Title

Development of over 30-years of high spatiotemporal resolution air pollution models and surfaces for California

Authors
Su, Jason G
Shahriary, Eahsan
Sage, Emma
et al.

Publication Date
2024-11-01

DOI
10.1016/j.envint.2024.109100

Environment International 193 (2024) 109100

Contents lists available at ScienceDirect

Environment International

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

Full length article

Development of over 30-years of high spatiotemporal resolution air pollution models and surfaces for California

Jason G. Su [a,*], Eahsan Shahriary [a], Emma Sage [a], John Jacobsen [a], Katherine Park [a], Arash Mohegh [b]

[a] School of Public Health, University of California, Berkeley Berkeley, CA 94720 the United States of America
[b] Research Division, California Air Resources Board, Sacramento, CA 95812, the United States of America

ARTICLE INFO

Keywords:
Land use regression
Deletion/substitution/addition
Air pollution
Nitrogen dioxide
Fine particulate matter
Ozone
Remote sensing

ABSTRACT

California's diverse geography and meteorological conditions necessitate models capturing fine-grained patterns of air pollution distribution. This study presents the development of high-resolution (100 m) daily land use regression (LUR) models spanning 1989–2021 for nitrogen dioxide (NO_2), fine particulate matter ($PM_{2.5}$), and ozone (O_3) across California. These machine learning LUR algorithms integrated comprehensive data sources, including traffic, land use, land cover, meteorological conditions, vegetation dynamics, and satellite data. The modeling process incorporated historical air quality observations utilizing continuous regulatory, fixed site saturation, and Google Streetcar mobile monitoring data. The model performance (adjusted R^2) for NO_2, $PM_{2.5}$, and O_3 was 84 %, 65 %, and 92 %, respectively.

Over the years, NO_2 concentrations showed a consistent decline, attributed to regulatory efforts and reduced human activities on weekends. Traffic density and weather conditions significantly influenced NO_2 levels. $PM_{2.5}$ concentrations also decreased over time, influenced by aerosol optical depth (AOD), traffic density, weather, and land use patterns, such as developed open spaces and vegetation. Industrial activities and residential areas contributed to higher $PM_{2.5}$ concentrations. O_3 concentrations exhibited no significant annual trend, with higher levels observed on weekends and lower levels associated with traffic density due to the scavenger effect. Weather conditions and land use, such as commercial areas and water bodies, influenced O_3 concentrations.

To extend the prediction of daily NO_2, $PM_{2.5}$, and O_3 to 1989, models were developed for predictors such as daily road traffic, normalized difference vegetation index (NDVI), Ozone Monitoring Instrument (OMI)–NO2, monthly AOD, and OMI-O3. These models enabled effective estimation for any period with known daily weather conditions.

Longitudinal analysis revealed a consistent NO_2 decline, regulatory-driven $PM_{2.5}$ decreases countered by wildfire impacts, and spatially variable O_3 concentrations with no long-term trend. This study enhances understanding of air pollution trends, aiding in identifying lifetime exposure for statewide populations and supporting informed policy decisions and environmental justice advocacy.

1. Introduction

Air pollution remains a persistent threat to public health (Fuller et al. 2022), requiring accurate methodologies to assess exposure and comprehend its complex spatiotemporal dynamics. In relating air pollution to health, Land Use Regression (LUR) models are typically used to develop spatiotemporal surfaces that align with the occurrence of a health outcome being studied. "Surfaces" in this context refers to spatially continuous data representations of air pollutant concentrations across a geographic area. LUR models estimate air pollution

Abbreviations: AOD, Aerosol Optical Depth; CARB, California Air Resources Board; Caltrans, California Department of Transportation; EOS, Earth Observing System; EPA, U.S. Environmental Protection Agency; ESRI, Environmental Systems Research Institute; LUR, Land Use Regression; MAIAC, Multi-angle Implementation of Atmospheric Correction; MODIS, Moderate Resolution Imaging Spectroradiometer; NASA, National Aeronautics and Space Administration; NLCD, National Land Cover Database; NO2, Nitrogen Dioxide; NTL, Nighttime Lights; NVDI, Normalized Difference Vegetation Index; O3, Ozone; OMI, Ozone Monitoring Instrument; PM2.5, Fine particulate matter with diameter ≤ 2.5 μm; USGS, United States Geological Survey; VKT, Vehicle Kilometers Traveled.

* Corresponding author.
E-mail address: jasonsu@berkeley.edu (J.G. Su).

https://doi.org/10.1016/j.envint.2024.109100
Received 27 June 2024; Received in revised form 22 October 2024; Accepted 24 October 2024
Available online 26 October 2024

concentrations at specific monitoring sites using geographic predictors, including land use, traffic volume, and environmental characteristics (Hoek et al. 2008; Ryan and LeMasters 2007). Several LUR models have been developed in California at regional level, mainly in Southern California, focusing on annual or multiple-year single surface prediction of pollutant concentrations (Jones et al. 2020; Lee et al. 2016; Moore et al. 2007; Ross et al. 2006; Su et al. 2009). Recently, machine learning (ML) algorithms have been used for air pollution modeling in California (Castelli et al. 2020; Reid et al. 2015), including the Deletion/Substitution/Addition (D/S/A) algorithm (Beckerman et al. 2013a; Su et al. 2015a). While D/S/A is fundamentally an ML approach, its application in air pollution research serves the same purpose as a LUR model by capturing the spatiotemporal variability of air pollution based on various land use and environmental predictors. However, D/S/A leverages the strengths of both LUR and ML techniques, providing better prediction accuracy and flexibility in handling complex, high-dimensional datasets (Beckerman et al. 2013a; Ren et al. 2020; Su et al. 2015b; Zhang et al. 2021). These traditional and ML integrated LUR models, however, either do not have statewide coverage, have coarser resolution (e.g., over 1 km resolution), or lack many years of continuous coverage (e.g., over 30 years) to identify the very fine-scale variations in pollutant concentrations for statewide multiple decade health studies.

With a land area of 423,970 km^2 and a multitrillion-dollar gross domestic product, the State of California in the U.S. would rank as the world's eighth-largest national economy if it were a nation (Mecklin 2014). California's distinctive geography and meteorological conditions result in pronounced spatial and temporal variations in air quality (Ostro et al. 2010). Sources and concentrations of air pollutants vary significantly across coastal regions, inland valleys, urban centers, and rural landscapes (Hu et al. 2014; Wikipedia contributors 2024). This variability requires models that capture fine-grained spatiotemporal patterns, enhancing the accuracy of exposure assessments (Brokamp et al. 2018; Jerrett et al. 2005).

High-resolution models are essential for understanding the complex relationships between pollutant exposures and health outcomes, including respiratory and cardiovascular diseases, adverse birth outcomes, and mortality (Di et al. 2017; Gauderman et al. 2015; Ha et al. 2014; Pope III et al. 2009a; Pope III et al. 2004; Pope III et al. 2015). Some health outcome studies, such as those investigating the impact of air pollution on life expectancy (Correia et al. 2013; Pope III et al. 2009b; Yin et al. 2020) and comprehending the lifelong consequences of air pollution (Pope 3rd 2000), necessitate extensive longitudinal studies such as those over a span exceeding 30 years. The traditional and ML integrated models provide the foundation for evidence-based policy interventions aimed at reducing pollution exposure misclassification and mitigating health risks, particularly in vulnerable populations (Craig et al. 2008; Giles et al. 2011; Kaufman et al. 2020). The identification of pollution hotspots and the elucidation of disparities in exposure also support environmental justice efforts, ensuring that policies are informed by a comprehensive understanding of both spatial and temporal variations in air quality (Houston et al. 2004; Liu and Marshall 2023; Morello-Frosch and Jesdale 2006; Morello-Frosch et al. 2002; Zou et al. 2014). Understanding the temporal aspects of air quality becomes crucial for immediate health outcomes and discerning the lifelong impact of air pollution on health.

Among air pollutants, fine particulate matter ($PM_{2.5}$) is of particular concern due to its ability to infiltrate the lungs and enter the bloodstream, contributing to the occurrence of respiratory and cardiovascular diseases (Fasola et al. 2020; Horne et al. 2018; Pope III et al. 2018). Nitrogen Dioxide (NO_2), a key indicator of traffic-related air pollution, has been associated with exacerbated respiratory conditions such as asthma (Naidoo 2019; Studnicka et al. 1997). Ozone (O_3), formed through photochemical reactions involving precursor emissions, plays an important role in the formation of ground-level ozone (smog), which is known to aggravate pulmonary disorders (Chuang et al. 2009; Kinney 1999; Rich et al. 2020). In California, the major health concerning criteria pollutants are NO_2, $PM_{2.5}$ and O_3.

In terms of spatial distribution, NO_2 exhibits a steep spatial gradient, with concentrations decreasing significantly as distance from emission sources increases (Monn et al. 1997; Su et al. 2009; Tack et al. 2017). $PM_{2.5}$, comprising both primary and secondary particles, displays a more gradual spatial attenuation (Wang et al. 2020). Conversely, O_3 distribution tends to exhibit an inverse spatial relationship to NO_2, influenced by the NOx titration effect (Kumar et al. 2008; Wang 2020). The distinct spatial characteristics and health impacts of $PM_{2.5}$, NO_2, and O_3 underscore their centrality to our study and highlight the importance of capturing their respective distribution patterns in exposure assessment models.

Mobile monitoring significantly improves the capture of detailed spatial and temporal variations in pollutant concentrations, leading to more accurate and comprehensive air pollution modeling. By incorporating data from Google Streetcar measurements, particularly those near highway roadways, we enhance spatiotemporal coverage beyond traditional regulatory air quality monitors. This approach allows for a finer resolution of data across diverse environments, contributing to a deeper understanding of exposure patterns and their potential health impacts.

In this research, we propose developing daily air pollution models of 30 m resolution across three decades using the D/S/A integrated LUR modeling technique (Beckerman et al. 2013b; Su et al. 2015a; Su et al. 2015b; Su et al. 2020) for NO_2, $PM_{2.5}$ and O_3. This approach incorporates diverse datasets, including traffic, land use, land cover, vegetation dynamics, meteorological conditions, satellite remote sensing data, and other data sources. The modeling approach integrates air pollution measurements data from government regulatory monitoring, fixed site saturation monitoring and Google Streetcar mobile monitoring. Additionally, we extend predictors to periods with no observations for ensuring the temporal continuity of the models, allowing for a comprehensive and consistent analysis of air pollution dynamics across an extended timeframe.

The research results will be used to help identify life-time air pollution exposure to statewide patients, including adverse birth outcomes and population life expectancy over 30 years, particularly for those vulnerable in California. These models and surfaces also provide the ability to identify historical environmental exposure disparities and trends due to their high spatial resolution. This work will also support future air pollution research studies that require high-precision air pollution surfaces over an extended period to help identify their association with major health outcomes of interest.

2. Methodology

2.1. 1. Acquiring and processing air pollution data from regulatory monitoring

We acquired and processed daily air pollution data and their spatial locations from the U.S. Environmental Protection Agency (https://aqs.epa.gov/aqsweb/airdata/download_files.html). The regulatory data measurements were obtained from monitoring sites equipped with standardized instruments for measuring air pollutants. Specifically, NO_2 was measured using instruments coded as 42602, which typically involve chemiluminescence techniques, recognized for their accuracy in detecting nitrogen dioxide levels in ambient air. $PM_{2.5}$ concentrations were measured using Federal Reference Method (FRM) or Federal Equivalent Method (FEM) instruments coded as 88101, which involve either gravimetric or continuous monitoring techniques to capture fine particulate matter in the air. Ozone (O_3) measurements were conducted using instruments coded as 44201, which commonly utilize ultraviolet photometry to accurately measure ozone concentrations. In California, the spatial distribution of the regulatory air quality monitoring data for NO_2, $PM_{2.5}$ and O_3 are presented in Fig. 1 (left for NO_2, middle for $PM_{2.5}$

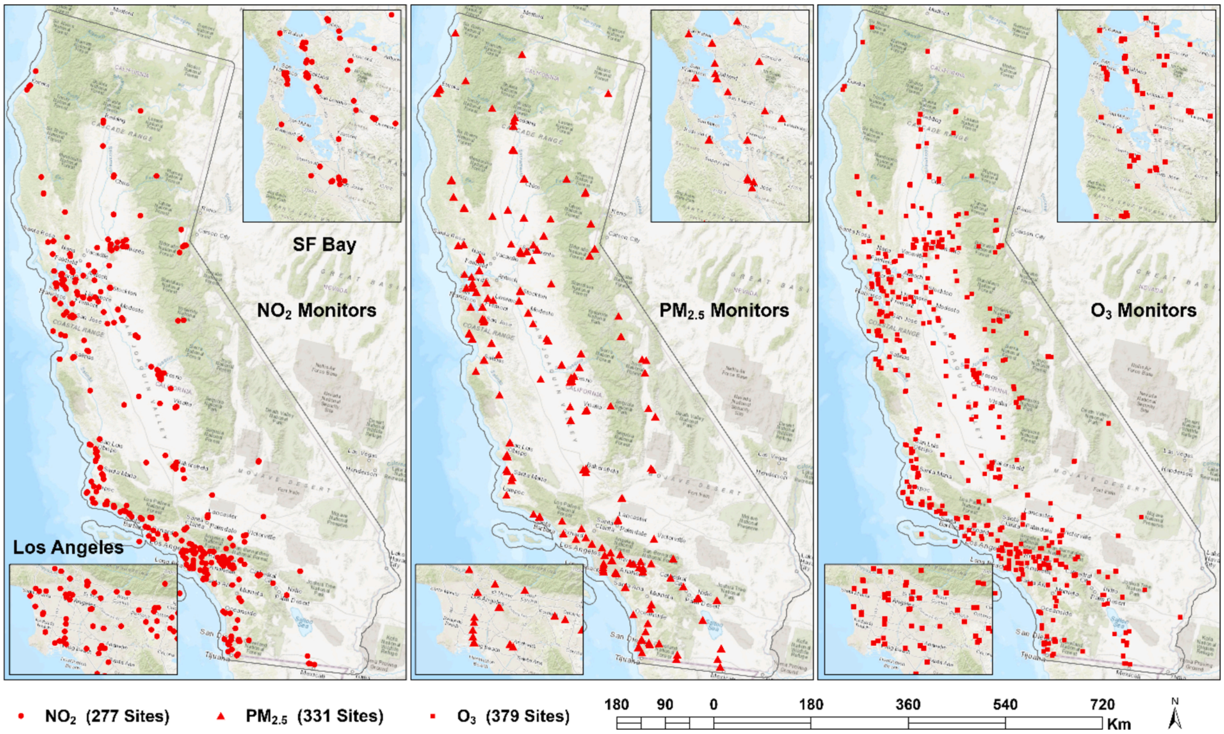

Fig. 1. The spatial distributions of the regulatory monitors for NO_2 (left panel), $PM_{2.5}$ (middle panel), and O_3 (right panel) over the observable time periods.

and right for O_3) and the respective unique number of regulatory sites is presented in Table 1.

The trend for NO_2 measurement sites shows a slight decline during the early 1990 s, with the number of unique sites decreasing from 151 in 1990 to 147 in 2000. This downward trend continued until 2006, when the number of monitoring sites reached its lowest point. After 2006, the number of unique NO_2 measurement sites fluctuated between 127 and 135, suggesting variability in monitoring efforts. Overall, there is no consistent upward or downward trend in NO_2 monitoring, indicating that the focus on this pollutant has varied over the years. The total number of unique NO_2 air quality monitors is 277. In contrast, the trend for $PM_{2.5}$ reveals a clear upward trajectory in the number of unique measurement sites. Starting with 183 sites in 1999, the number steadily increased to 252 by 2021. This growth is particularly evident from 2000 onward, demonstrating a growing recognition of the importance of this pollutant and dedicated resources to understanding and mitigating its impacts. The total number of unique $PM_{2.5}$ air quality monitors is 331. For O_3, the trend indicates a generally stable pattern with a gradual increase in monitoring sites over time. The number of unique O_3 measurement sites increased from 194 in 1990 to 198 in 2008, with some fluctuations throughout the years. Although the overall growth in O_3 monitoring efforts is less pronounced than that of $PM_{2.5}$, it still demonstrates a steady commitment to tracking this pollutant. The total number of unique O_3 air quality monitors is 379.

In our modeling process, we also applied fixed site saturation monitoring data in our analysis. A detailed description of the saturation monitoring data can be found in Su et al. (2020).

Table 1
The unique number of regulatory monitoring sites with the respective effective measurements of NO_2, $PM_{2.5}$ and O_3 across the study period.

Year	Number of Unique Sites		
	NO_2	$PM_{2.5}$	O_3
1989			182
1990	151		194
1991	150		201
1992	149		205
1993	159		199
1994	164		208
1995	163		
1996	159		
1997	156		
1998	154		
1999	148	183	
2000	147		
2001	153		
2006	127		186
2007	129	213	195
2008	136	221	198
2009	130	225	192
2010	132	228	194
2011	127	229	196
2012	132	248	198
2013	129	242	190
2014	132	246	189
2015	133	240	185
2016	135	238	185
2017	132	240	184
2018	129	246	180
2019	128	241	181
2020	124	247	182
2021	127	252	178
Total	277	331	379

2.2. Acquiring and processing air pollution data from from Google Streetcar monitoring

Google Streetcar had mobile monitoring of the three criteria pollutants across San Francisco Bay (counties of Alameda, Contra Costa, San Francisco and San Mateo), Los Angeles County, and Central Valley regions (see: https://www.google.com/earth/outreach/special-projects/air-quality/). The Google Streetcar mobile measurements for each region are highly spatially autocorrelated due to the intense sampling of air pollutants on its road network. To ensure that our models captured a wide range of variability in road traffic patterns while minimizing the influence of spatial autocorrelation, we selected 150 road segments for each region through a location-allocation algorithm (Kanaroglou et al.

2005). Spatial autocorrelation can lead to inflated model performance metrics and reduced generalizability by over-representing certain areas or patterns. By using the location-allocation algorithm, we distributed the selected road segments more evenly across each region, reducing clustering and ensuring that our models are better representative of the broader spatial patterns across California. This approach helped in developing more robust and interpretable models by preventing overfitting to localized traffic conditions. A total of 150 road segments with each road segment having at least 100 measurements was selected for each of the four regions: Alameda and Contra Costa; San Francisco and San Mateo; Los Angeles, and Central Valley. Each region had (1) 50 road segments selected from locations within 500 m of highways allowing truck traffic, or within 500 m of major California ports (i.e., goods movement corridors or GMCs), (2) 50 road segments selected from locations within 500 m of highways not allowing truck traffic or within 300 m of major roadways (i.e., non-goods movement corridors or NGMCs), and (3) locations not encompassed in the first and second parts (i.e., control areas or CTRLs). The detailed selection process is documented in the Supplementary file. The Google Streetcar measured NO_2 and O_3 concentrations in the unit of ppb – the same as regulatory monitoring; however, $PM_{2.5}$ concentrations were in total number of particles instead of the typical concentrations in μg m^{-3}. The daily concentration of $PM_{2.5}$ in μg m^{-3} of road segment i of traffic corridor k on day j was estimated through:

$$C_{i,j,k} = G_{i,j,k} * \widehat{R_{j,k}} / \widehat{G_{j,k}} \quad (1)$$

where $C_{i,j,k}$ and $G_{i,j,k}$ represent the converted and original measures. $\widehat{R_{j,k}}$ and $\widehat{G_{j,k}}$ are respectively the mean $PM_{2.5}$ concentrations in $\mu g/m^{-3}(-|-)$ from all the regulatory monitors and the mean $PM_{2.5}$ particle numbers from all the selected 50 road segments for day j in corridor k. The $PM_{2.5}$ concentrations were estimated separately for each region.

2.3. Acquiring and processing air pollution predictors from the observation period

For the predictors (Table 2), the availability of daily traffic data varied across 12 California Department of Transportation (Caltrans) districts (Figure S2), with the earliest traffic data available from 2000 to 2005. We used the data collected by the Caltrans Performance Measurement System (PeMS) to derive roadway daily traffic. PeMS data are collected in real-time from nearly 40,000 individual detectors spanning the freeway system across all major metropolitan areas of the State of California and provide an Archived Data User Service that provides over fifteen years of data for historical analysis. The detector measured traffic flow covered ~ 5 % highway segments and we summed hourly traffic to daily traffic for all the stations across California. The interconnected steps were then used to derive daily traffic for all the California highways. Please refer to the Supplementary file for the details of traffic assignment.

The land use data was derived from the 2019 statewide parcel data, combined by the California Air Resources Board (CARB) from individual County Assessor's Offices, and we considered them consistent across all the years. The land cover data was acquired from the National Land Cover Database (NLCD) at five-year intervals (2001, 2006, 2011, 2016, and 2019) (Yang et al. 2018). The assumption was that land cover remained constant until the subsequent available measurement. Vegetation dynamics were assessed through the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument-derived data, specifically the Normalized Difference Vegetation Index (NDVI) (Lunetta et al. 2022), computed at 16-day intervals since 2000. We assumed the vegetation index remained constant from its previous measurements within 16 days. Daily meteorological data were acquired from the GridMet dataset (Abatzoglou 2013), covering 1989 to 2021 at a 4 km spatial resolution. For satellite remote sensing data, daily measurements from the Ozone Monitoring Instrument (OMI) (Levelt et al. 2018) for NO_2 and O_3 were accessible from 2005 to 2021. The aerosol optical depth (AOD) data (Zhang et al. 2011) was available from 2000 to 2021.

2.4. Extending air pollution predictors to unobserved periods

Backcasting daily traffic.

The earliest traffic data available for California ranged from 2000 to 2004 (Table 2). The range of dates for traffic data availability is due to increased efforts by Caltrans to manage traffic across California. They initially focused on densely populated areas, such as the San Francisco Bay Area in District 4 and Los Angeles in District 7 (Figure S2), before expanding to other less populated districts. We used a linear mixed-effects model to estimate missing traffic for the unobserved period. The linear mixed-effects model allowed us to account for both fixed and random effects to accurately predict daily traffic. The fixed effect was the year, capturing any overarching trends in traffic volume over time.

Table 2
LUR predictors and available time periods in the modeling process.

Variables	Source	Spatial Resolution	Temporal Resolution	Time Period	Extension Period
Traffic[δ]	CalTrans	30 *m*	Daily	2005–2021	1989–2004
Land use[θ]	CARB	30 *m*	One time	2019	Use 2019
Land cover[¥]	NLCD	30 *m*	Every 5 years	2001–2019	Use 2001
Vegetation index (NDVI) [€]	MODIS	250 *m*	Every 16 days	2000–2021	1989–1999
Meteorological data[£]	GridMet	4 km	Daily	1989–2021	None
AOD data[ξ]	MAIAC	1 km	Daily	2000–2021	1989–1999
OMI-NO_2 data[ξ]	NASA's OMI	25 km	Daily	2005–2021	1989–2004
OMI-O_3 data[ξ]	NASA's OMI	25 km	Daily	2005–2021	1989–2004
Distance to highway and major roadways[ǂ]	ESRI	30 *m*	One time	2018	None
Distance to coast[ǂ]	USGS	30 *m*	One time	2015	None
Elevation from digital elevation model[ǂ]	USGS	30 *m*	One time	2015	None
Distance to ports[ǂ]	ESRI	30 *m*	One time	2018	None

ξ: MAIAC AOD data: Data from the Multi-angle Implementation of Atmospheric Correction (MAIAC) algorithm using MODIS Terra and Aqua satellites; OMI-NO_2 and OMI-NO_3 data are derived from the National Aeronautics and Space Administration Ozone Monitoring Instrument.

[δ] : Traffic data are derived from the California Department of Transportation (CalTrans).

[θ] : Land use data are provided by the California Air Resources Board (CARB), which combined the parcel data from all the 58 counties in California.

[¥] : Land cover data is derived from the NLCD (National Land Cover Database) provided by the U.S. Geological Survey (USGS).

[€] : The NDVI (Normalized Difference Vegetation Index) data is provided by MODIS (Moderate Resolution Imaging Spectroradiometer) from NASA's Earth Observing System (EOS).

[£] : The meteorological data is sourced from GridMet provided by the University of Idaho.

[ǂ] : Traditional predictors include distance to the nearest highway and major roadway derived from the ESRI Street data layer for 2018, distance to coast and elevation data derived from the USGS for 2015, and distance to major ports derived from the ESRI data layer for 2018.

Meanwhile, the random effects included the road segment's route, the county in which it was located, the specific month, season, and whether the day in question was a weekday or weekend. For each Caltrans district, we developed separate models that reflected the district-specific relationships between these factors and daily traffic. This district-specific modeling was crucial as traffic patterns can vary significantly across California's diverse regions. Once the models were established for each district, they were applied to estimate daily traffic on road segments for days where traffic data was missing, specifically targeting the years without observations.

Backcasting daily NDVI data:

No MODIS NDVI data is available before 2000, as indicated in Table 2. We used a multiple linear regression modeling technique to backcaste daily NDVI data. In this approach, we used long-term monthly average NDVI values as the baseline and modeled the relationship between daily NDVI values and various meteorological variables (such as precipitation, temperature, relative humidity, wind speed, and wind direction) as predictors. We recognize that meteorological conditions such as temperature, precipitation, and humidity directly impact vegetation growth and health. NDVI, which is a measure of vegetation density and health, can vary significantly with changes in these meteorological factors. For instance, higher temperatures and varying precipitation levels can affect plant growth cycles and chlorophyll content, thus influencing NDVI values. This regression model allowed us to estimate daily NDVI values for the period before 2000, extending back to 1989.

Backcasting daily OMI-NO_2 data:

In constructing the daily NO_2 model, we identified OMI-NO_2 satellite remote sensing data with the highest t-score, indicating its significant impact on predicting daily NO_2 values. However, OMI-NO_2's spatial resolution of 25 km led to edge effects along the resolution cells, and the data did not cover periods before 2005 (Table 2). To address these challenges, we incorporated NASA's (National Aeronautics and Space Administration) annual NO_2 re-analysis data at a 1 km resolution for 1990, 1995, 2000, and 2005–2020 (Anenberg et al. (2023)). This augmentation aimed to enhance the spatial resolution of OMI-NO_2 data and extend it back to 1989. For the missing years in the NO_2 re-analysis data, we employed multiple linear regression to estimate values based on available data from adjacent years, incorporating variables such as temperature, wind speed, and population density.. All the data were resampled to a spatial resolution of 1 km during the modeling process. We assumed that there are relationships between long-term average OMI-NO_2 values for specific days of the month (e.g., the 1st day) and for specific months (e.g., January), and the OMI-NO_2 values for the corresponding specific dates (e.g., January 1, 2005). Additionally, we assumed a long-term trend in OMI-NO_2 values and used variable year in a linear regression model to extend annual OMI-NO_2 values from 2005 to 2021 to 1989–2004. By incorporating annual, long-term monthly, and long-term daily (day 1–31) OMI-NO_2 data with annual NO_2 reanalysis data, we were able to accurately model and predict daily OMI-NO_2 values through a multiple regression model.. The modeling outcomes were then used to derive daily OMI-NO_2 values from 1989 to 2021, with an improved spatial resolution of 1 km.

Backcasting monthly AOD data:

The earliest dates available for AOD data were in 2000 (Table 2). In our modeling of $PM_{2.5}$, we opted to use monthly AOD median values in our modeling process due to extensive missing data from cloud impact at the daily level. To extend monthly AOD data to 1989, we used the annual $PM_{2.5}$ data of resolution 1 km for 1989–2016 from Washington University in St. Louis (Van Donkelaar et al. 2019). We assumed similarities in AOD values between a specific month (e.g., January 2005) and its long-term month (e.g., January) and year (e.g., 2005). Additionally, we assumed a long-term annual trend in AOD values and conducted grid-wise linear regression trend analysis to extend annual AOD values from 2000 to 2021 to 1989–1999. Integrating long-term year and month AOD values with Van Donkelaar et al. (2019) annual $PM_{2.5}$ data enabled the prediction of monthly AOD values. The modeling outcomes were then used to derive monthly AOD values from 1989 to 1999, with a spatial resolution of 1 km.

Backcasting daily OMI-O_3 data:

Like OMI-NO_2 data, OMI-O_3 data lacks coverage for periods before 2005 (Table 2). We utilized a linear regression modeling approach to estimate daily OMI-O_3 values based on their long-term daily (1–31), monthly (1–12), and yearly (2005–2021) values. Additionally, we included daily OMI-NO_2 data as a predictor in the model. Subsequently, the modeling results were used to extend daily OMI-O_3 data back to 1989. In the modeling process, yearly OMI-O_3 values for 1989–2004 were extrapolated through trend analysis, and daily OMI-NO_2 data for 1989–2004 were obtained using the above extension procedure.

2.5. Developig daily air pollution models through ML integrated LUR approach

The D/S/A algorithm initiates the selection process by starting with a base model, typically the intercept-only model unless a different starting point is specified. The algorithm then iteratively adds, deletes, or substitutes terms to improve the model's predictive performance. During each iteration, potential modifications to the model, such as adding polynomial terms or interaction effects, are evaluated based on a predefined criterion, usually the reduction of the cross-validated error or the improvement in some other model performance metric. The selection process continues iteratively, with the algorithm testing various combinations of terms and retaining the modifications that lead to the greatest improvement in model performance. This process is similar to a guided search through the space of possible models, where each step is evaluated to ensure it moves toward a better fit. The algorithm halts its iterations when no further modifications result in a significant improvement in the model's performance, according to the predefined stopping criteria. These criteria could include a threshold for the minimum improvement in cross-validated R-squared or reaching a maximum number of iterations (15 in our research). At this point, the model with the optimal combination of terms is selected as the final model, representing the best balance between complexity and predictive accuracy. To enhance the interpretability of our modeling results, we limited the predictors to linear terms and avoided interaction terms..

For regulatory and saturation monitoring data, each was treated independently, randomized, and divided into 10 equal folds without considering spatial or temporal constraints. The Google Streetcar data, which spanned multiple regions, was randomized and divided into 10 folds separately for each region. These region-specific folds were then merged with the corresponding folds from the other regions, as well as with the 10 randomized folds from the regulatory and saturation monitoring datasets. This approach ensured that each of the 10 folds contained a balanced mix of data from all monitoring types and regions. One subsample was then retained as validation data, while the remaining 9 subsamples served as training data during the modeling process. This cross-validation process was repeated 10 times, with each subsample used once as validation data.

In developing the daily LUR models for NO_2, $PM_{2.5}$, and O_3, we constructed respective models using only available observable dates for both predictors and air quality measures. No algorithms of temporal extensions to the predictors were applied during the modeling process. The modeling results, however, were applied to all the predictors across all the years to predict daily NO_2, $PM_{2.5}$ and O_3 concentrations for the 1989–2021 period.

3. Results

3.1. D/S/A integrated LUR models covering the available observational periods

Table S1-S3 present the daily LUR models, capturing the available observational time periods for NO_2, $PM_{2.5}$, and O_3. In the case of NO_2 (Table S1), the consistent year-after-year decline in concentrations observed during the study period was reflected in the variable "year" and this could be attributed to the regulatory efforts to reduce traffic NO_2 emissions. The recurrent pattern of lower concentrations during weekends compared to weekdays suggests potential reductions in human activities on roadways. Additionally, the positive correlation between higher OMI-NO_2 values and increased NO_2 concentrations underscores the significance of remote sensing observations in capturing spatial variability. Traffic density emerged as a significant factor, as areas with greater vehicular activity exhibited greater NO_2 emissions and higher concentrations. Moreover, weather conditions played a crucial role, with higher relative humidity, wind speed, and temperature contributing to lower NO_2 concentrations. Conversely, increased precipitation was linked to higher NO_2 levels, highlighting the interplay between meteorological conditions and NO_2 dynamics. Residential areas were found to have lower NO_2 concentrations, as well as in the developed open spaces. Low and high-intensity developments, on the other hand, were associated with greater NO_2 concentrations, indicating the positive association of urban development with NO_2 levels. The availability of green spaces, indicated by higher vegetation index, shrub cover, and wetlands—recognized as pollution sinks—was associated with lower NO2 concentrations. Conversely, a higher proportion of impervious surfaces was correlated with increased NO2 levels. Additionally, locations farther from ports displayed lower NO_2 concentrations, indicating elevated NO_2 levels near ports. The NO_2 model had an adjusted R^2 of 0.84 in variance explained.

For $PM_{2.5}$ (Table S2), throughout the study period, its concentrations consistently decreased, mirroring the trend observed for NO_2. The study identified a positive correlation between higher aerosol optical depth (AOD) values and elevated $PM_{2.5}$ concentrations, suggesting that increased aerosol presence in the atmosphere is associated with higher particulate matter levels. Increased traffic density emerged as a contributing factor to higher $PM_{2.5}$ concentrations, emphasizing the impact of vehicular emissions on air quality. Weather factors such as higher relative humidity, wind speed, and temperature were associated with lower $PM_{2.5}$ concentrations. Developed open spaces were linked to reduced $PM_{2.5}$ concentrations, and so were areas characterized by a higher vegetation index, shrub cover, barren land, and water bodies, emphasizing the role of natural features in mitigating air pollution. Barren land refers to areas that have little to no vegetation cover and is often characterized by exposed soil or rock (Homer et al. 2015). Industrial land use, however, was associated with higher $PM_{2.5}$ concentrations, pointing to the impact of industrial activities on particulate matter emissions. In contrast to NO_2, greater residential areas were linked to higher $PM_{2.5}$ concentrations, potentially attributed to background concentrations. In densely populated regions, the increased density of housing, traffic, and other activities can lead to elevated $PM_{2.5}$ background concentrations. Additionally, the urban heat island effect and limited air circulation in residential areas can hinder the dispersion of pollutants, allowing background $PM_{2.5}$ levels to rise. Additionally, locations farther from the coast were associated with higher $PM_{2.5}$ concentrations, indicating a spatial relationship between proximity to the coast and particulate matter levels. The final $PM_{2.5}$ model had a predictive performance of 0.65.

In contrast to the patterns observed for NO_2 and $PM_{2.5}$, O_3 concentrations exhibited predominantly opposing relationships (Table S3). The variable "year" did not show a significant association with O_3 concentrations, indicating the absence of an annual trend in O_3 levels. Weekends were characterized by higher O_3 concentrations than weekdays, revealing a distinct opposite temporal pattern. Higher OMI-O_3 values were linked to greater O_3 concentrations, emphasizing the positive association of remote sensing observations with measured ozone levels. Surprisingly, greater traffic was associated with lower O_3 concentrations, suggesting a nuanced photochemical process (i.e., scavenger effect, see details in discussion of Fig. 4) between vehicular emissions and ozone dynamics. Weather factors such as higher relative humidity, wind speed, and atmospheric pressure correlated with elevated O_3 concentrations, underscoring the influence of meteorological conditions on ozone levels. Land use patterns also played a role, with government & institutional, commercial, and waterbody areas associated with higher O_3 concentrations, while barren land, crops, and wetlands were linked to lower O_3 concentrations. Developed low, medium, and high-intensity developments were associated with lower ozone concentrations, suggesting potential lower concentrations in urban areas. Low-intensity development includes areas with sparse residential or commercial buildings, such as small towns or suburban neighborhoods. Medium-intensity development encompasses areas with more concentrated buildings and infrastructure, typically found in denser suburban or urban areas with moderate residential and commercial activities. High-intensity development represents the most densely built areas, including central business districts and urban centers with significant residential, commercial, and industrial structures (Homer et al. 2015). Moreover, greater distances from highways were associated with higher O_3 concentrations, highlighting a similar scavenger effect between proximity to highways and ozone levels. The final O_3 model had a predictive performance of 0.92.

3.2. Modeling and extending model predictors to 1989

To extend the prediction of daily NO_2, $PM_{2.5,}$ and O_3 beyond the observable time periods to 1989, models were developed for predictors such as daily road traffic, daily NDVI, daily OMI-NO_2, monthly AOD, and daily OMI-O_3. These models facilitated the extension of predictions back to 1989. Regarding daily road traffic (Table S5), the overall predictive performance (Conditional R^2) ranged from 0.33 to 077, with the fixed effect predictor "year" demonstrating relatively lower model performance compared to random effects variables like season, month, weekend, and county. Except for District 9, its fixed effect variable explained a 33.9 % variance. Daily NDVI predictions were based on 16-day NDVI and corresponding weather conditions during measurement days. As depicted in Figure S3a and Table S6, utilizing NDVI's long-term monthly means and daily weather conditions yielded an effective prediction (adjusted $R^2 = 0.98$) for any time period with known daily weather conditions.

For daily OMI-NO_2 (Figure S3b and Table S7), the inclusion of OMI-NO_2's long-term conditions (daily, monthly, and yearly) along with NASA NO_2 re-analysis annual data resulted in a model performance (adjusted R^2) of 0.81, enabling estimation back to 1989. Monthly AOD predictions (Figure S3c and Table S8) utilized long-term monthly AOD and Van Donkelaar et al. (2019) annual $PM_{2.5}$, effectively predicting monthly AOD values (adjusted $R^2 = 0.94$) and allowing estimation back to 1989. As for daily OMI-O_3 (Figure S3d and Table S9), the incorporation of OMI-O_3's long-term conditions (daily, monthly, and yearly) and daily OMI-NO_2 data yielded a model performance (adjusted R^2) of 0.99, making it practical for estimating daily OMI-O_3 back to 1989.

3.3. Daily air pollution surfaces covering 1989–2021

Once all the predictors with temporal components were extended to the year 1989, the NO_2, $PM_{2.5,}$ and O_3 models, presented respectively in Tables 2, 3, and 4, were run for those days missing predictions, and the final surfaces included daily NO_2, $PM_{2.5}$ and O_3 concentrations for California at a spatial resolution of 100 m for the years of 1989–2021.

Fig. 2 shows the aggregated annual concentration surfaces of NO_2 for four decennial years, including 1990, 2000, 2010, and 2020. The spatial patterns clearly show the decrease in NO_2 concentrations throughout the years, especially in the urban areas. To identify degrees of reduction throughout California, we used regulatory monitors for NO_2, $PM_{2.5,}$ and O_3 (Fig. 1) to identify average decennial concentrations for the State. This approach is reasonable given the state regulatory monitors are designed to ensure comprehensive spatial coverage, capturing the

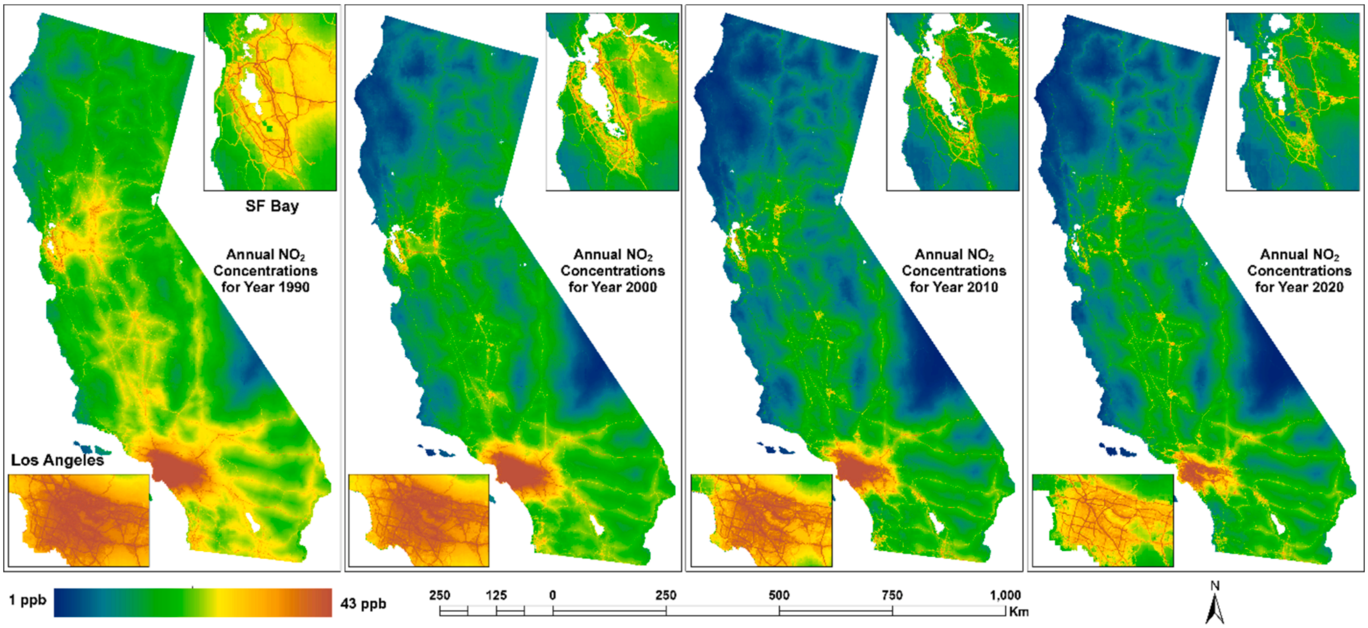

Fig. 2. Decennial years of NO_2 surfaces among the over 30- years study period.

diverse environmental conditions across the state, including coastal, inland, and mountainous regions. By incorporating monitoring points from both urban and rural areas, it enables the examination of the urban–rural gradient in air pollution. These holistic statewide air quality monitors also allow for the identification of spatial patterns, hotspots, and potential disparities in pollution concentrations. Though some points are duplicated due to multiple pollutants being measured at the same time, they reflect the importance of those points in geographic placement strategies. Moreover, utilizing data from 1410 monitoring sites enhances the statistical robustness of the analysis, providing a more accurate assessment of statewide air pollution levels. Using those 1410 locations, we found that the average NO_2 concentrations decreased from 18.1 ppb in 1990 to 14.1 ppb in 2000, and decreased to 9.7 ppb in 2010 and 8.0 ppb in 2020. For $PM_{2.5}$, similar trends were identified for the four decennial years but with a much smaller decrease (Fig. 3). A striking change in 2020 was that the $PM_{2.5}$ levels increased significantly in Central Valley while other places decreased, especially in Los Angeles, which experienced the greatest decline. We suspect the significant increase in $PM_{2.5}$ levels in Central Valley in 2020 was due to the significant impact of wildfires.(Keeley and Syphard 2021) Using the locations of the 1410 regulatory monitors, we found that the average $PM_{2.5}$ concentrations decreased from 14.2 μg m^{-3} in 1990 to 12.0 μg m^{-3} in 2000, and further decreased to 9.9 μg m^{-3} in 2010 but increased to 12.2 μg m^{-3} in 2020. The increase in wildfire frequency and intensity in California (Brown et al. 2023; Keeley and Syphard 2021; Li and Banerjee 2021) will further increase $PM_{2.5}$ levels, though regulatory actions have significantly reduced traffic and industry-related $PM_{2.5}$.

For O_3 (Fig. 4), we did not see any apparent trend, but we did identify that those urban metropolitan areas, such as the San Francisco Bay and Los Angeles Metro, had relatively lower O_3 concentrations compared to

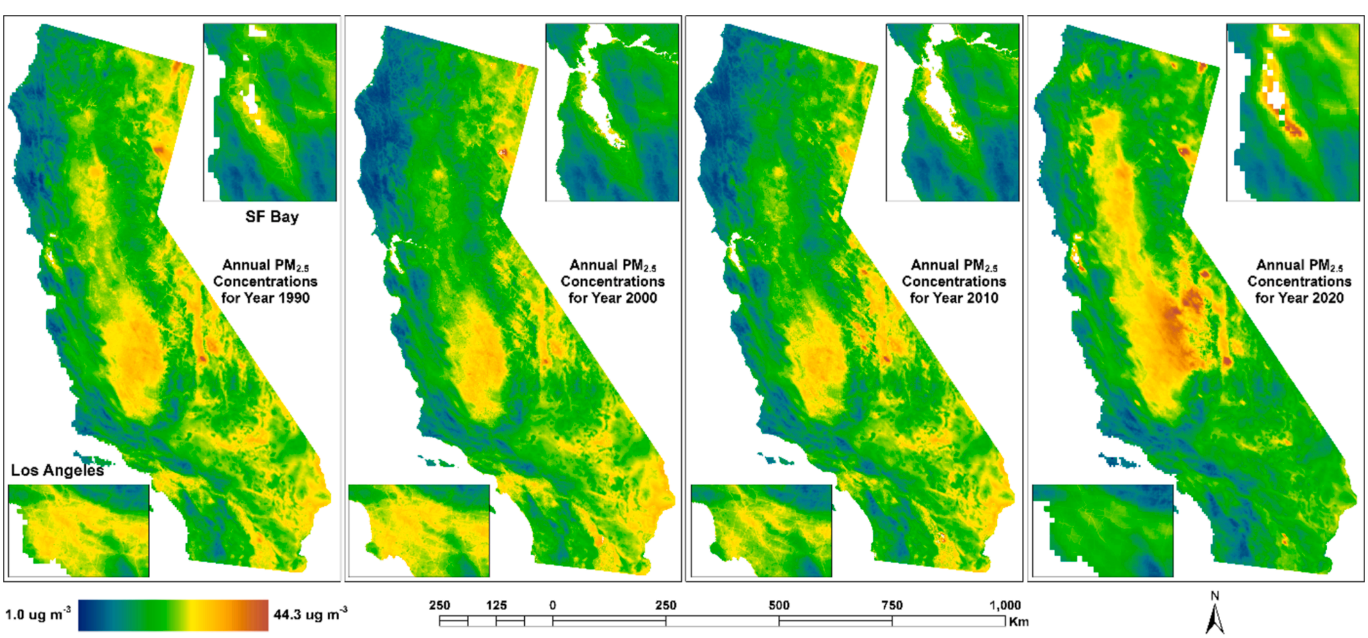

Fig. 3. Decennial years of $PM_{2.5}$ surfaces among the over 30- years study period.

rural areas. This is very likely due to the O_3 scavenger effect (Larsen and Sacramento 2003). The scavenger effect involves the removal or reduction of ozone from the atmosphere due to the presence of specific pollutants or conditions. These pollutants can act as scavengers by reacting with ozone molecules, leading to a decrease in overall ozone concentrations. Common scavengers of ozone include nitrogen oxides (NOx), carbon monoxide (CO), volatile organic compounds (VOCs), and particulate matter. In urban environments, where these pollutants are often abundant due to human activities such as combustion processes and industrial emissions, the scavenger effect can be more pronounced. Nitrogen oxides, particularly NO_2, can react with ozone in the presence of sunlight to form nitric oxide (NO) and oxygen (O_2). This process reduces the overall ozone levels in the atmosphere. VOCs and carbon monoxide can also participate in ozone-depleting reactions. These compounds can undergo photochemical reactions that consume ozone while generating other pollutants. Using a total of 1410 spatial points from regulatory monitors, we found that the overall O_3 level did not change significantly through those four decennial years: the average O_3 concentrations decreased from 38.2 ppb in 1990 to 37.8 ppb in 2000, and slightly increased to 38.1 ppb in 2010 and 39.3 ppb in 2020.

Further, we provided daily air pollution surfaces for NO_2, $PM_{2.5}$, and O_3 for January 1st, 2019 (Figure S4) and compared them with the corresponding nearest centennial annual surfaces (Figs. 2-4). We found that for NO_2, the daily surface closely matched the spatial patterns of the annual surface. For $PM_{2.5}$, the patterns were also similar, though there was a significant increase in the Sierra region (eastern part of the map), suggesting a potential impact from wildfires. For O_3, while the general patterns were consistent in Northern California, the LA metropolitan area in Southern California showed higher O_3 concentrations on the daily map, which were less prominent in the annual data. These comparisons indicate that while spatial patterns were largely consistent from daily to annual concentrations, there were notable differences in daily spatial patterns, particularly for $PM_{2.5}$ and O_3, likely due to the impact of temporal factors like wildfires and weather.

4. 8. Historical trend analysis covering 1989–2021

To assess the historical trends of the three pollutants, we extracted daily concentration values from 1,410 monitoring sites used in the study. Annual mean values were then calculated for each pollutant at these locations to capture long-term trends over the entire study period (Fig. 5). The analysis of NO_2 levels over the 30-year period reveals a significant decline. The trend equation, $y = -0.34x + 690.89$, with an $R^2 = 0.99$, indicates a strong negative correlation, suggesting a steady decrease in NO_2 concentrations over time. The trend for $PM_{2.5}$ also shows a decline, though less steep compared to NO_2. The trend equation, $y = -0.13x + 279.77$, and $R^2 = 0.72$, suggest a moderate reduction in $PM_{2.5}$ levels. Despite this decrease, recent years have seen spikes in $PM_{2.5}$ concentrations due to increased wildfire activity, which has influenced the overall trend. The 2018 Camp Fire was the deadliest and most destructive wildfire in California's history, burning over 153,000 acres and resulting in 85 deaths (Blackford 2024; Rooney et al. 2020). It completely devastated the town of Paradise. The 2020 Complex Fire was California's largest wildfire by acreage, burning over 1 million acres across multiple counties. It was composed of several fires that merged into one (Keeley and Syphard 2021; Safford et al. 2022). The 2021 Dixie Fire was the second-largest fire in California's history, which burned over 960,000 acres across five counties, destroying hundreds of structures and threatening many more. Unlike NO_2 and $PM_{2.5}$, O_3 concentrations show a slight upward trend over the study period, with the trend equation $y = 0.04x - 49.4$ and an $R^2 = 0.47$. Despite reductions in NO_2, the ozone levels have been influenced by factors such as increasing temperatures and changing atmospheric chemistry, which complicate ozone management.

5. Discussions and Conclusion

With advancements in technology, various ML algorithms have increasingly been applied to air pollution modeling, including neural network (Cabaneros et al. 2019), random forest (Kumar 2018), gradient boosting (Peng et al. 2023), support vector machines (Leong et al. 2020) and other techniques (Masood and Ahmad 2021), as well as models that combine multiple ML algorithms. Generally, ML algorithms demonstrate better predictive performance compared to traditional LUR models (Ren et al. 2020), although there are instances where LUR models perform better (Kerckhoffs et al. 2019). Additionally, models that integrate multiple ML algorithms tend to outperform those using individual algorithms. For example, Gocheva-Ilieva et al. (2020) reported a model performance of an adjusted R^2 of 0.749 for NO_2 and 0.836 for O_3 using random forest modeling, which increased to 0.945 and 0.978,

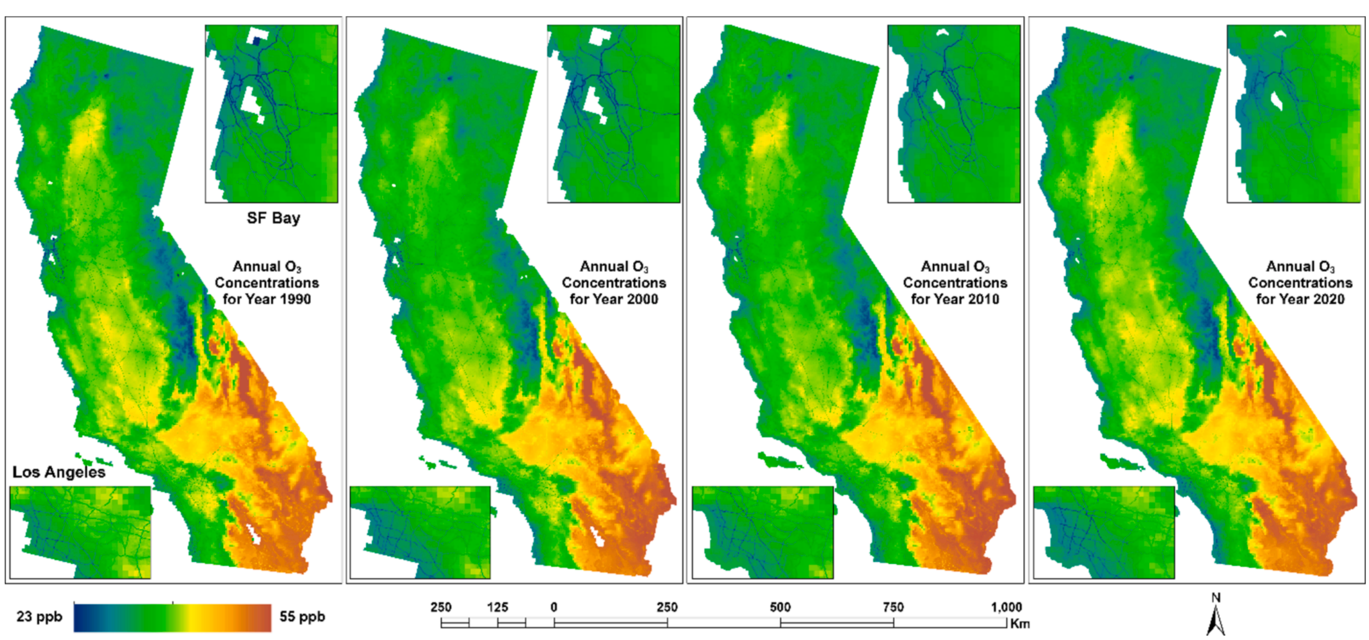

Fig. 4. Decennial years of O_3 surfaces among the over 30- years study period.

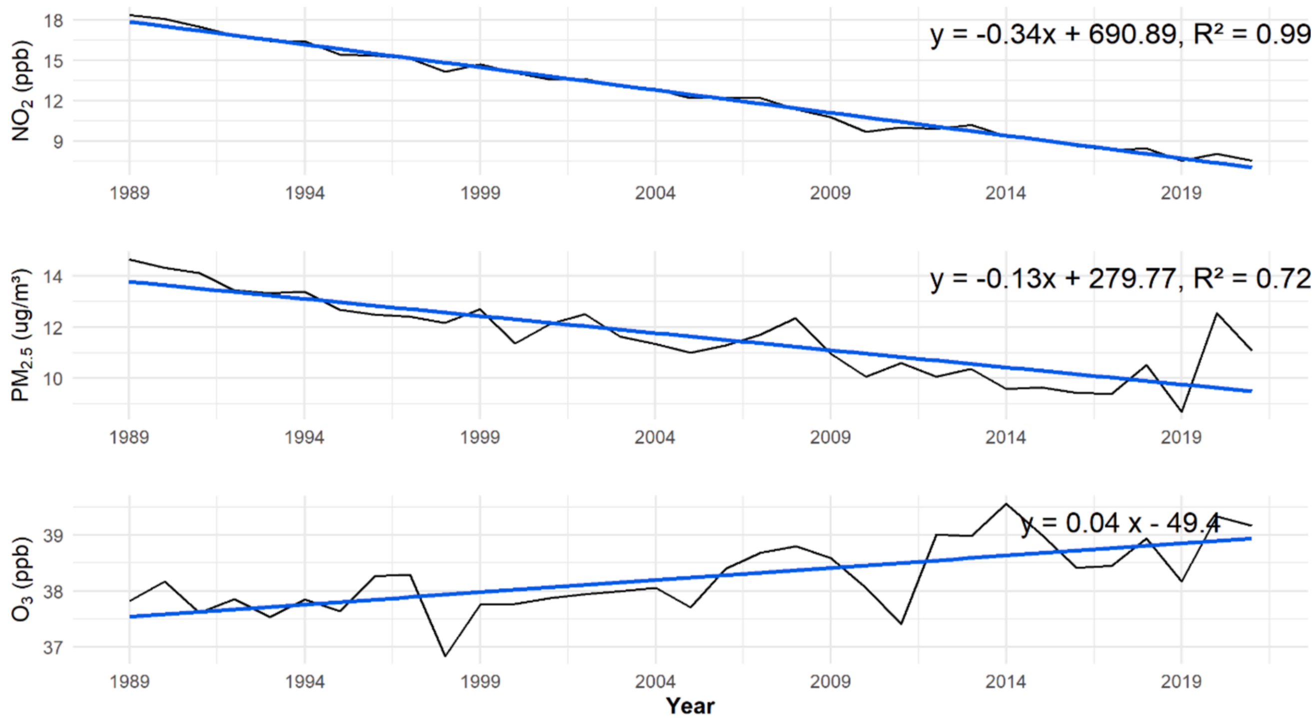

Fig. 5. The modeled historical trends of NO_2 (top), $PM_{2.5}$ (middle) and O_3 (bottom) in California over 30 years. A total of 1,410 points across California, including locations of regulatory stations, saturation monitors, and Google Streetcar mobile monitoring, were used to extract and aggregate modeled air pollution concentrations for the period spanning 1989 to 2021.

respectively, when AutoRegressive Integrated Moving Average (ARIMA) methodology was applied to the residuals of the random forest results. Similarly, Di et al. (2019a) applied an ensemble model combining neural networks, random forests, and gradient boosting to assess NO_2 levels across the U.S., achieving a cross-validated R^2 of 0.788 for daily predictions on 1-km grid cells from 2000 to 2016. In a related study, Di et al. (2019b) used a similar ensemble model for predicting $PM_{2.5}$ levels in the U.S. and obtained a 10-fold cross-validated R^2 of 0.86, outperforming individual models. Requia et al. (2020) further validated the improved performance of ensemble algorithms for O_3 modeling in the U.S., with an overall accuracy of 0.90. These ensemble modeling results in the U.S. are comparable to our model performance, with notably higher accuracy for $PM_{2.5}$. Our daily models, when applied at a 30 m grid resolution, explained 84 %, 65 %, and 92 % of the variations in measured concentrations for NO_2, $PM_{2.5}$, and O_3, respectively, in the 10-fold cross-validation process. Although we could have integrated additional predictors, such as regional factors, to enhance model accuracy, our primary objective was to capture small-area variations in pollutant concentrations. Despite the advantages of ensemble modeling, we opted to use the D/S/A integrated LUR model for our study, primarily due to its interpretability. While the D/S/A model had the potential to incorporate interactions between predictors and employ higher power functions for increased predictive accuracy, we deliberately focused on maintaining linear associations between predictors and measured concentrations and avoid complex interactions. This approach ensured that the expected direction of associations remained clear throughout the model development process. By clearly identifying the factors that significantly contribute to higher concentrations, our models provide valuable insights for policymakers, aiding in the development of effective mitigation strategies. Moreover, the implementation of ensemble models would have required considerably more computational power, particularly given our goal of generating a 100 m resolution daily surface across 33 years for each pollutant, totaling 12,052 days for a single pollutant. Considering the already sufficient predictive performance of our current models, we opted to use interpretable predictors that not only facilitate actionable insights for policymakers but also reduce computational requirements. This approach allowed us to achieve a balance between interpretability and efficiency, ensuring that our models are both practical and effective for informing air quality management decisions.

A primary consideration in this research is the need for a consistent set of predictors across the entire study period. Utilizing a stable framework allows us to assess the influence of these predictors on pollutant concentrations without the confounding effects that might arise from varying model specifications. Moreover, certain variables, such as land use characteristics and geographical features, do not change significantly over time, making it more appropriate to maintain a unified modeling approach. Additionally, while it is possible that model performance could vary across different years, focusing on a long-term model enables us to capture broader trends and patterns that are crucial for understanding air quality dynamics over time. This holistic perspective is essential, particularly in the context of evolving environmental policies and changes in monitoring practices.

While it shares some similarities with traditional stepwise regression in terms of iteratively modifying the model, D/S/A is not a stepwise regression model in the conventional sense. The D/S/A algorithm offers several advantages over traditional stepwise regression, particularly in its flexibility to handle non-linear relationships, interactions between variables, and high-dimensional data. Unlike stepwise regression, which often relies on p-values for variable selection, D/S/A uses a broader set of criteria that are better suited to the complex, high-dimensional nature of our data. Our model evaluation selection process includes cross-validation techniques, which help mitigate the risks associated with overfitting and ensure that the model's predictive performance is robust and generalizable. This approach provides a more reliable assessment of the model's validity compared to relying solely on p-values. It is also important to note that the p-values associated with individual predictors in Tables S2, S3, and S4 and the overall model performance in the D/S/A model were derived after the model was finalized. This process is the same as with linear mixed models, where coefficients and their significance are determined post-model selection. Thus, the p-values reported in Tables S2, S3, and S4 are valid and reflect the significance of the predictors within the context of the finalized model. The D/S/A algorithm's advanced approach, coupled with our V-folder cross-validation, ensures that the model remains robust and valid despite the complexities inherent in the data and the modeling process.

Previous studies have identified a decline in NO_2 and $PM_{2.5}$ concentrations in California (Lurmann et al. 2015; Su et al. 2020; Su et al. 2016) and found that stringent air quality regulations, such as the Clean Air Act (Lurmann et al. 2015; Van Vorst 1997) and California's mobile source regulations (Su et al. 2020), have played a significant role in reducing these pollutants. This study, using all the historical observations, has further confirmed the decrease in those pollutants.

In addition to the impact of policy regulations on the overall reductions in air pollutant concentrations, we found that environmental factors also contribute to pollutant levels. Vegetation was found to be negatively associated with pollutant concentrations, likely due to its ability to absorb pollutants and improve air quality through processes such as phytoremediation and the deposition of particulate matter on plant surfaces (Weyens et al. 2015). Areas with a higher percentage of impervious surfaces, such as roads and buildings, were positively associated with pollutant concentrations (Hatt et al. 2004). This is because impervious surfaces contribute to reduced natural filtration and increased runoff, which can carry pollutants into the air and water (Chithra et al. 2015). Additionally, impervious surfaces represent high levels of human activities such as those from vehicular emissions and industrial activities (Simpson et al. 2022). Traffic density was found to be positively associated with higher pollutant concentrations, especially NO_2 and $PM_{2.5}$ (Li et al. 2015). This is due to the direct emissions from vehicles, which are a major source of these pollutants. Higher temperatures were found to be associated with lower NO_2 concentrations but higher O_3 levels. Higher temperatures facilitate the photochemical reactions that use NO_2 to produce ground-level O_3, leading to decreased NO_2 and increased O_3 levels (Jhun et al. 2015). Wind speed was found to be negatively associated with pollutant concentrations. Stronger winds can disperse pollutants more effectively, diluting their concentrations in the atmosphere (Bhaskar and Mehta 2010). Higher elevations were found to be generally associated with lower concentrations of pollutants such as NO_2 and $PM_{2.5}$ (Su et al. 2020). This could be due to the lower density of emission sources at higher altitudes and more effective atmospheric dispersion. Additionally, pollutants tend to accumulate more in low-lying areas due to atmospheric settling and limited dispersion in valleys (Anderson et al. 2001).

While land use and land cover may appear similar, they represent distinct aspects of the environment, each providing unique insights for modeling. Land use refers to how humans utilize the land, such as residential, commercial, agricultural, or industrial purposes. These variables are critical for understanding sources of pollution linked to human activities. Land cover, on the other hand, describes the physical surface of the land, such as vegetation, water bodies, developed lands and impervious surfaces. It is particularly useful for identifying natural features that influence pollutant dispersion and deposition, such as forested areas that can absorb pollutants or urban heat islands that exacerbate pollution levels. The specific variables from land use and land cover are chosen based on their unique associations with measured pollutant concentrations. For instance, traffic density from land use data may directly correlate with NO_2 levels, while vegetation cover from land cover data may be more relevant for understanding variations in $PM_{2.5}$. These variables are treated as all other predictors, undergoing a holistic selection process where their inclusion is determined by their ability to improve the model's predictive accuracy. By integrating both land use and land cover variables, the model can achieve a more comprehensive and accurate assessment of pollutant sources and their impacts over time. This approach ensures that we capture the full range of factors influencing air quality, enhancing the robustness of our exposure assessments.

While Nighttime Lights (NTL) data is recognized as a valuable predictor in many exposure assessment studies, we did not include it in our analysis due to several specific considerations. Firstly, the spatial and temporal resolution of available NTL data may not align with the high-resolution modeling we employed, potentially leading to discrepancies or reduced accuracy in capturing fine-scale variations in pollutant concentrations. Additionally, NTL data primarily serves as a proxy for human activity, particularly in urban areas, which can be sufficiently captured through other land use variables, such as traffic density and building density, directly integrated into our model. These variables of 30 m spatial resolution offer more precise and context-specific information about pollutant sources related to human activities. Furthermore, land cover data, including % impervious surface and degree of development, inherently represents aspects of NTL data, capturing the extent of urbanization and built environments that are closely associated with light emissions at night. By incorporating these land cover variables of 30 m spatial resolution, we effectively accounted for the spatial patterns that NTL data might indicate. We also applied a rigorous variable selection process, focusing on predictors that demonstrated the strongest association with measured pollutant concentrations in our study area. In this process, other variables were identified as more critical for improving model performance and enhancing the accuracy of our exposure assessments. While NTL data has its merits, we determined that its inclusion would not significantly enhance our model's predictive performance given the spatiotemporal resolution we have from land use and land cover data. Therefore, we prioritized predictors that were most relevant to our study's goals, ensuring a robust and reliable assessment of pollutant exposure.The OMI NO_2 and O_3 datasets are characterized by a coarse resolution of 25 km, which significantly minimizes the occurrence of data gaps. In our analysis, we found that relatively few gaps were detected. To address any gaps that did arise, we implemented a two-round gap-filling algorithm, which involved linear interpolation techniques. The specifics of this process are as follows: If data at a pixel location was available for the day before and the day after a missing value, we calculated the mean of those two values to fill the gap. If only one adjacent day contained effective measurements, we utilized that value to fill the gap. We further use two days before and two days after for any remaining gaps and the data gaps were fully filled after that.

The decision to average AOD data rather than impute missing pixels was driven by practical considerations related to the inherent characteristics of AOD data in California. AOD values exhibit significant day-to-day variability, with large stretches of missing data across the state due to constant cloud impacts. This frequent absence of data diminishes the utility of many days' worth of AOD information across vast regions. Attempting to interpolate these missing values often results in large contiguous areas being assigned the same interpolated values, which may not accurately reflect the true AOD levels. Such interpolation could introduce substantial inaccuracies, undermining the reliability of the exposure assessments. Even after averaging the AOD data on a monthly basis, we still encountered some gaps that required additional processing. To address these remaining gaps, we employed multiple rounds of a 1pixel-by-1pixel smoothing algorithm, which helped fill the holes without compromising the integrity of the data. Moreover, California's climate is characterized by distinct fire and non-fire seasons. During the fire season, significant concentrations of wood smoke contribute to elevated $PM_{2.5}$ levels each month. Even when averaged, these concentrations remain notably higher than during the non-fire season. While daily wildfire concentrations can sometimes reach 300–400 $\mu g/m^3$, the modeling process would likely treat these extreme values as outliers. Averaging AOD data allows us to maintain a balanced representation of these seasonal variations without the distortion that might arise from the direct inclusion of such extreme values. By using monthly averages, we capture the general patterns of AOD while mitigating the risk of skewed results due to significant missing gaps due to cloud and outliers during extreme events.

We incorporated Google Streetcar mobile monitoring data in our research. The Google data complements existing monitoring efforts by filling critical gaps, particularly near highways and densely populated areas, where traditional monitoring stations are often underrepresented. This innovative approach provides a more comprehensive view of air quality dynamics, especially in urban environments where traffic and land use patterns are complex. By leveraging this data, we can better assess exposure patterns and their potential health impacts. The approach of integrating multiple air pollution monitoring types into air quality modeling not only strengthens our findings but also sets a example for future studies to incorporate similar mobile monitoring techniques in air quality research.

The decline in NO_2 concentrations observed in this study reflects the impact of long-term regulatory measures aimed at reducing traffic emissions. The incorporation of remote sensing data, such as OMI-NO_2, proved crucial in capturing spatial variability and enhancing model accuracy. The significant influence of traffic density and weather conditions on NO_2 levels underscores the importance of these factors in air pollution modeling. Moreover, the spatial patterns indicated that urban development and proximity to pollution sources, such as ports, play a critical role in NO_2 distribution.

The study's $PM_{2.5}$ models highlighted the effectiveness of regulatory actions in reducing particulate matter concentrations over time. The integration of AOD data provided valuable insights into the relationship between aerosol presence in the atmosphere and $PM_{2.5}$ levels. The models also demonstrated the mitigating effects of natural features, such as vegetation and water bodies, on $PM_{2.5}$ pollution. However, the increasing frequency and intensity of wildfires pose a significant challenge to sustaining these improvements, as they can lead to spikes in $PM_{2.5}$ levels, especially in vulnerable regions like the Central Valley.

Unlike NO_2 and $PM_{2.5}$, O_3 concentrations did not exhibit a clear long-term trend, reflecting the complex nature of ozone formation and depletion processes. The study's findings suggest that factors such as traffic density and land use patterns significantly influence O_3 levels, with the scavenger effect playing a notable role. The varying influence of meteorological conditions further complicates the prediction and management of O_3 concentrations.

The ability to extend the prediction of daily NO_2, $PM_{2.5}$, and O_3 levels back to 1989 enhances our understanding of long-term air pollution trends. By developing models for predictors such as daily road traffic, NDVI, OMI-NO2, monthly AOD, and OMI-O3, the study successfully estimated historical air pollution levels, providing a comprehensive temporal perspective.

The study, focused on California, leverages data and conditions unique to the state, which, while providing valuable insights, may limit the models' applicability to other regions without significant adjustments. The development of high-resolution (100 m) daily air pollution models over 33 years required substantial computational resources, leading to the use of the D/S/A integrated LUR modeling approach over more resource-intensive methods like ensemble learning. This choice, aimed at ensuring model interpretability and feasibility over an extended temporal scale, may have constrained the exploration of potentially more accurate techniques. The emphasis on linear relationships in the D/S/A integrated LUR models, while enhancing their utility for policymakers, limits the ability to capture complex, non-linear interactions that could improve predictive accuracy. Additionally, extending predictions back to 1989 involved the use of historical predictors and assumptions, introducing uncertainties, particularly for periods with sparse direct measurements, which may affect the accuracy of the backcasted data.Overall, the insights gained from this study are crucial for informing environmental policies and intervention strategies. The identification of pollution hotspots and temporal trends supports efforts to address environmental injustices and protect vulnerable communities. The integration of diverse datasets ensures the robustness of the models, capturing the complex interplay of factors affecting air quality. These findings can guide targeted regulatory actions and public health initiatives, emphasizing the need for continued monitoring and adaptive management in response to emerging challenges such as wildfires.

Funding

This work was supported by the California Air Resources Board (CARB) through funding 21RD004 and 22RD011.

CRediT authorship contribution statement

Jason G. Su: . **Eahsan Shahriary:** Writing – review & editing, Writing – original draft. **Emma Sage:** Writing – review & editing, Writing – original draft. **John Jacobsen:** Writing – review & editing, Writing – original draft. **Katherine Park:** Writing – review & editing, Writing – original draft. **Arash Mohegh:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

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

Acknowledgements

We would like to extend our sincere gratitude to the team at the California Air Resources Board (CARB) for their invaluable and continued support throughout the review and revision process of our project report. Their expertise, feedback, and collaboration played a pivotal role in shaping the content and quality of this manuscript. We deeply appreciate CARB's commitment to advancing environmental research and their dedication to enhancing the understanding of the impacts of air pollution on respiratory health. Their ongoing involvement has been instrumental in bringing this study to fruition. Co-author E. S. is now a research scientist at Stanford University, Palo Alto, United States.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2024.109100.

Data availability

Data will be made available on request.

References

Abatzoglou, J.T., 2013. Development of gridded surface meteorological data for ecological applications and modelling. International Journal of Climatology 33, 121–131.

Anderson, J.; Fernando, H.; Lee, S.; Grossman-Clarke, S.; Pardyjak, E.; Princevac, M. Urban fluid mechanics: air circulation and contaminant dispersion incities. Environmental fluid mechanics 2001;1.

Anenberg, S. Nitrogen Dioxide Surface-Level Annual Average Concentrations V1, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center (GES DISC). 10.5067/J99FI2U38YRN.; 2023.

Beckerman, B.S., Jerrett, M., Martin, R.V., van Donkelaar, A., Ross, Z., Burnett, R.T., 2013a. Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California. Atmospheric Environment 77, 172–177.

Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.-J., van Donkelaar, A., Ross, Z., Su, J., Burnett, R.T., 2013b. A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of PM2.5 in the Contiguous United States. Environmental Science & Technology 47, 7233–7241.

Bhaskar, B.V., Mehta, V.M., 2010. Atmospheric particulate pollutants and their relationship with meteorology in Ahmedabad. Aerosol and Air Quality Research 10, 301–315.

Blackford, A.C. The impact of the 2018 camp fire on land-atmosphere interactions. 2024.

Brokamp, C., Jandarov, R., Hossain, M., Ryan, P., 2018. Predicting daily urban fine particulate matter concentrations using a random forest model. Environmental Science & Technology 52, 4173–4179.

Brown, P.T., Hanley, H., Mahesh, A., Reed, C., Strenfel, S.J., Davis, S.J., Kochanski, A.K., Clements, C.B., 2023. Climate warming increases extreme daily wildfire growth risk in California. Nature 621, 760–766.

Cabaneros, S.M., Calautit, J.K., Hughes, B.R., 2019. A review of artificial neural network models for ambient air pollution prediction. Environmental Modelling & Software 119, 285–304.

Castelli, M., Clemente, F.M., Popovič, A., Silva, S., Vanneschi, L., 2020. A machine learning approach to predict air quality in California. Complexity 2020, 8049504.

Chithra, S., Nair, M.H., Amarnath, A., Anjana, N., 2015. Impacts of impervious surfaces on the environment. International Journal of Engineering Science Invention 4, 27–31.

Chuang, G.C., Yang, Z., Westbrook, D.G., Pompilius, M., Ballinger, C.A., White, C.R., Krzywanski, D.M., Postlethwait, E.M., Ballinger, S.W., 2009. Pulmonary ozone exposure induces vascular dysfunction, mitochondrial damage, and atherogenesis. American Journal of Physiology-Lung Cellular and Molecular Physiology 297, L209–L216.

Correia, A.W., Pope III, C.A., Dockery, D.W., Wang, Y., Ezzati, M., Dominici, F., 2013. The effect of air pollution control on life expectancy in the United States: an analysis of 545 US counties for the period 2000 to 2007. Epidemiology (cambridge, Mass) 24, 23.

Craig, L., Brook, J.R., Chiotti, Q., Croes, B., Gower, S., Hedley, A., Krewski, D., Krupnick, A., Krzyzanowski, M., Moran, M.D., 2008. Air pollution and public health: a guidance document for risk managers. Journal of Toxicology and Environmental Health, Part A 71, 588–698.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., 2019a. Assessing NO2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. Environmental Science & Technology 54, 1372–1384.

Di, Q.; Amini, H.; Shi, L.; Kloog, I.; Silvern, R.; Kelly, J.; Sabath, M.B.; Choirat, C.; Koutrakis, P.; Lyapustin, A. An ensemble-based model of PM2. 5 concentration across the contiguous United States with high spatiotemporal resolution. Environment international 2019b;130:104909.

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., Dominici, F., Schwartz, J.D., 2017. Air pollution and mortality in the Medicare population. New England Journal of Medicine 376, 2513–2522.

Fasola, S., Maio, S., Baldacci, S., La Grutta, S., Ferrante, G., Forastiere, F., Stafoggia, M., Gariazzo, C., Viegi, G., Group, B.C., 2020. Effects of particulate matter on the incidence of respiratory diseases in the pisan longitudinal study. International Journal of Environmental Research and Public Health 17, 2540.

Fuller, R., Landrigan, P.J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., 2022. Pollution and health: a progress update. The Lancet Planetary Health 6, e535–e547.

Gauderman, W.J., Urman, R., Avol, E., Berhane, K., McConnell, R., Rappaport, E., Chang, R., Lurmann, F., Gilliland, F., 2015. Association of improved air quality with lung development in children. New England Journal of Medicine 372, 905–913.

Giles, L.V., Barn, P., Künzli, N., Romieu, I., Mittleman, M.A., van Eeden, S., Allen, R., Carlsten, C., Stieb, D., Noonan, C., 2011. From good intentions to proven interventions: effectiveness of actions to reduce the health impacts of air pollution. Environmental Health Perspectives 119, 29–36.

Gocheva-Ilieva, S.G., Ivanov, A.V., Livieris, I.E., 2020. High performance machine learning models of large scale air pollution data in urban area. Cybernetics and Information Technologies 20, 49–60.

Ha, S., Hu, H., Roussos-Ross, D., Haidong, K., Roth, J., Xu, X., 2014. The effects of air pollution on adverse birth outcomes. Environmental Research 134, 198–204.

Hatt, B.E., Fletcher, T.D., Walsh, C.J., Taylor, S.L., 2004. The influence of urban density and drainage infrastructure on the concentrations and loads of pollutants in small streams. Environmental Management 34, 112–124.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 42, 7561–7578.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. Photogrammetric Engineering & Remote Sensing 81, 345–354.

Horne, B.D., Joy, E.A., Hofmann, M.G., Gesteland, P.H., Cannon, J.B., Lefler, J.S., Blagev, D.P., Korgenski, E.K., Torosyan, N., Hansen, G.I., 2018. Short-term elevation of fine particulate matter air pollution and acute lower respiratory infection. American Journal of Respiratory and Critical Care Medicine 198, 759–766.

Houston, D., Wu, J., Ong, P., Winer, A., 2004. Structural disparities of urban traffic in Southern California: implications for vehicle-related air pollution exposure in minority and high-poverty neighborhoods. Journal of Urban Affairs 26, 565–592.

Hu, J.; Zhang, H.; Chen, S.-H.; Wiedinmyer, C.; Vandenberghe, F.; Ying, Q.; Kleeman, M. J. Predicting primary PM2. 5 and PM0. 1 trace composition for epidemiological studies in California. Environmental science & technology 2014;48:4971-4979.

Jerrett, M., Burnett, R.T., Ma, R., Pope III, C.A., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E.E., 2005. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 16, 727–736.

Jhun, I., Coull, B.A., Zanobetti, A., Koutrakis, P., 2015. The impact of nitrogen oxides concentration decreases on ozone trends in the USA. Air Quality, Atmosphere & Health 8, 283–292.

Jones, R.R., Hoek, G., Fisher, J.A., Hasheminassab, S., Wang, D., Ward, M.H., Sioutas, C., Vermeulen, R., Silverman, D.T., 2020. Land use regression models for ultrafine particles, fine particles, and black carbon in Southern California. Science of the Total Environment 699, 134234.

Kanaroglou, P.S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M.A., Gilbert, N.L., Brook, J.R., 2005. Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. Atmospheric Environment 39, 2399–2409.

Kaufman, J.D., Elkind, M.S., Bhatnagar, A., Koehler, K., Balmes, J.R., Sidney, S., Burroughs Peña, M.S., Dockery, D.W., Hou, L., Brook, R.D., 2020. Guidance to reduce the cardiovascular burden of ambient air pollutants: a policy statement from the American Heart Association. Circulation 142, e432–e447.

Keeley, J.E., Syphard, A.D., 2021. Large California wildfires: 2020 fires in historical context. Fire Ecology 17, 1–11.

Kerckhoffs, J.; Hoek, G.; Portengen, L.t.; Brunekreef, B.; Vermeulen, R.C. Performance of prediction algorithms for modeling outdoor air pollution spatial surfaces. Environmental science & technology 2019;53:1413-1421.

Kinney, P.L. The pulmonary effects of outdoor ozone and particle air pollution. Seminars in respiratory and critical care medicine: Copyright© 1999 by Thieme Medical Publishers, Inc.; 1999.

Kumar, D., 2018. Evolving Differential evolution method with random forest for prediction of Air Pollution. Procedia Computer Science 132, 824–833.

Kumar, U., Prakash, A., Jain, V., 2008. A photochemical modelling approach to investigate O3 sensitivity to NOx and VOCs in the urban atmosphere of Delhi. Aerosol and Air Quality Research 8, 147–159.

Larsen, L.C.; Sacramento, C. The ozone weekend effect in California: evidence supporting NOx emission reductions. CARB report, http://www arb ca gov 2003.

Lee, H.J., Chatfield, R.B., Strawa, A.W., 2016. Enhancing the applicability of satellite remote sensing for PM2. 5 estimation using MODIS deep blue AOD and land use regression in California. United States. Environmental Science & Technology 50, 6546–6555.

Leong, W., Kelani, R., Ahmad, Z., 2020. Prediction of air pollution index (API) using support vector machine (SVM). Journal of Environmental Chemical Engineering 8, 103208.

Levelt, P.F., Joiner, J., Tamminen, J., Veefkind, J.P., Bhartia, P.K., Stein Zweers, D.C., Duncan, B.N., Streets, D.G., Eskes, H., van der A, R, 2018. The Ozone Monitoring Instrument: overview of 14 years in space. Atmospheric Chemistry and Physics 18, 5699–5745.

Li, S., Banerjee, T., 2021. Spatial and temporal pattern of wildfires in California from 2000 to 2019. Sci Rep-Uk 11, 8779.

Li, Q.; Chen, C.; Deng, Y.; Li, J.; Xie, G.; Li, Y.; Hu, Q. Influence of traffic force on pollutant dispersion of CO, NO and particle matter (PM2. 5) measured in an urban tunnel in Changsha, China. Tunnelling and Underground Space Technology 2015;49: 400-407.

Liu, J., Marshall, J.D., 2023. Spatial decomposition of air pollution concentrations highlights historical causes for current exposure disparities in the United States. Environmental Science & Technology Letters 10, 280–286.

Lunetta, R.S.; Knight, J.F.; Ediriwickrema, J.; Lyon, J.G.; Worthy, L.D. Land-cover change detection using multi-temporal MODIS NDVI data. Geospatial Information Handbook for Water Resources and Watershed Management, Volume II: CRC Press; 2022.

Lurmann, F., Avol, E., Gilliland, F., 2015. Emissions reduction policies and recent trends in Southern California's ambient air quality. Journal of the Air & Waste Management Association 65, 324–335.

Masood, A., Ahmad, K., 2021. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. Journal of Cleaner Production 322, 129072.

Mecklin, J., 2014. California here we come? B Atom Sci 70, 24–25.

Monn, C., Carabias, V., Junker, M., Waeber, R., Karrer, M., Wanner, H.-U., 1997. Small-scale spatial variability of particulate matter< 10 μm (PM10) and nitrogen dioxide. Atmospheric Environment 31, 2243–2247.

Moore, D., Jerrett, M., Mack, W., Künzli, N., 2007. A land use regression model for predicting ambient fine particulate matter across Los Angeles. CA. Journal of Environmental Monitoring 9, 246–252.

Morello-Frosch, R., Jesdale, B.M., 2006. Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in US metropolitan areas. Environmental Health Perspectives 114, 386–393.

Morello-Frosch, R., Pastor Jr, M., Porras, C., Sadd, J., 2002. Environmental justice and regional inequality in southern California: implications for future research. Environmental Health Perspectives 110, 149–154.

Naidoo, R.N., 2019. NO2 increases the risk for childhood asthma: a global concern. The Lancet Planetary Health 3, e155–e156.

Ostro, B., Rauch, S., Green, R., Malig, B., Basu, R., 2010. The effects of temperature and use of air conditioning on hospitalizations. American Journal of Epidemiology 172, 1053–1061.

Peng, Z., Zhang, B., Wang, D., Niu, X., Sun, J., Xu, H., Cao, J., Shen, Z., 2023. Application of machine learning in atmospheric pollution research: A state-of-art review. Science of the Total Environment, 168588.

Pope 3rd, C., 2000. Epidemiology of fine particulate air pollution and human health: biologic mechanisms and who's at risk? Environmental Health Perspectives 108, 713–723.

Pope III, C.A., Burnett, R.T., Thurston, G.D., Thun, M.J., Calle, E.E., Krewski, D., Godleski, J.J., 2004. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. Circulation 109, 71–77.

Pope III, C.A., Burnett, R.T., Krewski, D., Jerrett, M., Shi, Y., Calle, E.E., Thun, M.J., 2009a. Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke: shape of the exposure-response relationship. Circulation 120, 941–948.

Pope III, C.A., Ezzati, M., Dockery, D.W., 2009b. Fine-particulate air pollution and life expectancy in the United States. New Engl J Med 360, 376–386.
Pope III, C.A., Turner, M.C., Burnett, R.T., Jerrett, M., Gapstur, S.M., Diver, W.R., Krewski, D., Brook, R.D., 2015. Relationships between fine particulate air pollution, cardiometabolic disorders, and cardiovascular mortality. Circulation Research 116, 108–115.
Pope III, C.A., Cohen, A.J., Burnett, R.T., 2018. Cardiovascular disease and fine particulate matter: lessons and limitations of an integrated exposure–response approach. Circulation Research 122, 1645–1647.
Reid, C.E., Jerrett, M., Petersen, M.L., Pfister, G.G., Morefield, P.E., Tager, I.B., Raffuse, S.M., Balmes, J.R., 2015. Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning. Environmental Science & Technology 49, 3887–3896.
Ren, X., Mi, Z., Georgopoulos, P.G., 2020. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. Environment International 142, 105827.
Requia, W.J., Di, Q., Silvern, R., Kelly, J.T., Koutrakis, P., Mickley, L.J., Sulprizio, M.P., Amini, H., Shi, L., Schwartz, J., 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. Environmental Science & Technology 54, 11037–11047.
Rich, D.Q., Thurston, S.W., Balmes, J.R., Bromberg, P.A., Arjomandi, M., Hazucha, M.J., Alexis, N.E., Ganz, P., Zareba, W., Thevenet-Morrison, K., 2020. Do ambient ozone or other pollutants modify effects of controlled ozone exposure on pulmonary function? Annals of the American Thoracic Society 17, 563–572.
Rooney, B., Wang, Y., Jiang, J.H., Zhao, B., Zeng, Z.-C., Seinfeld, J.H., 2020. Air quality impact of the Northern California camp fire of November 2018. Atmospheric Chemistry and Physics 20, 14597–14616.
Ross, Z., English, P.B., Scalf, R., Gunier, R., Smorodinsky, S., Wall, S., Jerrett, M., 2006. Nitrogen dioxide prediction in Southern California using land use regression modeling: potential for environmental health analyses. Journal of Exposure Science & Environmental Epidemiology 16, 106–114.
Ryan, P.H., LeMasters, G.K., 2007. A review of land-use regression models for characterizing intraurban air pollution exposure. Inhalation Toxicology 19, 127–133.
Safford, H.D., Paulson, A.K., Steel, Z.L., Young, D.J., Wayman, R.B., 2022. The 2020 California fire season: A year like no other, a return to the past or a harbinger of the future? Global Ecology and Biogeography 31, 2005–2025.
Simpson, I.M., Winston, R.J., Brooker, M.R., 2022. Effects of land use, climate, and imperviousness on urban stormwater quality: A meta-analysis. Science of the Total Environment 809, 152206.
Studnicka, M., Hackl, E., Pischinger, J., Fangmeyer, C., Haschke, N., Kuhr, J., Urbanek, R., Neumann, M., Frischer, T., 1997. Traffic-related NO2 and the prevalence of asthma and respiratory symptoms in seven year olds. European Respiratory Journal 10, 2275–2278.
Su, J.G., Jerrett, M., Beckerman, B., Wilhelm, M., Ghosh, J.K., Ritz, B., 2009. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. Environmental Research 109, 657–670.
Su, J.G., Jerrett, M., Meng, Y.-Y., Pickett, M., Ritz, B., 2015b. Integrating smart-phone based momentary location tracking with fixed site air quality monitoring for personal exposure assessment. Science of the Total Environment 506, 518–526.
Su, J.G., Hopke, P.K., Tian, Y., Baldwin, N., Thurston, S.W., Evans, K., Rich, D.Q., 2015a. Modeling particulate matter concentrations measured through mobile monitoring in a deletion/substitution/addition approach. Atmos Environ 122, 477–483.
Su, J.G., Meng, Y.-Y., Pickett, M., Seto, E., Ritz, B., Jerrett, M., 2016. Identification of effects of regulatory actions on air quality in goods movement corridors in California. Environmental Science & Technology 50, 8687–8696.
Su, J.G., Meng, Y.-Y., Chen, X., Molitor, J., Yue, D., Jerrett, M., 2020. Predicting differential improvements in annual pollutant concentrations and exposures for regulatory policy assessment. Environ Int 143, 105942.
Tack, F., Merlaud, A., Iordache, M.-D., Danckaert, T., Yu, H., Fayt, C., Meuleman, K., Deutsch, F., Fierens, F., Van Roozendael, M., 2017. High-resolution mapping of the NO 2 spatial distribution over Belgian urban areas based on airborne APEX remote sensing. Atmospheric Measurement Techniques 10, 1665–1688.
Van Donkelaar, A., 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. Environ Sci Technol 53, 2595–2611.
Van Vorst, W.D., 1997. Impact of the California clean air act. International Journal of Hydrogen Energy 22, 31–38.
Wang, Y.; Bechle, M.J.; Kim, S.-Y.; Adams, P.J.; Pandis, S.N.; Pope III, C.A.; Robinson, A. L.; Sheppard, L.; Szpiro, A.A.; Marshall, J.D. Spatial decomposition analysis of NO2 and PM2. 5 air pollution in the United States. Atmospheric environment 2020;241: 117470.
Wang, W. Investigations of the atmospheric oxidative capacity with chemical ionization mass spectrometry and chemical box model. 2020.
Weyens, N., Thijs, S., Popek, R., Witters, N., Przybysz, A., Espenshade, J., Gawronska, H., Vangronsveld, J., Gawronski, S.W., 2015. The role of plant–microbe interactions and their exploitation for phytoremediation of air pollutants. International Journal of Molecular Sciences 16, 25576–25604.
Wikipedia contributors. “California.” Wikipedia, The Free Encyclopedia, August 20, 2024. https://en.wikipedia.org/wiki/California.; 2024.
Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., 2018. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. ISPRS Journal of Photogrammetry and Remote Sensing 146, 108–123.
Yin, P., Brauer, M., Cohen, A.J., Wang, H., Li, J., Burnett, R.T., Stanaway, J.D., Causey, K., Larson, S., Godwin, W., 2020. The effect of air pollution on deaths, disease burden, and life expectancy across China and its provinces, 1990–2017: an analysis for the Global Burden of Disease Study 2017. The Lancet Planetary Health 4, e386–e398.
Zhang, P.; Ma, W.; Wen, F.; Liu, L.; Yang, L.; Song, J.; Wang, N.; Liu, Q. Estimating PM2. 5 concentration using the machine learning GA-SVM method to improve the land use regression model in Shaanxi, China. Ecotoxicology and Environmental Safety 2021; 225:112772.
Zhang, H., Lyapustin, A., Wang, Y., Kondragunta, S., Laszlo, I., Ciren, P., Hoff, R., 2011. A multi-angle aerosol optical depth retrieval algorithm for geostationary satellite data over the United States. Atmospheric Chemistry and Physics 11, 11977–11991.
Zou, B., Peng, F., Wan, N., Mamady, K., Wilson, G.J., 2014. Spatial cluster detection of air pollution exposure inequities across the United States. PLoS One 9, e91917.