Changes in Fine Particulate Matter ($PM_{2.5}$) Levels in the United States: A Case Study of 1999 vs 2012

ELHARRAR MEHDI[a,*]

[a]*Department of Public Health Data Science, University of XYZ, 123 Research Street, CityName, 12345, StateName, Morocco*

Abstract

This study compares $PM_{2.5}$ concentrations in the United States between 1999 and 2012 using national monitoring data. For clarity in descriptive trend estimation, analyses are restricted to valid measurements (observations with non-missing, non-negative $PM_{2.5}$ values). We describe national and state-level distributions, quantify changes in central tendency and dispersion, and identify states with the largest improvements or deteriorations. Results show a clear downward shift in median $PM_{2.5}$ between 1999 and 2012, with heterogeneous state-level patterns. We discuss implications for exposure assessment and highlight sensitivity checks to evaluate the impact of data exclusions.

Keywords: $PM_{2.5}$, air pollution, data quality, United States, environmental epidemiology

1. Introduction

Airborne fine particulate matter with aerodynamic diameter ≤ 2.5 μm ($PM_{2.5}$) is a well-established risk factor for cardiovascular and respiratory disease [1]. Over the last two decades the United States has implemented a range of regulatory and technological interventions aimed at reducing emissions from industry, transportation, and power generation; monitoring networks have also expanded and evolved [2]. Quantifying how $PM_{2.5}$ distributions changed between 1999 and 2012 provides a useful overview of national progress and highlights spatial heterogeneity in exposure reductions.

This paper presents a descriptive comparison of $PM_{2.5}$ data from two benchmark years (1999 and 2012). To focus the analysis on reliable measurements, we restrict analyses to valid observations (non-missing, non-negative $PM_{2.5}$ values) and summarize changes at national and state levels. We present distributional comparisons (boxplots, densities), state-level trajectories, and maps, and we identify states with the largest percent changes. Where relevant, we flag limitations introduced by the data-cleaning choices and propose sensitivity checks [3].

1.1. Primary Objective

- Describe how $PM_{2.5}$ levels changed between 1999 and 2012 at national and state levels using valid monitoring observations.

1.2. Research Questions

1. Did median $PM_{2.5}$ decline nationally from 1999 to 2012?
2. Which states experienced the largest decreases or increases?
3. Are changes geographically structured (regional patterns)?
4. Do larger reductions occur where monitoring density changed?

*Corresponding author: elharrar.elmehdi@uni.edu.com

1.3. Secondary Analyses

- Seasonal patterns: Compare monthly or seasonal distributions, if sufficient data are available.
- Robustness checks: (a) Include previously excluded negative values as NA (b) Exclude states with very few monitors and repeat key summaries

2. Methods

2.1. Data cleaning and selection

Data were sourced from the U.S. Environmental Protection Agency (EPA) Air Quality System (AQS) for the years 1999 and 2012. Observations with missing or negative $PM_{2.5}$ values were excluded from the primary analysis; these exclusions are reported in Table 1 (number and percent removed).

All subsequent analyses were performed on the cleaned dataset. Given the expected non-normal distribution of air pollution data, we selected the median as a robust measure of central tendency. To quantify the uncertainty in the change of medians between 1999 and 2012, we employed a percentile bootstrap procedure with R = 10,000 resamples. The analysis was performed with a fixed random seed (set.seed(2025)) to ensure reproducibility. We calculated the 95% confidence interval (CI) for the difference in medians ($median_{2012}$ - $median_{1999}$) using the 2.5th and 97.5th percentiles of the bootstrap distribution.

3. Results

3.1. Data processing

The initial dataset contained 1,421,708 observations. The data cleaning process, summarized in Table 1, resulted in the exclusion of 7.9% of the data, primarily due to missing $PM_{2.5}$ values. The final analytical dataset comprised 1,308,884 valid observations.

metric	value	percentage
total_rows	1421708	100 %
removed_missing	86350	6.074 %
removed_negative	26474	1.862 %
kept_rows	1308884	92.064 %

Table 1: Data cleaning summary (counts & % of total)

3.2. Change in National Median $PM_{2.5}$

A comparison of national $PM_{2.5}$ distributions reveals a clear downward shift between 1999 and 2012. Figure 1 visually illustrates this change, showing a lower median and a tighter interquartile range in 2012 compared to 1999. Both distributions exhibit a strong right skew, with numerous high-concentration outliers.

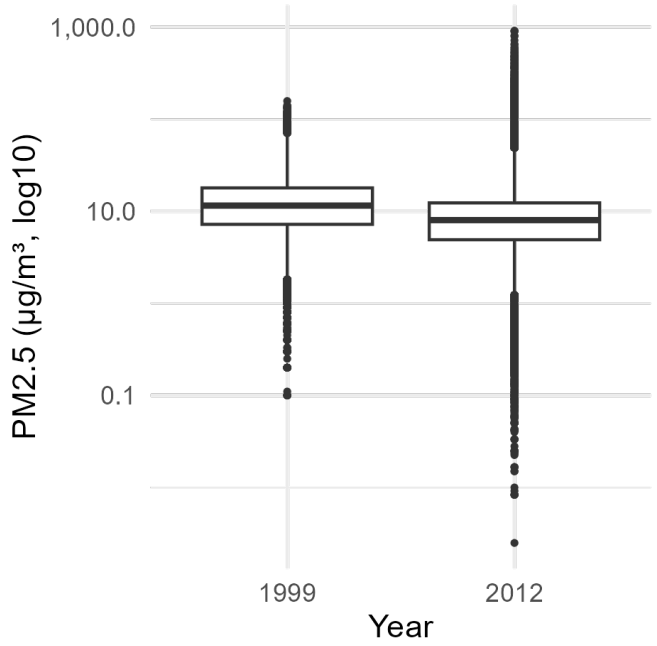

Figure 1: Distribution of $PM_{2.5}$: 1999 vs 2012

The precise change is quantified in Table 1. The national median $PM_{2.5}$ concentration decreased from 11.50 $\mu g\,m^{-3}$ in 1999 to 7.90 $\mu g\,m^{-3}$ in 2012. Our primary inferential analysis focuses on the difference between these medians. The bootstrap estimate for this change (2012 − 1999) was -3.60 $\mu g\,m^{-3}$, with a 95% percentile bootstrap confidence interval of $[-3.60, -3.50]$ $\mu g\,m^{-3}$. Because the confidence interval lies entirely below zero, this decline is statistically significant.

Statistic	1999	2012
Sample Size (n)	104,204	1,204,680
Median $PM_{2.5}$ ($\mu g/m^3$)	11.50	7.90
95% Bootstrap CI	[11.40, 11.54]	[7.90, 8.00]
Median Difference (2012 - 1999):	**-3.60**	*(95% CI: [-3.60, -3.50])*

Table 2: National $PM_{2.5}$ statistics for 1999 and 2012.

4. Discussion

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

5. Conclusions

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

References

[1] Yu-Fei Xing, Yue-Hua Xu, Min-Hua Shi, and Yi-Xin Lian. The impact of pm2.5 on the human respiratory system. *Journal of Thoracic Disease*, 8(1):E69, 2016. doi: 10.3978/j.issn.2072-1439.2016.01.19. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4740125/. PMID: 26904255.

[2] Chad Milando, Lei Huang, and Stuart Batterman. Trends in PM2.5 emissions, concentrations and apportionments in Detroit and Chicago. *Atmospheric environment (Oxford, England : 1994)*, 129:197–209, 2016. ISSN 1352-2310. doi: 10.1016/j.atmosenv.2016.01.012. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5603263/.

[3] Wilmar Hernandez, Alfredo Mendez, Angela Maria Diaz-Marquez, and Rasa Zalakeviciute. Robust analysis of pm2.5 concentration measurements in the ecuadorian park la carolina. *Sensors (Basel)*, 19(21):4648, 2019. doi: 10.3390/s19214648. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6864519/.