# How have U.S. $CO_2$ spatial patterns shifted between 2000 and 2024, and how do urbanization and population density influence these changes?*

Cristina Su Lam

December 30, 2024

## Contents



*Code available at: https://github.com/cristinaasu/CO2_US

# 1 Introduction

The spatial distribution of carbon dioxide ($CO_2$) concentrations has significant implications for environmental policy, climate change mitigation, and public health. Understanding these patterns is critical for addressing regional disparities in emissions and developing targeted interventions. This report investigates spatial patterns in $CO_2$ concentrations across U.S. states, leveraging data from January 2024 and comparing it to historical data from January 2000. The analysis focuses on identifying spatial clusters of emissions, assessing changes over time, and exploring the influence of urbanization and population density on spatial autocorrelation.

To address these questions, the study employs a combination of exploratory spatial data analysis (ESDA), and spatial modeling techniques. The analysis compares three modeling approaches: Ordinary Least Squares (OLS), Spatial Autoregressive (SAR), and Conditional Autoregressive (CAR) models. The results reveal a significant reduction in overall emissions between 2000 and 2024, with persistent hotspots identified in urban industrial regions such as California and Nevada, in addition to notable progress in mitigating emissions in the Midwest and Northeast clusters.

# 2 Methods

## 2.1 Data Collection and Cleaning

This analysis utilized daily mean $CO_2$ concentration data for January 2024 and January 2000, sourced from United States Environmental Protection Agency (2024). Additional datasets on population density (World Population Review, 2024) and urbanization (United States Census Bureau, 2020) were integrated to examine their relationship with $CO_2$ levels.

Preprocessing was conducted using R Core Team (2023), including cleaning, filtering, and merging steps to ensure consistency and relevance for spatial modeling. Key packages such as 'spdep' (Bivand, 2022), 'geoR' (Ribeiro Jr et al., 2024), and 'spatialreg' (Bivand et al., 2021) were integral for constructing spatial weights matrices, calculating Moran's I, and implementing spatial regression models like SAR and CAR, which address spatial dependencies in the data. Visualization was enhanced with 'ggplot2' Wickham (2016) and 'ggspatial' (Dunnington, 2023), enabling the creation of high-quality, spatially detailed maps that incorporate contextual elements such as basemaps and scalebars to better interpret spatial patterns.

To align demographic data with $CO_2$ records, the datasets were merged using a left join, matching state names. This facilitated a targeted analysis of spatial dependencies and demographic factors in 2024, focusing on current emission patterns, with January 2000 data serving as a historical benchmark for comparison, offering insights into changes in spatial clustering and trends over two decades.

The states of Alaska, Hawaii, and Puerto Rico were excluded to narrow the analysis to the contiguous U.S., ensuring consistency in geographic and demographic characteristics. The District of Columbia was omitted due to insufficient data. Filters were applied to retain records with Sample Duration equal to "8-HR RUN AVG END HOUR," Parameter Name equal to "Carbon monoxide," and Pollutant Standard equal to "CO 8-hour 1971". These criteria ensured standardized measurements, with the 8-hour run average offering a stable metric that reflects sustained emissions levels and facilitates reliable comparisons across states.

The variable Mean $CO_2$ was created by grouping observations by state name and local date, averaging $CO_2$ values across counties to align with the study's state-level focus. To classify states, a binary variable, Urban, was generated: states with the majority of their population in urban areas were labeled as urban (Urban = 1), while others were classified as rural (Urban = 0). This classification enabled comparisons between urban and rural states, offering insights into urbanization's influence on $CO_2$ concentrations.

A sample of the aggregated dataset for January 2024 is shown below in Table 1. The January 2000 dataset is similar but lacks the Urban and Population Density variables.

Table 1. Preview of Aggregated CO2 Data for January 2024

| State Name | Date Local | Mean CO2 | Geometry | Urban | Population Density |
|---|---|---|---|---|---|
| Arizona | 2024-01-01 | 0.5605264 | MULTIPOLYGON (((-114.8163 ... | 1 | 25.48203 |
| Arizona | 2024-01-02 | 0.4329167 | MULTIPOLYGON (((-114.8163 ... | 1 | 25.48203 |
| Arizona | 2024-01-03 | 0.4102205 | MULTIPOLYGON (((-114.8163 ... | 1 | 25.48203 |

## 2.2 Exploratory Spatial Analysis

Global Moran's I and Local Moran's I were employed for spatial data exploration due to their effectiveness in detecting and quantifying spatial autocorrelation. Global Moran's I provides a single summary statistic that evaluates whether spatial clustering or dispersion exists across the entire dataset. In contrast, Local Moran's I identifies specific clusters or outliers, such as high-high hotspots—regions with elevated emissions surrounded by similarly high values—or low-low cold spots, where consistently low emissions dominate both the focal area and its neighbors. These metrics enable a detailed understanding of regional variations in $CO_2$ concentrations.

To establish spatial relationships, Queen and Rook contiguity-based spatial weights matrices were constructed. As shown in Figure 1, Queen contiguity captures broader spatial interactions by including neighbors that share borders or vertices. In contrast, Rook contiguity, depicted in Figure 2, restricts connections to neighbors with shared borders only. These structures were used to compute Moran's I statistics and assess spatial autocorrelation in $CO_2$ concentrations.

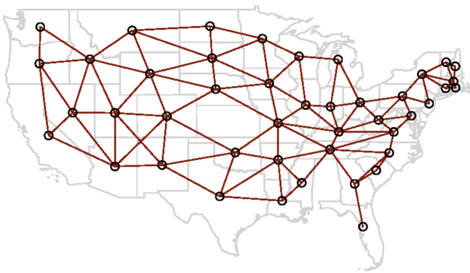

Figure 1. Queen contiguity network

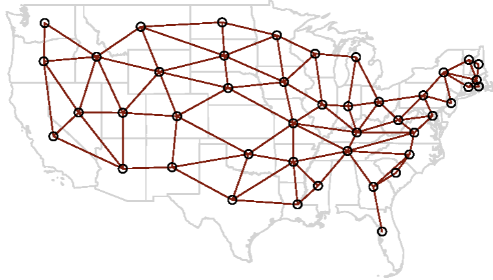

Figure 2. Rook contiguity network

Key assumptions for Moran's I include independence of observations and proper specification of the spatial weights matrix to ensure accurate representation of spatial relationships. Comparisons of the Moran's I statistic and associated p-values helped determine the strength and significance of clustering under each weights scheme. These analyses guided the selection of the most appropriate spatial weights matrix for subsequent modeling, ensuring a rigorous framework for understanding spatial patterns.

## 2.3 Spatial Modeling

The OLS model served as a baseline to establish linear relationships between $CO_2$ concentrations, population density, and urbanization. Although straightforward, this model assumes independent and identically distributed residuals, linearity, homoscedasticity (constant variance of residuals), and normally distributed errors. However, the presence of spatial dependencies in the data violates these assumptions, as confirmed by residual analysis showing significant spatial autocorrelation through Moran's I. This necessitated the use of spatial models to address these dependencies.

To address spatial autocorrelation, the SAR Lag-Error model was employed. This model combines two components: the spatial lag term, which captures interactions between neighboring states, and the spatial error term, which accounts for unobserved spatially correlated factors. Proper specification of the spatial weights matrix is crucial for accurately modeling these dependencies. This approach effectively incorporates regional influences by accounting for spatial relationships through either neighboring values (lag) or correlated errors.

The CAR model further refined spatial analysis by introducing random effects based on neighborhood structures. Unlike the SAR model, which captures broader regional effects, the this model assumes localized spatial interactions where directly adjacent regions influence each other. By using spatial weights matrices, such as queen or rook contiguity, the CAR model excels at capturing detailed spatial dependencies and localized variations. This makes it particularly suited for datasets with fine-grained spatial structures.

Model diagnostics, including AIC, $R^2$, and Moran's I for residual spatial independence, were used to evaluate each model. These diagnostics ensured that assumptions were rigorously tested and the most suitable framework was selected to explain spatial variations in $CO_2$ concentrations.

# 3 Results

## 3.1 Temporal Trends in CO2 Concentrations

Table 3 and Table 2 summarize the changes in $CO_2$ concentrations between January 2000 and January 2024, highlighting significant progress over two decades. In 2000, the mean $CO_2$ concentration was 0.8251 with a standard deviation of 0.4108, indicating considerable variability across states. By 2024, the mean declined to 0.2675 with a reduced standard deviation of 0.1312, reflecting substantial reductions in emissions and a more uniform distribution across regions.

Table 2. Summary Statistics for Jan 2024

| Statistic | Mean $CO_2$ | Population Density |
|---|---|---|
| Mean | 0.2675 | 83.3521 |
| SD | 0.1312 | 109.1149 |
| Min | 0.0000 | 2.3322 |
| Max | 0.8736 | 489.3665 |

Table 3. Summary Statistics for Jan 2000

| Statistic | Mean | SD | Min | Max |
|---|---|---|---|---|
| Mean $CO_2$ | 0.8251 | 0.4108 | 0 | 2.3382 |

The geographical maps in Figure 3, and Figure 4, further illustrate these shifts in spatial distribution. In 2000, $CO_2$ emissions were predominantly concentrated in western states such as Utah and Arizona, while northeastern regions displayed comparatively lower levels. By 2024, these hotspots had significantly diminished, with Oklahoma emerging as the state with the highest mean $CO_2$ emissions. This transition underscores the effectiveness of nationwide mitigation efforts, although regional disparities remain.

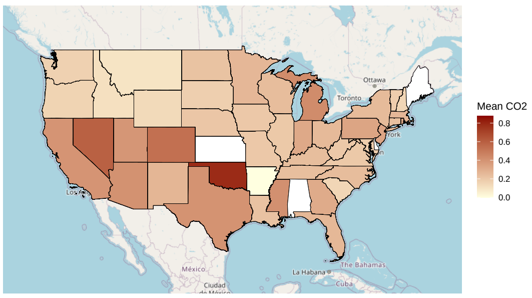


Figure 3. Mean $CO_2$ concentrations in Jan 2024 (white indicates missing data)

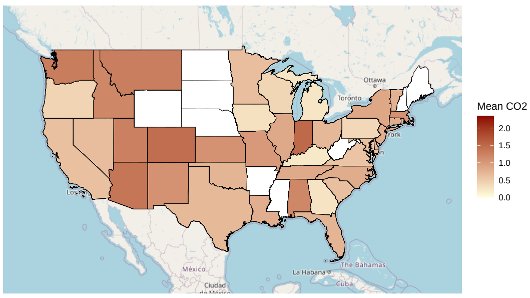


Figure 4. Mean $CO_2$ concentrations in Jan 2024 (white indicates missing data)

The scatter plot in Figure 5 reinforces these findings, showing most states experienced markedly lower $CO_2$ concentrations by 2024. Nevada and Arizona, in particular, demonstrated significant reductions, reflecting targeted efforts and the influence of urbanization on emissions. Despite the overall decline, the uneven pace of reductions highlights the need for continued region-specific policies to address persistent disparities and sustain progress in emissions mitigation.

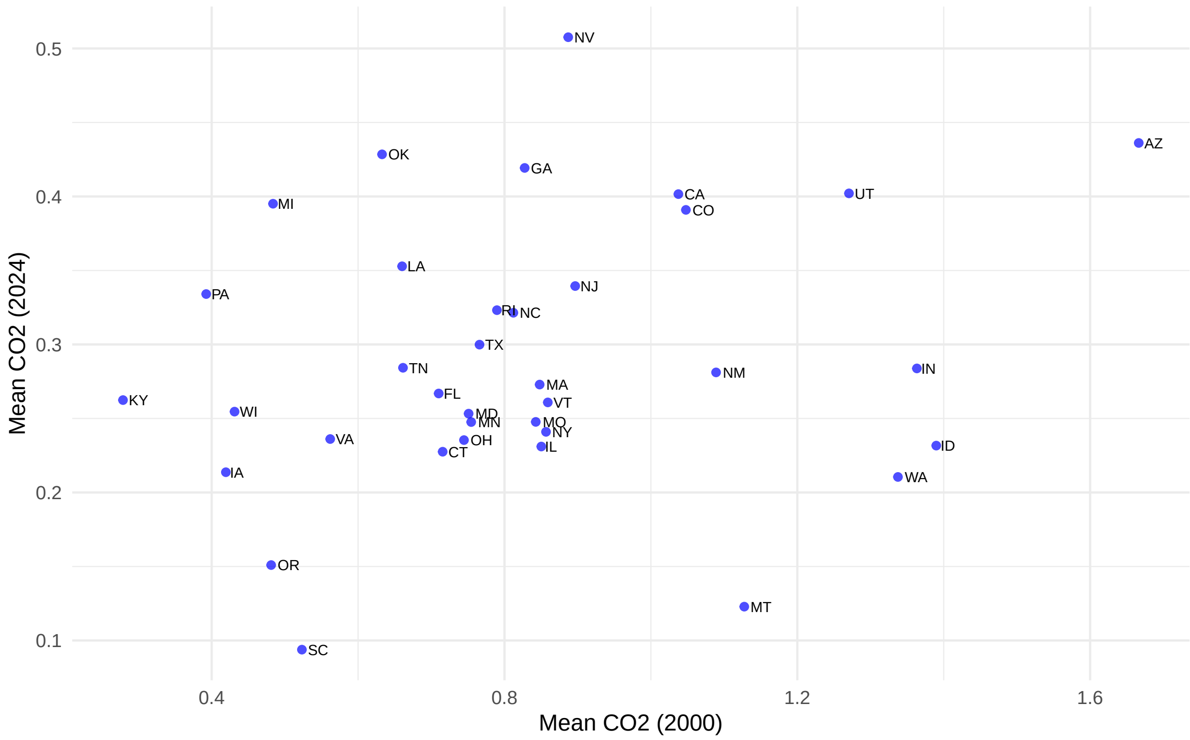


Figure 5. State-level comparison of mean $CO_2$ concentrations between 2000 and 2024

## 3.2 Spatial Patterns and Clustering

The Global Moran's I analysis confirmed significant spatial autocorrelation in $CO_2$ concentrations for both 2000 and 2024, with Queen contiguity yielding prominent clustering patterns as shown in Table 4. This matrix captured extensive spatial interactions, aligning with the hypothesis that $CO_2$ emissions are influenced by regional dynamics rather than just adjacent connections. In contrast, Rook contiguity, as seen in Table 5, produced slightly lower Moran's I values, reflecting its narrower definition of spatial relationships. Despite this, both methods highlighted significant spatial clustering, underscoring regional dependencies.

Table 4. Moran's I - Queen Weights

| Variable | Statistic | P-value |
|---|---|---|
| Mean $CO_2$ Jan 2000 | 0.2051 | $< 0.001$ |
| Mean $CO_2$ Jan 2024 | 0.1686 | $< 0.001$ |

Table 5. Moran's I - Rook Weights

| Variable | Statistic | P-value |
|---|---|---|
| Mean $CO_2$ Jan 2000 | 0.2043 | $< 0.001$ |
| Mean $CO_2$ Jan 2024 | 0.1676 | $< 0.001$ |

The Local Moran's I analysis, applied with Queen weights, provided detailed insights into localized spatial patterns over the two decades. Figure 6, in 2024, persistent hotspots (high-high clusters) were concentrated in industrial hubs like California and Nevada, reflecting urban and industrial activities. Conversely, cold spots (low-low clusters) were prominent in eastern regions. Comparing 2000 and 2024, as seen in Figure 7, several high-high clusters, such as those in Washington and Idaho, dissipated by 2024, suggesting significant progress in emission mitigation. However, the emergence of new high-low and low-high outliers underscored evolving regional disparities in $CO_2$ emissions.

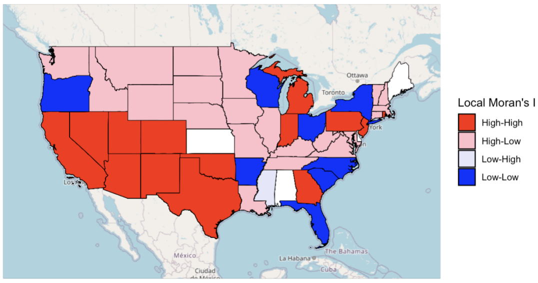


Figure 6. Local Moran's I clusters for Jan 2024

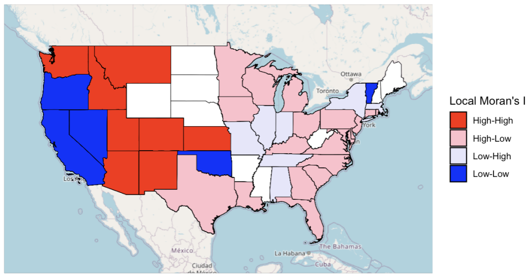


Figure 7. Local Moran's I clusters for Jan 2000

Together, these spatial analyses reveal the shifting dynamics of $CO_2$ emissions across the U.S., offering valuable insights into the effectiveness of mitigation efforts and identifying areas that require continued policy intervention and research. Persistent clustering and emerging patterns emphasize the importance of regional strategies for addressing disparities and promoting sustainable environmental outcomes.

## 3.3 Spatial Models Regression

The OLS model yielded an AIC of -1674.558 and an $R^2$ of 0.0097, reflecting its limited ability to explain variations in $CO_2$ concentrations and address spatial dependencies. As shown in Table 6, population density was positively associated with $CO_2$ ($p = 0.0004$), while urbanization showed no significant effect ($p = 0.674$). The spatial clustering of residuals, illustrated in Figure 8, highlights regions with high discrepancies, particularly in the Midwest and South, and confirms significant spatial autocorrelation (0.166, $p < 0.001$) through Moran's I. These results emphasize the inadequacy of the OLS model for capturing spatial patterns, underscoring the need for advanced spatial regression models to better account for regional dynamics influencing $CO_2$ concentrations.

Table 6. Coefficient estimates from the OLS model

| Variables | Estimate | P-value |
|---|---|---|
| Intercept | 0.2523 | $< 0.0001$ |
| Population Density | 0.0001156 | 0.0004 |
| Urban | 0.005965 | 0.6740 |

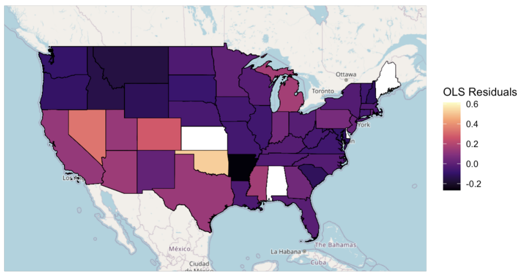


Figure 8. Residuals from the OLS model

To address spatial relationships, the SAR Lag-Error model, utilizing Queen's contiguity weights, showed significant improvements over the OLS model, achieving a lower AIC of -1946.356 and reduced spatial autocorrelation in residuals, as reflected by Moran's I (0.0159, $p < 0.001$). As summarized in Table 7, population density and urbanization were both statistically significant predictors of $CO_2$ concentrations, with population density showing a positive association and urbanization a negative one. These results, visualized in Figure 9, indicate that the SAR model better captures spatial dependencies than OLS, offering insights into regional dynamics.

Table 7. Coefficient estimates from the SAR model

| Variables | Estimate | P-value |
|---|---|---|
| Intercept | 0.3839 | 0.2065 |
| Population Density | 0.0002259 | <0.0001 |
| Urban | -0.04083 | 0.0053 |

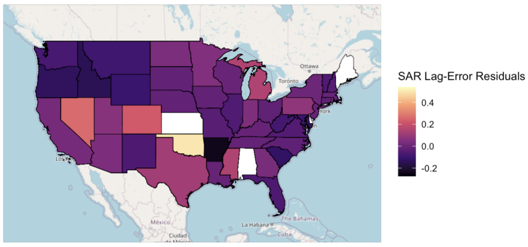


Figure 9. Residuals from the SAR model

The CAR model refined the spatial analysis, outperforming both OLS and SAR with the lowest AIC of -13290.796, highlighting its ability to account for complex spatial dependencies. Residual Moran's I (-0.032, $p = 0.8156 > 0.01$) indicated no significant spatial autocorrelation, as we fail to reject the null hypothesis, providing a robust framework for understanding $CO_2$ concentrations. As shown in Table 8 population density exhibited a stronger positive association with $CO_2$ concentrations (coefficient = 0.006, $p < 0.0001$), consistent with the expectation that denser regions experience higher emissions due to concentrated human activities. The negative influence of urbanization was also more pronounced (coefficient = -1.15, $p < 0.0001$), underscoring the critical role of urban planning and policies in reducing emissions. Figure 10 further illustrates the absence of residual clustering, demonstrating the model's capacity to capture spatial patterns while offering a more nuanced understanding of regional dynamics.

Table 8. Coefficient estimates from the CAR model

| Variables | Estimate | P-value |
|---|---|---|
| Intercept | -4.4615 | <0.0001 |
| Population Density | 0.0059387 | <0.0001 |
| Urban | -1.1487 | <0.0001 |

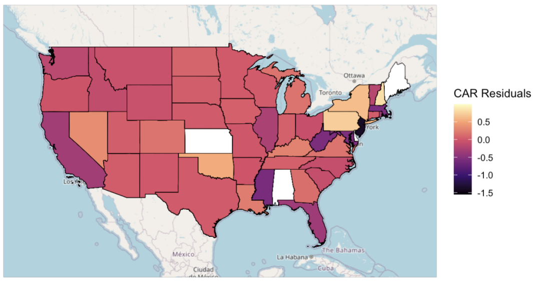


Figure 10. Residuals from the CAR model

In summary, the CAR model demonstrated the best fit among the three approaches, capturing spatial dependencies more effectively and eliminating residual clustering. Its findings reinforce the importance of targeted mitigation strategies in densely populated areas while emphasizing the potential benefits of urbanization in reducing emissions.

These results provide valuable insights for policymakers aiming to address spatially heterogeneous $CO_2$ emissions and prioritize interventions in high-density regions.

# 4 Limitations and Future Directions

This analysis has several limitations that warrant consideration. First, the study's reliance on January data for 2000 and 2024 may not capture seasonal variations or temporal shifts in $CO_2$ emissions, which could influence spatial patterns. Additionally, using population density and urbanization as proxies for demographic and socioeconomic factors may oversimplify the complex drivers of emissions, potentially excluding critical variables. Incorporating more detailed socioeconomic data, such as household income, education levels, and employment distribution, alongside industrial metrics like energy consumption, manufacturing activity, or transportation emissions, could provide a more comprehensive understanding of emission drivers. The spatial weights matrices, based on Queen and Rook contiguity, assume uniform relationships between states and may not fully capture nuanced regional interactions. Future research could explore alternative matrices, such as k-nearest neighbors (k-NN), inverse distance weights, or adaptive distance-based weights, to better represent complex spatial dependencies. Furthermore, the state-level focus may obscure local heterogeneities, limiting the granularity of insights. High-resolution analyses at the county or city level would further enhance the understanding of localized emission trends and mitigation strategies, enabling policymakers to design interventions tailored to specific regions.

# 5 Conclusion

This study reveals significant shifts in the spatial distribution of $CO_2$ concentrations across U.S. states between 2000 and 2024, with notable reductions in overall emissions and changes in regional clustering patterns. The CAR model emerged as the most robust framework, effectively capturing spatial dependencies and providing a nuanced explanation of $CO_2$ variations. These findings highlight the value of spatial econometric approaches in environmental research, emphasizing the need to account for regional dynamics in emissions analysis. Persistent hotspots in urban-industrial hubs, such as California and Nevada, underline the necessity of targeted policy interventions to mitigate emissions in these high-impact areas. Meanwhile, the dissipation of some clusters, particularly in the Midwest and Northeast, signals progress in emissions control, potentially driven by advancements in technology, regulatory measures, or shifts in industrial practices. This analysis provides a strong basis for future research and policymaking, advocating for strategies that prioritize emissions reductions while addressing spatial disparities to achieve sustainable environmental outcomes.

# References


Bivand, R. (2022). R packages for analyzing spatial data: A comparative case study with areal data. *Geographical Analysis 54*(3), 488–518.

Bivand, R., G. Millo, and G. Piras (2021). A review of software for spatial econometrics in r. *Mathematics 9*(11), 1276.

Dunnington, D. (2023). *ggspatial: Spatial Data Framework for ggplot2*. R package version 1.1.9.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ribeiro Jr, P. J., P. Diggle, O. Christensen, M. Schlather, R. Bivand, and B. Ripley (2024). *geoR: Analysis of Geostatistical Data*. R package version 1.9-4.

United States Census Bureau (2020). Decennial census 2020.

United States Environmental Protection Agency (2024). Air data: Download files.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

World Population Review (2024). State densities.