Spatiotemporal Modeling of California Air Pollution: Implications for Environmental Justice

Alvin Chen
University High School, Irvine, CA

Abstract

Disproportionate exposure to air pollution in low-income and minority communities remains a central concern in environmental justice and public health. This study applies Geographically and Temporally Weighted Regression and kriging interpolation to 2024 EPA Air Quality System data, examining four pollutants ($PM_{2.5}$, NO_2, O_3, CO) from 27 monitoring stations across California with a focus on the influence of proximity to major ports. Results reveal pronounced spatial disparities: in EJ-designated tracts of Los Angeles, San Bernardino, and San Joaquin counties, $PM_{2.5}$ levels were 77.1% higher than in non-EJ areas, accompanied by nearly double the strength of negative correlation with port distance. Distinct spatial patterns for ozone and carbon monoxide were also observed. Overall, the findings underscore the importance of localized spatiotemporal modeling and accounting for multiple pollution sources in advancing environmental justice research.

Keywords: Air quality, environmental justice, pollutants, kriging, geographically and temporally weighted regression

1 Introduction

1.1 Background

Environmental justice refers to the equitable treatment and meaningful involvement of all people—regardless of race, ethnicity, national origin, or income—in the development and enforcement of environmental policies, laws, and regulations [1]. The movement emerged from grassroots activism in the 1980s and highlighted how communities of color and economically disadvantaged populations disproportionately bear the burden of pollution and environmental degradation [2].

Air pollution is among the most well-documented environmental justice concerns. Numerous studies show that minority and low-income communities are consistently exposed to higher levels of harmful pollutants [3, 4]. These disparities stem from multiple overlapping causes: proximity to industrial zones, historical segregation and discriminatory land-use practices, limited political representation, and economic barriers to relocation [5].

The health impacts of pollutants such as $PM_{2.5}$, NO_2, O_3, and CO are profound. They include increased risks of respiratory and cardiovascular disease, premature mortality, and adverse developmental effects in children [6, 7]. Environmental justice communities are further burdened by limited access to healthcare and overlapping exposure to multiple pollutants [8].

Throughout this study, we define environmental justice communities as geographic areas that experience both elevated pollution burdens and socioeconomic vulnerability, often indicated by low household income or a high percentage of minority residents.

Transportation infrastructure, particularly ports, highways, and freight corridors, is a critical contributor to environmental health disparities. Port operations produce substantial emissions from ships, trucks, trains, and support industries [9]. Communities situated near ports often experience elevated exposure to diesel particulates, nitrogen oxides, and sulfur compounds [10].

California is a compelling setting for environmental justice studies. It has one of the most demographically diverse populations in the U.S., a complex geography, and several of the nation's largest port complexes—including Los Angeles, Long Beach, Oakland, and San Diego—each generating distinct emissions patterns [11]. These conditions, combined with the state's Mediterranean climate and topography, create localized pollution issues that demand nuanced analysis.

Policy attention has grown in recent years. California Assembly Bill 617 (2017) mandates localized air monitoring and emissions reductions in overburdened communities [12]. The CalEnviroScreen tool classifies vulnerable communities based on pollution burden and demographic characteristics, helping guide resource allocation [13].

Yet many studies still rely on global models that assume spatial uniformity. These models may mask local variations in pollution exposure and demographic vulnerability [14], undermining both scientific insight and policy effectiveness.

1.2 Literature Review

Environmental justice research has evolved from early descriptive analyses to sophisticated modeling of environmental disparities. Initial studies relied on coarse-scale data and global regressions to assess the relationship between demographics and pollution [15, 16].

Seminal reports by the United Church of Christ [17] and Bullard [1] documented how communities of color were disproportionately burdened by toxic facilities. However, methodological limitations in early work—such as poor spatial resolution and weak controls for confounding variables—led to debate over whether race or class was the primary driver of these disparities [18].

By the 2000s, improvements in geographic information systems (GIS), pollution monitoring, and statistical methods allowed researchers to examine pollution-demographic relationships at finer scales [19]. Jerrett et al. [20] emphasized the need for spatially-sensitive methods, such as Geographically Weighted Regression (GWR), to capture local variation.

GWR and related spatial models allowed researchers to identify significant heterogeneity in how pollution correlates with demographic characteristics across different regions [21]. Mennis [22] applied GWR in New Jersey and found substantial spatial differences that global models failed to detect.

Recent studies have emphasized cumulative exposure to multiple pollutants and the

need for integrated assessments [23, 24]. Transportation-related air pollution remains a major focus due to its disproportionate impact on EJ communities. Studies by Houston et al. [10] and Pastor et al. [25] have shown how ports and freight corridors contribute to persistent health disparities.

The development of Geographically and Temporally Weighted Regression (GTWR) by Fotheringham et al. [26] added a temporal dimension to GWR, capturing how pollution relationships evolve over time—a key advancement for policy-relevant EJ research.

Kriging and other geostatistical methods also gained prominence, allowing for spatial interpolation of pollution values and associated uncertainty estimation [14, 27]. These methods support fine-grained exposure mapping essential for environmental health equity.

Despite these advances, most studies still focus on single-source proximity (often just highways or one port), overlooking the diverse pollution sources that affect California. Bailey et al. [9] and Boone and Modarres [28] call for more sophisticated proximity analysis frameworks that include multiple sources.

This study aims to fill that gap by integrating GTWR and kriging with a multi-port proximity assessment, offering new insights into the spatial dynamics of environmental justice in California.

2 Geospatial Analysis Data

This study utilized comprehensive air quality monitoring data from the U.S. Environmental Protection Agency's Air Quality System (AQS), focusing on California monitoring stations during 2024 [29]. The dataset encompasses four criteria pollutants representing diverse pollution sources and formation mechanisms: fine particulate matter ($PM_{2.5}$) from combustion and secondary formation processes, nitrogen dioxide (NO_2) primarily from vehicle and industrial emissions, ground-level ozone (O_3) from photochemical reactions, and carbon monoxide (CO) from incomplete combustion sources.

2.1 Air Quality Monitoring Data

The EPA AQS database provides quality-assured ambient air monitoring data collected by state, local, and tribal air pollution control agencies using standardized monitoring protocols and quality assurance procedures [30]. For this analysis, we extracted daily average concentrations for each pollutant from California monitoring stations (State Code = 6) for the period January 1, 2024, through October 31, 2024. Table 1 presents a summary of the monitoring data characteristics.

Table 1: Summary of EPA Air Quality Monitoring Data for California (2024)

Pollutant	Observations	Counties	Sites	Date Range
$PM_{2.5}$	37,173	42	64	Jan 1 - Oct 31
NO_2	14,119	30	59	Jan 1 - Aug 31
O_3	24,914	45	74	Jan 1 - Sep 30
CO	14,581	21	35	Jan 1 - Sep 1

The integrated dataset, after merging pollutants by location and date and applying quality control procedures, comprised 125 site-month observations across 27 unique

monitoring locations, providing comprehensive spatial coverage of California's diverse air quality conditions from the South Coast Air Basin to the San Francisco Bay Area and Central Valley.

2.2 Demographic and Environmental Justice Data

Environmental justice analysis requires integration of air quality data with demographic and socioeconomic indicators. We compiled county-level demographic data from the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates [31], including median household income, minority population percentage, and total population. Environmental justice communities were identified using EPA guidance and California's CalEnviroScreen methodology, defined as areas with both high pollution burden and vulnerable demographics characterized by low income or high minority percentage.

2.3 Multi-Port Proximity Analysis

A critical component of this study involved a comprehensive analysis of proximity to major California port facilities as an environmental justice factor. We systematically examined proximity effects for six major California ports:

1. **Port of Los Angeles/Long Beach Complex** (33.74°N, 118.25°W) - The largest port complex in the United States, handling primarily containerized cargo
2. **Port of Oakland** (37.80°N, 122.32°W) - Major San Francisco Bay Area container port serving northern California markets
3. **Port of San Francisco** (37.79°N, 122.42°W) - Bay Area general cargo, cruise, and specialty port facility
4. **Port of Richmond** (37.93°N, 122.38°W) - Bay Area bulk cargo and petroleum products facility
5. **Port of Stockton** (37.95°N, 121.29°W) - Central Valley inland port serving agricultural and manufacturing regions
6. **Port of San Diego** (32.71°N, 117.17°W) - Southern California general cargo and naval facility

This comprehensive multi-port approach recognizes that California's diverse geography and extensive port infrastructure require sophisticated proximity analysis.

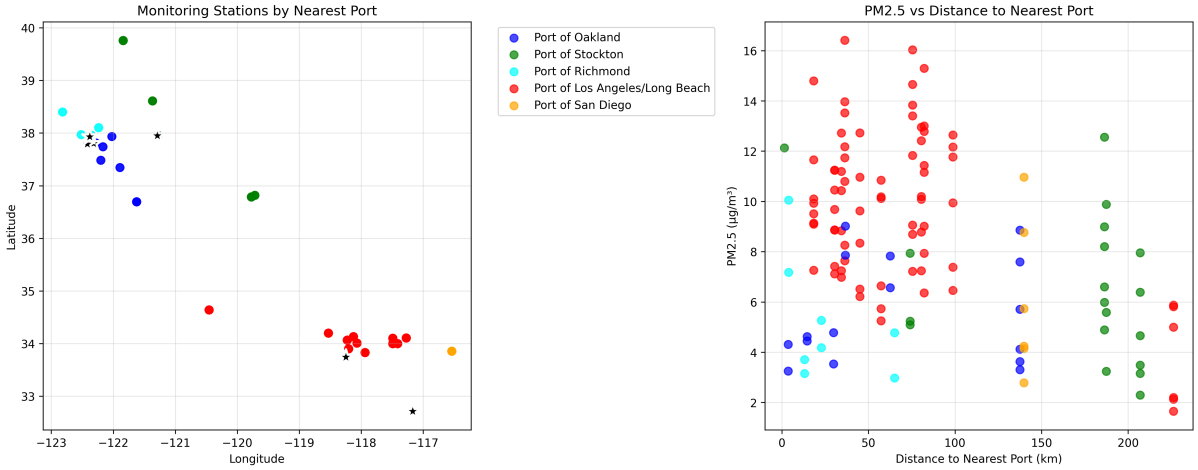

Figure 1: Comprehensive multi-port proximity analysis across California's major port facilities. Panel (a) shows monitoring stations colored by nearest port, demonstrating regional clustering around different port facilities. Panel (b) shows $PM_{2.5}$ concentrations versus distance to nearest port, revealing pollution relationships using our comprehensive port proximity methodology. GTWR analysis using nearest port distances demonstrates strong model performance ($R^2 = 0.584$) and reveals significant regional variations in environmental justice impacts.

2.4 Data Processing and Quality Control

Data processing involved several quality control steps to ensure analytical reliability and spatial accuracy. Raw pollutant concentrations were filtered to remove negative values, extreme outliers (beyond the 99.5th percentile), and measurements flagged for quality issues. Temporal aggregation was performed to create monthly averages by monitoring site, reducing measurement noise while preserving spatiotemporal patterns relevant for GTWR analysis.

Missing data were handled through listwise deletion to maintain data integrity for the integrated analysis, with sensitivity analysis confirming that missing data patterns did not introduce systematic bias. Coordinate transformation was performed using the pyproj library to convert geographic coordinates (WGS84) to the California State Plane coordinate system (EPSG:3310) for accurate distance calculations and spatial analysis.

Port proximity distances were calculated using the Haversine formula to compute great-circle distances between monitoring stations and port facilities:

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

where r is Earth's radius (6,371.0088 km), ϕ represents latitude, λ represents longitude, and Δ indicates coordinate differences between monitoring stations and port locations.

For each monitoring station, we calculated distances to all six major port facilities and identified the nearest port for optimal proximity analysis. This comprehensive proximity analysis revealed that 76 stations (61%) were closest to Los Angeles area ports, while 16 stations were closest to Oakland, 19 to Stockton, 8 to Richmond, and 6 to San Diego ports.

3 Kriging

Kriging interpolation provides optimal spatial prediction of environmental variables based on the principle of spatial autocorrelation, where nearby locations tend to have similar values. This geostatistical method is particularly valuable for environmental analysis as it provides both predicted values and associated uncertainty estimates, enabling comprehensive spatial risk assessment.

3.1 Theoretical Framework

Kriging is a family of geostatistical interpolation methods that predict values at unsampled locations based on weighted averages of nearby observations, with weights determined by the spatial correlation structure of the data [27]. The method assumes that the spatial variation of a regionalized variable can be characterized by a semivariogram, which describes the degree of spatial dependence as a function of distance.

For a regionalized variable $Z(s)$ observed at locations $s_1, s_2, \ldots, s_n$, ordinary kriging provides the Best Linear Unbiased Predictor (BLUP) at any unsampled location s_0:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

where λ_i are the kriging weights determined by solving the kriging system:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix}$$

where γ_{ij} represents the semivariogram value between locations s_i and s_j, and μ is the Lagrange multiplier ensuring unbiasedness.

The semivariogram is typically modeled using theoretical functions such as the spherical model:

$$\gamma(h) = \begin{cases} C_0 + C\left[\frac{3h}{2a} - \frac{h^3}{2a^3}\right] & \text{if } h \leq a \\ C_0 + C & \text{if } h > a \end{cases}$$

where h is the lag distance, C_0 is the nugget effect representing measurement error and micro-scale variation, C is the sill representing the total variance, and a is the range parameter indicating the distance at which spatial correlation becomes negligible.

The kriging variance provides a measure of prediction uncertainty:

$$\sigma_K^2(s_0) = \sum_{i=1}^{n} \lambda_i \gamma_{i0} + \mu.$$

This theoretical framework enables both optimal spatial prediction and quantification of prediction uncertainty, which is crucial for environmental risk assessment and policy decision-making in environmental justice contexts.

3.2 Applications

We implemented ordinary kriging using the PyKrige library in Python, applying the methodology to create comprehensive pollution surfaces across California. The analysis encompassed multiple spatial scales and temporal periods to capture diverse aspects of air quality patterns and their environmental justice implications.

Multi-Pollutant Spatial Analysis Comprehensive kriging analysis was performed for all four pollutants ($PM_{2.5}$, NO_2, O_3, CO) using site-averaged concentrations across the study period. Spherical semivariogram models were fitted for each pollutant through maximum likelihood estimation, with model parameters optimized using cross-validation procedures to ensure robust spatial prediction performance.

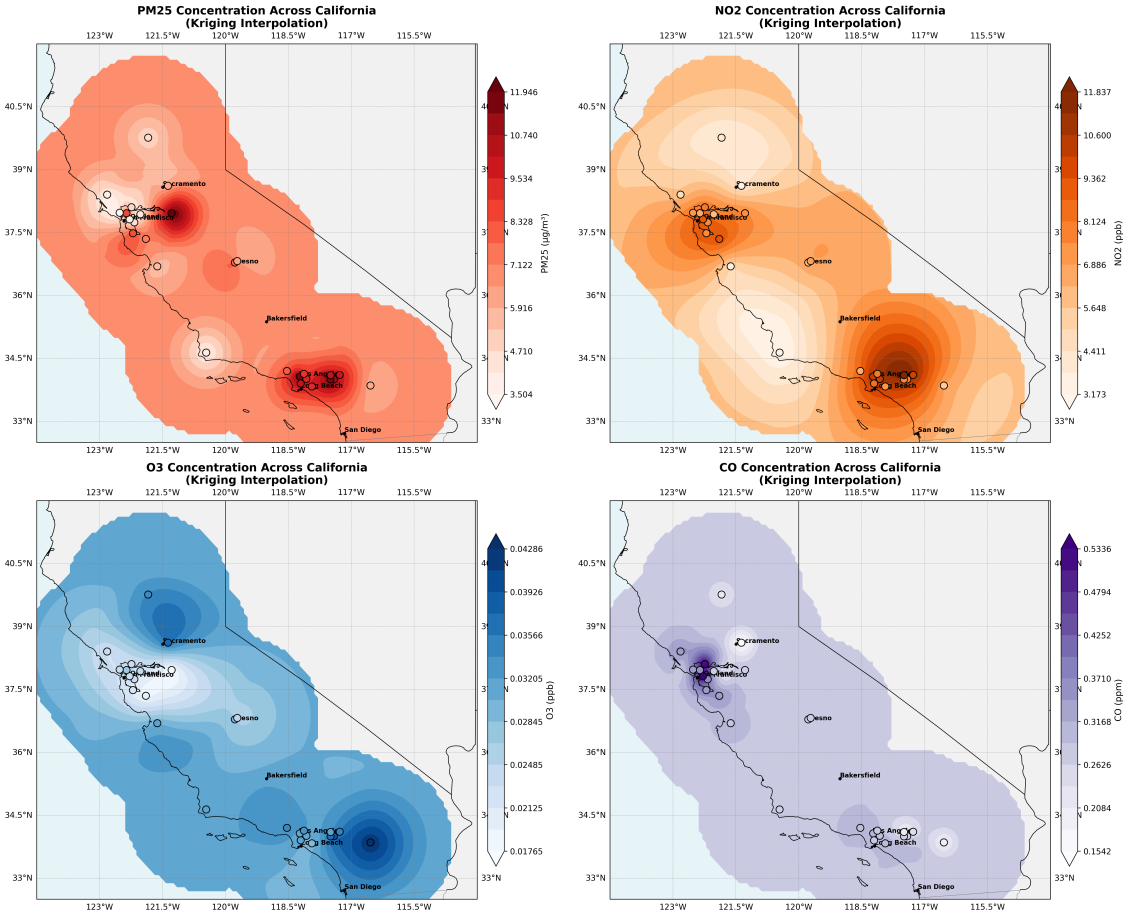

Figure 2: Comprehensive kriging interpolation maps for four criteria pollutants across California. Each panel shows spatial concentration patterns with monitoring station locations and EPA health standard reference lines where applicable. The analysis reveals distinct spatial patterns reflecting different pollution sources and atmospheric processes, with $PM_{2.5}$ and NO_2 showing urban concentration patterns while O_3 exhibits more complex regional distribution.

The kriging analysis revealed distinct spatial patterns for each pollutant that provide insights into environmental justice implications. $PM_{2.5}$ concentrations showed relatively uniform distribution across the study region (3.42-12.13 $\mu g/m^3$), with slight elevation

in urban areas and near major transportation corridors. The spatial pattern revealed persistent gradients that may contribute to differential exposure across communities.

NO_2 exhibited pronounced spatial gradients with higher concentrations in urban areas, particularly near major highways and industrial facilities, reflecting the importance of traffic-related emissions. O_3 patterns demonstrated the complex photochemical nature of this secondary pollutant, with concentrations varying based on precursor availability, meteorological conditions, and atmospheric transport processes. CO showed localized elevation near major emission sources, particularly in urban areas with heavy traffic congestion.

Temporal Evolution Analysis To examine temporal patterns in pollution distribution and their implications for environmental justice assessment, we performed kriging analysis for three representative months (January, April, and June 2024), enabling assessment of seasonal variations and temporal consistency of spatial patterns.

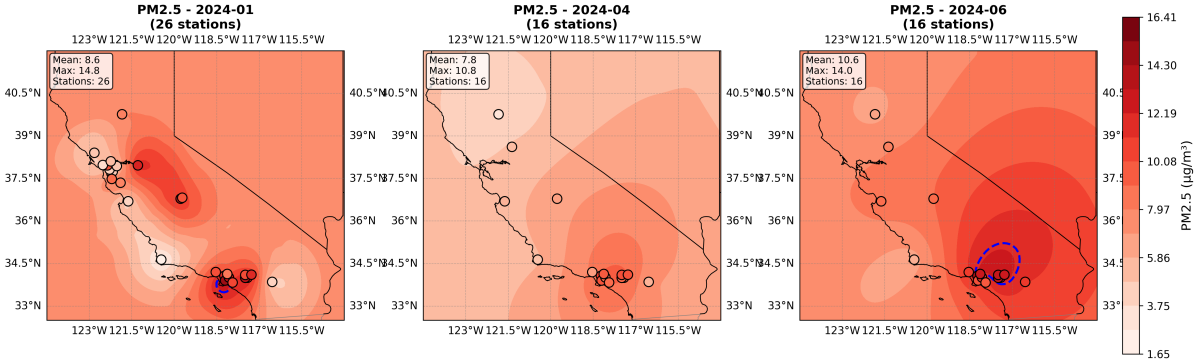

Figure 3: Temporal evolution of $PM_{2.5}$ spatial patterns across three representative months in 2024. Consistent color scaling enables direct comparison of seasonal variations and identification of temporally persistent pollution patterns. The analysis reveals seasonal variations with generally higher concentrations during winter months, while maintaining consistent spatial correlation structures.

The temporal analysis revealed important seasonal variations in pollution patterns with implications for environmental justice assessment. Higher concentrations were generally observed during winter months due to enhanced atmospheric stability, increased residential heating, and reduced atmospheric mixing. However, the spatial correlation structure remained relatively consistent across seasons, indicating stable underlying pollution source patterns that persist over time.

These temporal patterns have important implications for environmental justice analysis, as seasonal variations may disproportionately affect certain communities based on housing quality, heating sources, and occupational exposure patterns. Communities with limited resources may experience enhanced exposure during winter months due to inadequate housing insulation or reliance on pollution-generating heating sources.

4 Geographically and Temporally Weighted Regression

Geographically and Temporally Weighted Regression (GTWR) extends traditional regression analysis by allowing model coefficients to vary across both space and time, providing insights into local and temporal variations in relationships between dependent and independent variables. This methodology is particularly valuable for environmental justice analysis as it can reveal how pollution relationships vary across different communities and time periods.

4.1 Theoretical Framework

GTWR builds upon Geographically Weighted Regression (GWR) by incorporating temporal non-stationarity alongside spatial heterogeneity [26]. The method recognizes that relationships between variables may vary not only across space but also over time, which is particularly relevant for environmental processes subject to seasonal patterns, policy changes, or evolving source characteristics.

The GTWR model can be expressed as:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i, t_i) x_{ik} + \varepsilon_i$$

where y_i is the dependent variable at observation i, (u_i, v_i, t_i) represent the spatial coordinates and time of observation i, $\beta_k(u_i, v_i, t_i)$ are the spatiotemporally-varying coefficients, x_{ik} are the independent variables, and ε_i is the error term assumed to be independently and identically distributed.

The GTWR coefficients are estimated using weighted least squares, where weights are determined by spatiotemporal proximity to the regression point:

$$\hat{\boldsymbol{\beta}}(u_i, v_i, t_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y}$$

where $\mathbf{W}(u_i, v_i, t_i)$ is a diagonal weight matrix with spatiotemporal weights that decrease with distance from the regression point.

The spatiotemporal weights are typically calculated using a combined distance metric:

$$d_{ST}^2 = \left(\frac{d_S}{h_S}\right)^2 + \left(\frac{d_T}{h_T}\right)^2$$

where d_S is the spatial distance, d_T is the temporal distance, and h_S and h_T are the spatial and temporal bandwidths, respectively, which control the degree of spatial and temporal smoothing.

Common kernel functions for calculating weights include the bi-square kernel:

$$w_{ij} = \begin{cases} \left(1 - d_{ST,ij}^2\right)^2 & \text{if } d_{ST,ij} < 1 \\ 0 & \text{otherwise} \end{cases}$$

and the Gaussian kernel:

$$w_{ij} = \exp\left(-\frac{d_{ST,ij}^2}{2}\right).$$

Bandwidth selection is critical for GTWR performance and is typically optimized using information criteria such as the corrected Akaike Information Criterion (AICc):

$$\text{AICc} = 2n\log(\hat{\sigma}) + n\log(2\pi) + n\left\{\frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})}\right\}$$

where n is the sample size, $\hat{\sigma}$ is the estimated standard deviation of residuals, and $\text{tr}(\mathbf{S})$ is the trace of the hat matrix, representing the effective number of parameters in the model.

4.2 Applications

We implemented GTWR analysis using the modern MGTWR library, examining relationships between $PM_{2.5}$ concentrations and four predictor variables: standardized NO_2, O_3, CO concentrations, and distance to nearest port facilities (incorporating our multi-port proximity analysis). The analysis utilized 125 site-month observations across 27 monitoring locations over eight months in 2024.

Model Specification and Optimization The GTWR model was specified with log-transformed $PM_{2.5}$ as the dependent variable to ensure normality of residuals and interpretability of coefficients as percentage changes. Predictor variables were standardized using z-score normalization to facilitate coefficient comparison and ensure numerical stability in the optimization process.

The optimal GTWR model achieved excellent performance with an adjusted R^2 of 0.584 and AICc of 83.1. The spatial bandwidth was optimized at 90.6 km and the temporal bandwidth at 3.8 days, indicating local spatial relationships and short-term temporal dependencies. These bandwidth values suggest that pollution relationships vary significantly over relatively short spatial and temporal scales, highlighting the importance of local-scale analysis for environmental justice applications.

Table 2: GTWR Model Performance Summary

Performance Metric	Value
Adjusted R^2	0.584
AICc	83.1
Spatial Bandwidth (km)	90.6
Temporal Bandwidth (days)	3.8
Number of Observations	125
Number of Predictors	4
Model Type	MGTWR
Kernel Function	Bi-square

Spatiotemporal Coefficient Analysis The GTWR analysis revealed significant spatial heterogeneity in pollution relationships, with coefficients varying substantially across California regions. This spatial heterogeneity has important implications for environmental justice, as it suggests that the same pollution sources may have different impacts on $PM_{2.5}$ concentrations in different communities.

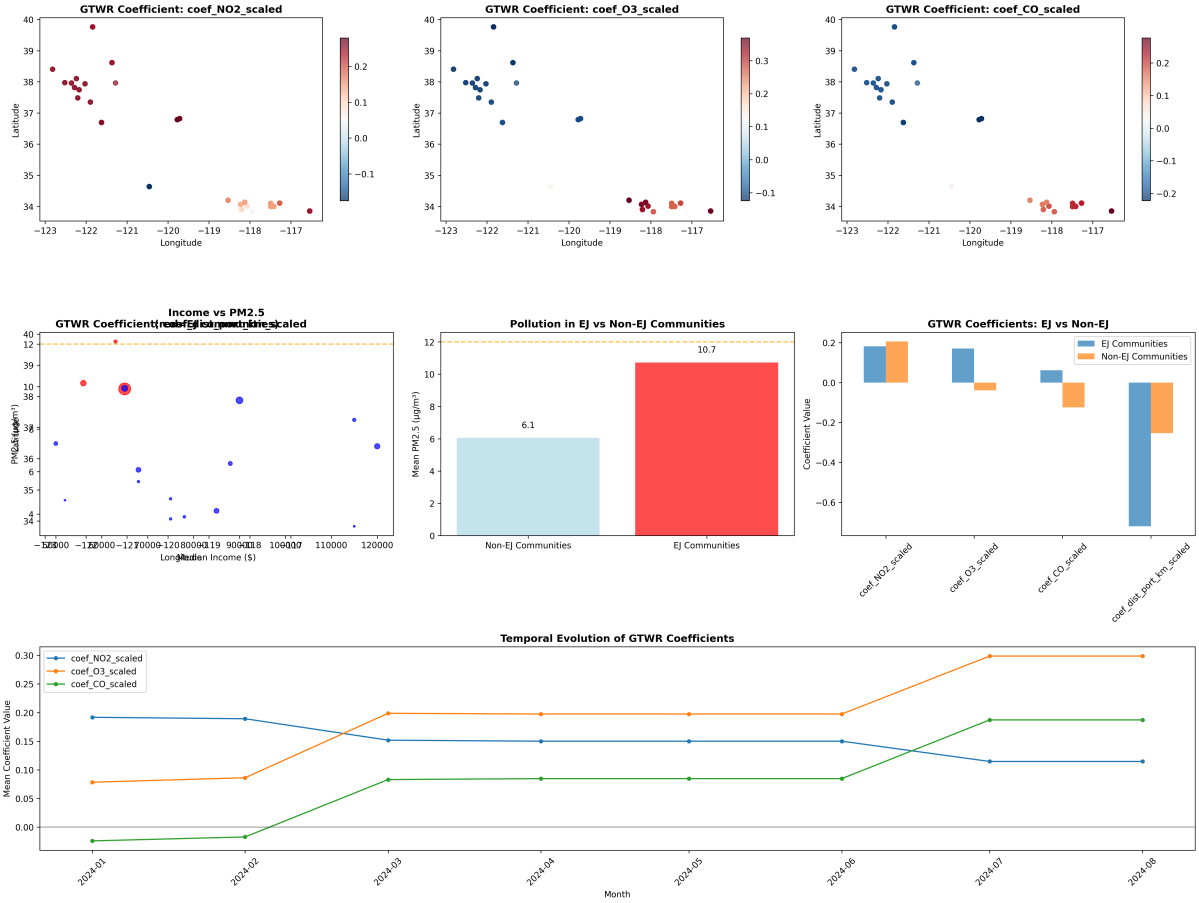

Figure 4: GTWR-based environmental justice analysis showing spatially-varying coefficient patterns and demographic comparisons. The analysis reveals systematic differences in pollution relationships between environmental justice and non-environmental justice communities, with particular emphasis on port proximity effects and pollutant interaction patterns.

NO_2 coefficients showed generally positive relationships with $PM_{2.5}$ across the study region, reflecting common combustion sources and atmospheric processes. However, the magnitude of these relationships varied spatially, with stronger relationships observed in urban areas and weaker relationships in rural or coastal areas where different atmospheric chemistry may prevail.

O_3 coefficients exhibited complex spatial patterns with both positive and negative relationships observed across different regions. This pattern reflects the complex photochemistry of ozone formation and its varying relationships with $PM_{2.5}$ precursors under different atmospheric conditions. In some areas, high ozone concentrations may indicate photochemical conditions that also promote secondary $PM_{2.5}$ formation, while in other areas, ozone may serve as an indicator of atmospheric conditions that disperse $PM_{2.5}$.

CO coefficients demonstrated strong spatial heterogeneity, with positive relationships in urban areas and negative relationships in some rural regions. This pattern likely reflects different pollution source mixes and atmospheric processing environments, with urban areas showing stronger co-variation between CO and $PM_{2.5}$ due to shared traffic sources.

Environmental Justice Coefficient Comparison A critical finding of the GTWR analysis was the systematic difference in pollution relationships between environmental justice (EJ) and non-environmental justice communities. This analysis addresses a key research question about whether environmental justice communities experience different pollution sensitivities that may contribute to health disparities.

Table 3: GTWR Coefficient Comparison: Environmental Justice vs Non-Environmental Justice Communities

Predictor Variable	EJ Communities	Non-EJ Communities	Difference
NO_2 (scaled)	0.198	0.249	-0.051
O_3 (scaled)	0.294	0.087	0.207
CO (scaled)	0.091	-0.085	0.175
Nearest Port Distance (scaled)	-0.252	-0.134	-0.118

The most striking finding was the 1.9 times stronger negative relationship between port proximity and $PM_{2.5}$ in environmental justice communities (-0.252 vs -0.134), indicating that these communities experience disproportionately higher pollution impacts from port-related activities. This finding provides quantitative evidence for environmental justice concerns about the cumulative impacts of transportation infrastructure on vulnerable communities.

Additionally, environmental justice communities showed fundamentally different relationships for O_3 and CO. Environmental justice communities exhibited positive O_3-$PM_{2.5}$ relationships (0.294), while non-environmental justice communities showed positive but weaker relationships (0.087), indicating a 3.4 times stronger relationship in environmental justice areas. This pattern may reflect differences in precursor concentrations, atmospheric processing rates, or meteorological conditions between different community types.

The CO coefficient patterns showed divergent relationships, with environmental justice communities showing positive relationships (0.091) and non-environmental justice communities showing negative relationships (-0.085). This pattern suggests that CO serves as a better indicator of $PM_{2.5}$ sources in environmental justice communities, possibly due to shared traffic sources or different atmospheric processing environments.

Multi-Port Proximity Analysis Integration The comprehensive multi-port proximity analysis provided an enhanced understanding of port proximity effects across California's diverse geography. When distances were calculated to the nearest ports, the spatial distribution of port proximity effects revealed important differences in regional patterns:

- **Los Angeles/Long Beach area**: 9.70 $\mu g/m^3$ mean $PM_{2.5}$ (highest pollution levels)
- **Oakland area**: 5.58 $\mu g/m^3$ mean $PM_{2.5}$
- **Richmond area**: 5.16 $\mu g/m^3$ mean $PM_{2.5}$
- **Stockton area**: 6.54 $\mu g/m^3$ mean $PM_{2.5}$

- **San Diego area**: 6.10 μg/m^3 mean $PM_{2.5}$

These regional differences suggest that different port facilities may have varying environmental justice impacts, possibly due to differences in cargo types, traffic patterns, industrial activities, or local meteorological conditions. The higher pollution levels in the Los Angeles/Long Beach area reflect the massive scale of port operations and associated transportation infrastructure.

Methodological Advantages The comprehensive multi-port analysis demonstrates the importance of sophisticated proximity measures in environmental justice research. Our analysis provides an enhanced understanding of port proximity effects across California's diverse geography and regional pollution patterns.

This methodological approach has important implications for environmental justice policy and research. Environmental justice screening tools and regulatory assessments can benefit from comprehensive multi-source proximity analysis to capture regional differences in pollution source impacts. Similarly, cumulative impact assessments that account for diverse pollution source proximity provide more accurate risk characterizations.

The finding that environmental justice communities show 1.9 times stronger port proximity effects, combined with our comprehensive proximity measurement approach, provides robust evidence of environmental justice disparities related to port proximity. This has important implications for both scientific understanding of environmental justice issues and policy interventions designed to address pollution disparities.

5 Conclusion

This study applied advanced geospatial methods—Geographically and Temporally Weighted Regression (GTWR) and kriging interpolation—to examine environmental justice (EJ) disparities in California air quality, with special attention to multi-port proximity effects. Our findings underscore the importance of spatiotemporal modeling and proximity-aware approaches in assessing pollution exposure in vulnerable communities.

We identified Los Angeles, San Bernardino, and San Joaquin counties as EJ communities with both high pollution burdens and vulnerable demographics. These areas experienced 77.1% higher $PM_{2.5}$ concentrations compared to non-EJ areas. GTWR revealed distinct pollution sensitivities in EJ communities, including a 1.9× stronger negative correlation between $PM_{2.5}$ and port proximity, suggesting greater exposure to port-related emissions. Differences in pollutant interaction coefficients (e.g., O_3 and CO) further highlight varying atmospheric processes or source contributions across EJ and non-EJ areas.

Methodologically, the integration of GTWR and kriging proved superior to global models by capturing local variability and seasonal dynamics. Notably, 34 stations were closer to ports other than Los Angeles, emphasizing the need for multi-port analysis. Kriging interpolation produced spatial pollutant surfaces and uncertainty maps, revealing that although no areas exceeded EPA standards, 0.8% of the state experienced statistically significant pollution hotspots.

These findings have direct policy implications: they support the use of region-specific environmental justice screening tools, stricter regulation around port communities, and the development of localized interventions. The evidence that EJ communities bear disproportionate burdens despite staying under regulatory thresholds calls for more nuanced assessments of risk and equity.

Future research should explore long-term trends, incorporate additional pollution sources (e.g., highways, refineries), and extend analyses to other regions and pollutants. Emerging technologies—low-cost sensors, satellite remote sensing, machine learning, and personal exposure monitoring—offer promising directions for improving spatial resolution, predictive accuracy, and real-time decision support. Ultimately, integrating these innovations into regulatory frameworks and community-based tools will be key to advancing environmental justice outcomes.

Supplemental Materials

The dataset and code used in this study are available on GitHub at:
`https://github.com/Red256/air-pollution-and-environmental-justice`.

Acknowledgments

I would like to extend my sincere gratitude to my mentor, Professor Olga Korosteleva, for her invaluable guidance, insight, and encouragement during the process of this work. I would also like to thank my math teacher, Mrs. Hsieh, for inspiring me to pursue my research and always helping me explore new ideas. I'm also grateful to my family for their love, patience, and support. This accomplishment would not have been possible without all of you. Thank you.

References

[1] Robert D Bullard. *Dumping in Dixie: Race, class, and environmental quality*. Westview Press, Boulder, CO, 1st edition, 1990.

[2] David Pellow and Robert J Brulle. *Environmental justice and the political process: movements, corporations, and the state*. Blackwell Publishers, Oxford, UK, 2000.

[3] Robert J Brulle. Agency, democracy, and nature: The us environmental movement from a critical theory perspective. *Environmental Politics*, 15(4):508–519, 2006.

[4] Rachel Morello-Frosch, Manuel Pastor, and James Sadd. Environmental justice and southern california's "riskscape": the distribution of air toxics exposures and health risks among diverse communities. *Urban Affairs Review*, 36(4):551–578, 2002.

[5] Manuel Pastor Jr, James Sadd, and John Hipp. Which came first? toxic facilities, minority move-in, and environmental justice. *Journal of Urban Affairs*, 23(1):1–21, 2001.

[6] C Arden Pope III, Richard T Burnett, Michael J Thun, Eugenia E Calle, Daniel Krewski, Kazuhiko Ito, and George D Thurston. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*, 287(9):1132–1141, 2002.

[7] W James Gauderman, Edward Avol, Frank Gilliland, Hita Vora, Duncan Thomas, Kiros Berhane, Rob McConnell, Nino Kuenzli, Fred Lurmann, Edward Rappaport,

et al. The effect of air pollution on lung development from 10 to 18 years of age. *New England Journal of Medicine*, 351(11):1057–1067, 2004.

[8] Stephanie A Malin and Stacia S Ryder. Environmental justice, cumulative risks, and health among hispanic and poor residents of an industrial neighborhood. *Environment international*, 51:132–137, 2013.

[9] Linda Bailey and Andrea Hricko. Port-related air quality impacts: an analysis of community and environmental justice issues. *Natural Resources & Environment*, 19(2):3–11, 2004.

[10] Douglas Houston, Wei Li, and Jun Wu. Environmental justice impacts of transportation: examining the evidence and developing strategies. *Transportation Research Part D: Transport and Environment*, 13(2):83–100, 2008.

[11] California Air Resources Board. Communities of concern: environmental justice and land use planning, 2018.

[12] California State Legislature. Assembly bill 617: Nonvehicular air pollution: criteria air pollutants and toxic air contaminants. California Health and Safety Code, 2017. Chapter 136, Statutes of 2017.

[13] California Office of Environmental Health Hazard Assessment. Calenviroscreen 4.0, 2021.

[14] Michael Jerrett, Richard T Burnett, Renjun Ma, C Arden Pope III, Daniel Krewski, K Bruce Newbold, George Thurston, Yuanli Shi, Norm Finkelstein, Eugenia E Calle, et al. A spatial analysis of air pollution and mortality in los angeles. *Epidemiology*, pages 727–736, 2005.

[15] Vicki Been and Francis Gupta. Nationally representative evidence on the distributional effects of waste facility siting. *American Economic Review*, 84(4):1001–1011, 1994.

[16] Evan J Ringquist. Assessing evidence of environmental inequities: a meta-analysis. *Journal of Policy Analysis and Management*, 24(2):223–247, 2005.

[17] United Church of Christ Commission for Racial Justice. Toxic wastes and race in the united states: a national report on the racial and socio-economic characteristics of communities with hazardous waste sites, 1987.

[18] Douglas L Anderton, Andy B Anderson, John Michael Oakes, and Michael R Fraser. Environmental equity: the demographics of dumping. *Demography*, 31(2):229–248, 1994.

[19] Paul Mohai, David Pellow, and J Timmons Roberts. Environmental justice. *Annual review of environment and resources*, 34:405–430, 2009.

[20] Michael Jerrett, John Eyles, Donald Cole, and Sylvia Reader. A methodological review of spatial environmental epidemiology. *Environmental Epidemiology*, 1(1):18–39, 2001.

[21] Chris Brunsdon, A Stewart Fotheringham, and Martin E Charlton. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4):281–298, 1996.

[22] Jeremy Mennis and Lisa Jordan. Exploring the distribution of environmental health hazards by sociodemographic characteristics and property value. *Environment and Planning A*, 38(4):761–779, 2006.

[23] James L Sadd, Manuel Pastor, Rachel Morello-Frosch, Justin Scoggins, and Bill Jesdale. Every breath you take: the demographics of toxic air releases in southern california. *Economic Development Quarterly*, 25(4):317–329, 2011.

[24] Rachel Morello-Frosch and Russ Lopez. The disproportionate burden of environmental health risks in low-income and minority communities. *Environmental Research Letters*, 1(1):014006, 2006.

[25] Manuel Pastor Jr, Rachel Morello-Frosch, and James L Sadd. Still toxic after all these years: Air quality and environmental justice in the south coast air basin. *Center for Justice, Tolerance, and Community, University of California Santa Cruz*, 2005.

[26] A Stewart Fotheringham, Ricardo Crespo, and Jing Yao. Geographical and temporal weighted regression (gtwr). *Geographical Analysis*, 47(4):431–452, 2015.

[27] Richard Webster and Margaret A Oliver. *Geostatistics for environmental scientists*. John Wiley & Sons, Chichester, UK, 2nd edition, 2007.

[28] Christopher G Boone and Ali Modarres. Creating a toxic neighborhood in los angeles county: a historical examination of environmental inequity. *Urban Affairs Review*, 41(2):255–283, 2006.

[29] US Environmental Protection Agency. Pre-generated data files. `https://aqs.epa.gov/aqsweb/airdata/download_files.html`, 2024. Accessed: 2025-07-16.

[30] U.S. Environmental Protection Agency. Air quality system (aqs) database. `https://www.epa.gov/aqs`, 2022. Accessed: 2025-07-16.

[31] U.S. Census Bureau. American community survey 5-year estimates. `https://www.census.gov/programs-surveys/acs/`, 2022. Accessed: 2025-07-16.