous and pollution-burdened basins.

9.6.4 DUST management under changing climate conditions

Although DUST contributes only a modest share to statewide $PM_{2.5}$, its strong meteorological dependence and localized prominence—particularly in Fresno, the southern SJV, and the desert regions near the Arizona border—make it a growing concern under a warming, drier climate. The decomposition results show that DUST is the only component for which meteorology explains a substantial fraction of variability, indicating high sensitivity to wind, soil moisture, and land disturbance.

Effective mitigation therefore hinges on adaptive land and soil management rather than traditional emission controls. Priority measures include maintaining soil moisture on fallow farmland, stabilizing unpaved roads, implementing windbreaks and vegetative buffers, and enforcing construction-site DUST suppression. Integrating these measures with California's drought and land-conservation policies will help limit DUST resuspension and prevent climate-driven degradation of baseline air quality in agricultural and desert regions.

9.6.5 Integrating wildfire and anthropogenic strategies

The decomposition analysis highlights that wildfire smoke increasingly dominates episodic extremes, while the anthropogenic residual defines the chronic baseline exposure that determines long-term health risk. OC and EC in particular exhibit distinct "dual-source" behavior—wildfire-driven spikes superimposed on persistent urban and valley backgrounds—underscoring the need for coordinated management across sectors.

Sustaining progress thus requires policies that couple forest-fuel management and prescribed burning with continued urban and industrial emission reductions. Strategic alignment between CalFire, CARB, and regional air districts can ensure that fuel treatments and smoke management plans are optimized to minimize public exposure without undermining long-term air-quality gains. Enhanced satellite-based smoke monitoring and predictive modeling should be institutionalized to guide real-time public-health advisories and air-quality episode response.

9.6.6 Regional prioritization and equity considerations

The non-fire results reaffirm that the SJV and Southern California remain the most persistent $PM_{2.5}$ hotspots and coincide with areas of high social and environmental vulnerability. These regions face overlapping burdens: NO_3^- and OC dominance in the SJV, EC and OC in SoCAB, and recurring DUST contributions in the southern interior.

Future emission-control policy should therefore adopt a regionally differentiated and equity-centered approach. Priorities include enhancing ambient monitoring networks in disadvantaged communities, linking clean-air and clean-energy investments to local job creation, and expanding incentive programs for low-income residents to transition to cleaner technologies (e.g., electric vehicles, residential heating upgrades). Integrating exposure disparity metrics—such as CalEnviroScreen percentiles—into attainment planning will help ensure that emission reductions translate into tangible, equitable improvements in public health.

Although this study focuses on the five major $PM_{2.5}$ components, the decomposed statewide means indicate a residual ~1.5 µg/m³ unaccounted for by SO_4^{2-}, NO_3^-, OC, EC, and dust, relative to the total $PM_{2.5}$ mean of ~5.9 µg/m³. This fraction likely reflects other minor or uncharacterized species—such as ammonium, trace metals, sea salt, and secondary organics—not explicitly resolved in the current dataset. While individually small, these components collectively contribute to overall mass closure and may play localized roles near industrial or coastal environments. Future extensions of this work could integrate these minor fractions to refine compositional closure and improve source attribution at finer scales.

9.7 Summary

This section disentangled the respective roles of meteorology, wildfire smoke, and anthropogenic emissions in shaping two decades of $PM_{2.5}$ and component variability across California, providing a refined perspective on the sources of long-term air-quality change.

The meteorological normalization analysis demonstrated that year-to-year weather variability explains only a small share (typically < 10 %) of total $PM_{2.5}$, SO_4^{2-}, NO_3^-, OC, and EC fluctuations, confirming that most interannual and decadal trends are not meteorology-driven. The sole exception is dust, whose variability is strongly coupled to wind, precipitation, and soil-moisture conditions, with variance shares exceeding 30 – 40 % in the SJV and Southern California. These results establish that the dominant long-term declines observed statewide primarily reflect emission-control and policy impacts rather than favorable meteorology.

After removing meteorological and wildfire influences, the residual fields revealed clear long-term patterns of anthropogenic influence. Persistent hotspots were observed in the SJV and Southern California, where OC and NO_3^- were the dominant components in both regions. Substantial declines occurred across all regions and species, with the steepest reductions during the mid-2000s and slower improvement thereafter. SO_4^{2-} and EC exhibited strong decreases in Southern California, reflecting the effectiveness of sulfur-content regulations and diesel-emission controls; however, SO_4^{2-} began to plateau or slightly rebound after 2016, and EC showed signs of rebound after 2018, particularly in the Sacramento and SF Bay areas. OC decreased markedly statewide—from about 4–5 μg/m³ to ~2 μg/m³ in major basins—yet remained the largest contributor to total $PM_{2.5}$, especially in the LA Basin and SJV, where both primary combustion and secondary organic aerosol formation persist. NO_3^- declined significantly in later 2000s and plateaued in recent years across most urban regions, underscoring the difficulty of controlling secondary formation in ammonia-rich and VOC-rich environments such as the Valley. DUST exhibited region-specific persistence, with relatively high levels in the Fresno urban cluster and southern desert areas.

Together, these findings demonstrate substantial statewide improvements in air quality driven by past emission controls, yet persistent and emerging challenges remain. Despite major reductions in SO_4^{2-} and EC, both show signs of rebound in recent years, while NO_3^- and OC continue to dominate in key regions, DUST remains regionally persistent, and wildfire smoke has become an increasing episodic threat. Sustaining progress will require continued attention to these re-emerging and persistent sources to secure lasting and equitable air-quality improvements across California.

10. References

Abatzoglou, J.T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology, Vol.33*, 121-131

Amini, H., Danesh-Yazdi, M., Di, Q., Requia, W., Wei, Y., Abu-Awad, Y., Shi, L., Franklin, M., Kang, C.-M., & Wolfson, J. (2022). Hyperlocal super-learned PM2. 5 components across the contiguous US

Appel, K.W., Napelenok, S.L., Foley, K.M., Pye, H.O.T., Hogrefe, C., Luecken, D.J., Bash, J.O., Roselle, S.J., Pleim, J.E., Foroutan, H., Hutzell, W.T., Pouliot, G.A., Sarwar, G., Fahey, K.M., Gantt, B., Gilliam, R.C., Heath, N.K., Kang, D., Mathur, R., Schwede, D.B., Spero, T.L., Wong, D.C., & Young, J.O. (2017). Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1. *Geosci. Model Dev., 10*, 1703-1732

Aryal, Y., & Evans, S. (2022). Decreasing Trends in the Western US Dust Intensity With Rareness of Heavy Dust Events. *Journal of Geophysical Research: Atmospheres, 127*, e2021JD036163

CA, P., M, E., & DW, D. (2009). Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine, Vol.360*, 376-386

Carlton, A.G., Wiedinmyer, C., & Kroll, J.H. (2009). A review of Secondary Organic Aerosol (SOA) formation from isoprene. *Atmospheric Chemistry and Physics, 9*, 4987-5005

CARB (2021). Camp fire air quality data analysis. *https://ww2.arb.ca.gov/resources/documents/camp-fire-air-quality-data-analysis (Access on March 23 2025)*

Casey, J.A., Kioumourtzoglou, M.-A., Padula, A., González, D.J.X., Elser, H., Aguilera, R., Northrop, A.J., Tartof, S.Y., Mayeda, E.R., Braun, D., Dominici, F., Eisen, E.A., Morello-Frosch, R., & Benmarhnia, T. (2024). Measuring long-term exposure to wildfire $PM_{2.5}$ in California: Time-varying inequities in environmental burden. *Proceedings of the National Academy of Sciences, Vol.121*, e2306729121

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794)

Chau, K., Franklin, M., & Gauderman, W.J. (2020). Satellite-Derived $PM_{2.5}$ Composition and Its Differential Effect on Children's Lung Function. *Remote Sensing, Vol.12*, 1028

Cohen, A., Brauer, M., Burnett, R., Anderson, H., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., & Dandona, R. (2017). 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 (London, England), Vol.389*, 1907-1918

Colleen E Reid, E.M.C., Melissa M Maestas, Gina Li (2021). Daily $PM_{2.5}$ concentration estimates by county, ZIP code, and census tract in 11 western states 2008-2018. *Scientific data, Vol.8*, 112

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L., & Schwartz, J. (2019). An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. *Environment international, Vol.130*, 104909

Di, Q., Koutrakis, P., & Schwartz, J. (2016). A hybrid prediction model for $PM_{2.5}$ mass and components using a chemical transport model and land use regression. *Atmospheric Environment, Vol.131*, 390-399

Dimple Pruthi, Q.Z., Wenhao Wang, Yang Liu (2024). Multiresolution Analysis of HRRR Meteorological Parameters and GOES-R AOD for Hourly $PM_{2.5}$ Prediction. *Environmental Science & Technology, Vol.58*, 20040-20048

Ding, K., Huang, X., Ding, A., Wang, M., Su, H., Kerminen, V.-M., Petäjä, T., Tan, Z., Wang, Z., Zhou, D., Sun, J., Liao, H., Wang, H., Carslaw, K.h.o.o.-.-.-X., Wood, R., Zuidema, P., Rosenfeld, D., Kulmala, M., Fu, C., Pöschl, U., Cheng, Y., & Andreae, M. (2021). Aerosol-boundary-layer-monsoon interactions amplify semi-direct effect of biomass smoke on low cloud formation in Southeast Asia. *Nature Communications, Vol.12*, 6416

Donkelaar, A.v., Martin, R.V., Ford, B., Li, C., Pappin, A.J., Shen, S., & Zhang, D. (2024). North American Fine Particulate Matter Chemical Composition for 2000–2022 from Satellites, Models, and Monitors: The Changing Contribution of Wildfires. *ACS ES&T Air, Vol.1*, 1589-1600

Donkelaar, A.v., Martin, R.V., Li, C., & Burnett, R.T. (2019). Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science & Technology, Vol.53*, 2595-2611

Faraz Enayati Ahangar, P.P., Sina Hasheminassab, Scott A. Epstein, ... Jason Low (2021). Long-term trends of $PM_{2.5}$ and its carbon content in the South Coast Air Basin: A focus on the impact of wildfires. *Atmospheric Environment, Vol.255*, 118431

Franklin, M., Kalashnikova, O.V., & Garay, M.J. (2017). Size-resolved particulate matter concentrations derived from 4.4 km-resolution size-fractionated Multi-angle Imaging

SpectroRadiometer (MISR) aerosol optical depth over Southern California. *Remote Sensing of Environment, Vol.196*, 312-323

Geng, G., Meng, X., He, K., & Liu, Y. (2020). Random forest models for $PM_{2.5}$ speciation concentrations using MISR fractional AODs. *Environmental Research Letters, 15*, 034056

Geng, G., Zhang, Q., Tong, D., Li, M., Zheng, Y., Wang, S., & He, K. (2017). Chemical composition of ambient PM 2. 5 over China and relationship to precursor emissions during 2005–2012. *Atmospheric Chemistry and Physics, 17*, 9187-9203

Gibson, M.D., Kundu, S., & Satish, M. (2013). Dispersion model evaluation of PM2. 5, NOx and SO2 from point and major line sources in Nova Scotia, Canada using AERMOD Gaussian plume air dispersion model. *Atmospheric Pollution Research, 4*, 157-167

Guo, H., Liu, J., Froyd, K.D., Roberts, J.M., Veres, P.R., Hayes, P.L., Jimenez, J.L., Nenes, A., & Weber, R.J. (2017). Fine particle pH and gas–particle phase partitioning of inorganic species in Pasadena, California, during the 2010 CalNex campaign. *Atmos. Chem. Phys., 17*, 5703-5719

H Orru, K.L.E., B Forsberg (2017). The Interplay of Climate Change and Air Pollution on Health. *Current Environmental Health Reports, Vol.4*, 504-513

Hand, J.L., Schichtel, B.A., & Malm, W.C. (2014). Spatial and seasonal patterns in urban influence on regional concentrations of speciated aerosols across the United States. *JOURNAL OF GEOPHYSICAL RESEARCH-ATMOSPHERES, Vol.119*, 12832-12849

Hand, J.L., Schichtel, B.A., Pitchford, M., Malm, W.C., & Frank, N.H. (2012). Seasonal composition of remote and urban fine particulate matter in the United States. *Journal of Geophysical Research:Atmospheres, Vol.117*, 0

He, Q., Wang, W., Song, Y., Zhang, M., & Huang, B. (2023a). Spatiotemporal high-resolution imputation modeling of aerosol optical depth for investigating its full-coverage variation in China from 2003 to 2020. *Atmospheric Research, Vol.281*, 106481(106481-106415)

He, Q., Ye, T., Wang, W., Luo, M., Song, Y., & Zhang, M. (2023b). Spatiotemporally continuous estimates of daily 1-km $PM_{2.5}$ concentrations and their long-term exposure in China from 2000 to 2020. *Journal of Environmental Management, Vol.342*, 118145

He, Q., Zhang, M., Song, Y., & Huang, B. (2021). Spatiotemporal assessment of PM2. 5 concentrations and exposure in China from 2013 to 2017 using satellite-derived data. *Journal of Cleaner Production, 286*, 124965

Hoogh, K.d., Wang, M., Adam, M., Badaloni, C., Beelen, R., Birk, M., Cesaroni, G., Cirach, M., Declercq, C., Dėdelė, A., Dons, E., Nazelle, A.d., Eeftens, M., Eriksen, K., Eriksson, C., Fischer, P., Gražulevičienė, R., Gryparis, A., Hoffmann, B., Jerrett, M., Katsouyanni, K., Iakovides, M., Lanki, T., Lindley, S., Madsen, C., Mölter, A., Mosler, G., Nádor, G., Nieuwenhuijsen, M., Pershagen, G., Peters, A., Phuleria, H., Probst-Hensch, N., Raaschou-Nielsen, O., Quass, U., Ranzi, A., Stephanou, E., Sugiri, D., Schwarze, P., Tsai, M.-Y., Yli-Tuomi, T., Varró, M.J., Vienneau, D., Weinmayr, G., Brunekreef, B., & Hoek, G. (2013). Development of Land Use Regression Models for Particle Composition in Twenty Study Areas in Europe. *Environmental Science & Technology, Vol.47*, 5778-5786

Hsu, C.-Y., Wu, C.-D., Hsiao, Y.-P., Chen, Y.-C., Chen, M.-J., & Lung, S.-C.C. (2018). Developing Land-Use Regression Models to Estimate $PM_{2.5}$-Bound Compound Concentrations. *Remote Sensing, Vol.10*, 1971

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. *Environmental Science & Technology, Vol.51*, 6936-6944

Kendall, M. G. & Stuart, A. The Advanced Theory of Statistics Vol. 3, 410–414 (Macmillan, 1983).

Kotchenruther, R.A. (2020). Recent changes in winter $PM_{2.5}$ contributions from wood smoke, motor vehicles, and other sources in the Northwest U.S. *Atmospheric Environment, 237*, 117724

Krall, J., Mulholland, J., Russell, A., Balachandran, S., Winquist, A., Tolbert, P., Waller, L., & Sarnat, S. (2017). Associations between Source-Specific Fine Particulate Matter and Emergency Department Visits for Respiratory Disease in Four US Cities. *Environmental Health Perspectives, Vol.125*, 97-103

Kumar, N., Park, R.J., Jeong, J.I., Woo, J.-H., Kim, Y., Johnson, J., Yarwood, G., Kang, S., Chun, S., & Knipping, E. (2021). Contributions of international sources to $PM_{2.5}$ in South Korea. *Atmospheric Environment, 261*, 118542

Li, L., Franklin, M., Girguis, M., Lurmann, F., Wu, J., Pavlovic, N., Breton, C., Gilliland, F., & Habre, R. (2020). Spatiotemporal Imputation of MAIAC AOD Using Deep Learning with Downscaling. *Remote Sens Environ, Vol.237*, 111584

Liu, S., Geng, G., Xiao, Q., Zheng, Y., Liu, X., Cheng, J., & Zhang, Q. (2022). Tracking Daily Concentrations of $PM_{2.5}$ Chemical Composition in China since 2000. *Environmental Science & Technology, 56*, 16517-16527

Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems, 30*

Mailloux, N.A., Abel, D.W., Holloway, T., & Patz, J.A. (2022). Nationwide and Regional $PM_{2.5}$-Related Air Quality Health Benefits From the Removal of Energy-Related Emissions in the United States. *GeoHealth, 6*, e2022GH000603

Malm, W.C., Schichtel, B.A., & Pitchford, M.L. (2011). Uncertainties in $PM_{2.5}$ Gravimetric and Speciation Measurements and What We Can Learn from Them(Article). *Journal of the Air and Waste Management Association, Vol.61*, 1131-1149

McClure, C.D., & Jaffe, D.A. (2018). US particulate matter air quality improves except in wildfire-prone areas. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, Vol.115*, 7901-7906

Md Mostafijur Rahman, G.T. (2022). A hybrid satellite and land use regression model of source-specific $PM_{2.5}$ and $PM_{2.5}$ constituents. *Environment international, Vol.163*, 107233

Meng, X., Garay, M.J., Diner, D.J., Kalashnikova, O.V., Xu, J., & Liu, Y. (2018a). Estimating $PM_{2.5}$ speciation concentrations using prototype 4.4 km-resolution MISR aerosol properties over Southern California. *Atmospheric Environment, Vol.181*, 70-81

Meng, X., Hand, J.L., Schichtel, B.A., & Liu, Y. (2018b). Space-time trends of $PM_{2.5}$ constituents in the conterminous United States estimated by a machine learning approach, 2005–2015(Article). *Environment international, Vol.121*, 1137-1147

P, S., RN, F., L, L., D, S., CR, N., J, L., N, W., CD, C., SE, P., Z, Y., P, C., & Q, Z. (2022). PM 2.5 composition and sources in the San Joaquin Valley of California: A long-term study using ToF-ACSM with the capture vaporizer. *Environmental pollution (Barking, Essex : 1987), Vol.292*, 118254

Potter, N.A., Meltzer, G.Y., Avenbuan, O.N., Raja, A., & Zelikoff, J.T. (2021). Particulate Matter and Associated Metals: A Link with Neurotoxicity and Mental Health. *Atmosphere, Vol.12*, 425

Prospero, J.M., & Lamb, P.J. (2003). African droughts and dust transport to the Caribbean: Climate change implications. *Science, 302*, 1024-1027

Sicheng Li, B.G., Ye Jiang, Xing Wang, Lin Chen, Xue Wang, Ting Chen, La Yang, Yangzong Silang, Feng Hong, Jianzhong Yin, Hualiang Lin, Xing Zhao (2023). Long-term Exposure to Ambient $PM_{2.5}$ and Its Components Associated With Diabetes: Evidence From a Large Population-Based Cohort From China. *Diabetes care, Vol.46*, 111-119

Solomon, P.A., Crumpler, D., Flanagan, J.B., Jayanty, R.K.M., Rickman, E.E., & McDade, C.E. (2014). U.S. National PM 2.5 Chemical Speciation Monitoring Networks—CSN and IMPROVE: Description of networks. *Journal of the Air & Waste Management Association (Taylor & Francis Ltd), Vol.64*, 1410-1438

Sorooshian, A., Wonaschütz, A., Jarjour, E.G., Hashimoto, B.I., Schichtel, B.A., & Betterton, E.A. (2011). An aerosol climatology for a rapidly growing arid region (southern Arizona): Major aerosol species and remotely sensed aerosol properties. *Journal of Geophysical Research:Atmospheres, Vol.116*, 0

Thornton, M.M., Shrestha, R., Wei, Y., Thornton, P.E., & Kao, S.C. (2022). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4 R1. In: ORNL Distributed Active Archive Center

Weatherhead, E.C., Reinsel, G.C., Tiao, G.C., Meng, X.L., Choi, D., Cheang, W.K., Keller, T., DeLuisi, J., Wuebbles, D.J., & Kerr, J.B. (1998). Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *Journal of Geophysical Research: Atmospheres (1984–2012), 103*, 17149-17161

Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., Lyapustin, A., Liu, L., Wu, H., & Song, Y. (2020). Improved 1 km resolution $PM_{2.5}$ estimates across China using enhanced space–time extremely randomized trees. *Atmospheric Chemistry and Physics, Vol.20*, 3273-3289

Wei, J., Wang, J., Li, Z., Kondragunta, S., Anenberg, S., Wang, Y., ... & Ichoku, C. (2023). Long-term mortality burden trends attributed to black carbon and PM2· 5 from wildfire emissions across the continental USA from 2000 to 2020: a deep learning modelling study. *The Lancet Planetary Health*, 7(12), e963-e975.

Woody, M.C., Baker, K.R., Hayes, P.L., Jimenez, J.L., Koo, B., & Pye, H.O.T. (2016). Understanding sources of organic aerosol during CalNex-2010 using the CMAQ-VBS. *Atmos. Chem. Phys., 16*, 4081-4100

WHO. (2021). *WHO global air quality guidelines: particulate matter (PM2. 5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization

Xiaozhe Yin, M.F., Masoud Fallah-Shorshani, Martin Shafer, Rob McConnell, Scott Fruin (2022). Exposure models for particulate matter elemental concentrations in Southern California. *Environment international, Vol.165*, 107247

Yang, Q., Yuan, Q., & Li, T. (2022). Ultrahigh-resolution $PM_{2.5}$ estimation from top-of-atmosphere reflectance with machine learning: Theories, methods, and applications. *Environmental pollution, Vol.306*, 119347

Zhou, Z.-H., & Feng, J. (2019). Deep forest. *National Science Review, 6*, 74-86