Journal of Research High School

Vol. 2025 (12) 964 – 970
ISSN 2688-3651 [online]

Evaluating Source-Receptor Relationships of PM2.5 in the Upper New Jersey Area Using Potential Source Contribution Function

Aidan Jung[1] *

[1]Bergen Catholic High School, Oradell, NJ, USA
*Corresponding Author: aidanj0103@gmail.com

Advisor: Myoungwoo Kim, information@bluebellenvirobotics.com

Received January 13, 2025; Revised July 23, 2025; Accepted September 9, 2025

Abstract

$PM_{2.5}$, defined as fine particulate matter with a diameter of less than $_{2.5}$ micrometers, has emerged as a serious environmental concern in the United States. Exposure to unhealthy amounts of $PM_{2.5}$ may result in asthma, heart disease, lung disease, or even premature mortality. This research focused on methods of identifying local, mid-range, and long-range transport of $PM_{2.5}$, essential for the development of statewide air quality implementation policies. The tools employed on this research are HYSPLIT4 back-trajectory model, cluster analysis, Potential Source Contribution Function (PSCF), and Weighted-PSCF (WPSCF). According to data generated from back-trajectory endpoint-based statistics, the North and Northwest accounted for approximately 40% of the $PM_{2.5}$ concentration in upper New Jersey, with the majority of the transport coming from long-distance sources like smoke from Canadian wildfires. The northeast, east, and southeast regions contribute 11% of total $PM_{2.5}$ emissions. From the west, urban and industrial emissions contribute 29% of total $PM_{2.5}$. PSCF displayed the mid-range source receptor relationship of Northern New Jersey and its surrounding states, while the four together, cluster analysis, PSCF, and WPSCF provide data revealing that 20% $PM_{2.5}$ contribution came from the south and southwest, indicating the impact of local sources like Newark Liberty International Airport, adjacent ship channels, industrial areas, and the New York Metropolitan area. This research can be used by lawmakers and public health experts to develop environmental policies that address local, mid-range, and distant $PM_{2.5}$ sources benefiting the lives of millions of residents in upper New Jersey and beyond.

Keywords: Fine particulate matter, Long-range transport, HYSPLIT4, Back-trajectory analysis, Cluster analysis, Potential Source Contribution Function analysis

1. Introduction

$PM_{2.5}$ are fine particle matters less than $_{2.5}$ micrometers in diameter. Due to their tiny size, they can be easily inhaled, going from the nose to the throat and all the way down to the lungs creating a multitude of health problems like heart and lung disease as stated by the EPA (Particulate Matter (PM) Basics). Just as easily as they can move around, they are, likewise, easily able to be produced. Their common sources include factories, mobile emissions, construction sites, natural phenomena such as wildfires, and agricultural activities (Particulate Matter (PM) Basics).

This study aimed to address the critical question of how adjacent states and localized industrial and urban emissions impact the concentration of $PM_{2.5}$ in Upper New Jersey, specifically Fort Lee, by utilizing multiple source apportionment methods. In this study, the HYSPLIT4 air back-trajectory model, Potential Source Contribution Function (PSCF) analysis, and spatial kriging analysis method were used to apportion the source origins contributing to the upper NJ areas. This information helps policymakers improve public health policies, benefiting both the state and its neighbors.

Vol. 2025 (12) 964 – 970
ISSN 2688-3651 [online]

2. Methodology and Previous Works

To identify the source contributors of $PM_{2.5}$, weather data and $PM_{2.5}$ concentration data were collected from National Climate Data Center (NCDC) at U.S. NOAA and the Daily Air Quality Monitoring Data Center at the United States Environmental Protection Agency, respectively. All the trajectory endpoint data used in these methods were generated by NOAA's HYSPLIT4 Back Trajectory Model. This research employed five types of air quality-related models and statistical techniques: Kolmogorov-Zurbenko (KZ) analysis, HYSPLIT4 back-trajectory model, cluster analysis, Potential Source Contribution Function (PSCF), and Weighted PSCF (WPSCF).

2.1 The Kolmogorov-Zurbenko (KZ) filter analysis

The KZ filter analysis method was conducted to evaluate long-term trends of $PM_{2.5}$. KZ filter is a statistical method used for removing noise from time series data, particularly effective in detecting and estimating trends and long-term changes in environmental data. KZ filter's equation can be seen below.

$$y[n] = \left(\frac{1}{M}\right) \cdot sum\big(x[n - 1 + 1 + x[x - i + 2] + \cdots + x[x - i + M]]\big)$$

Equation 1. KZ filter equation where x[n], y[n], M, and N represents the input time series data, output smoothed data, order of the filter (i.e., the number of data points in each moving average window), and length of the time series data respectively.

This equation tracks long-term pollutant trends to guide state emissions reduction plans. The dataset spans from 2000 to 2023, covering more than 20 years of daily $PM_{2.5}$ observations. Due to KZ filter analysis' ability to reduce random short-term events that could otherwise affect results, it earns its credibility as a long-term trend analysis method. Using Fort Lee's data, the following was the KZ filter graph produced by Python.

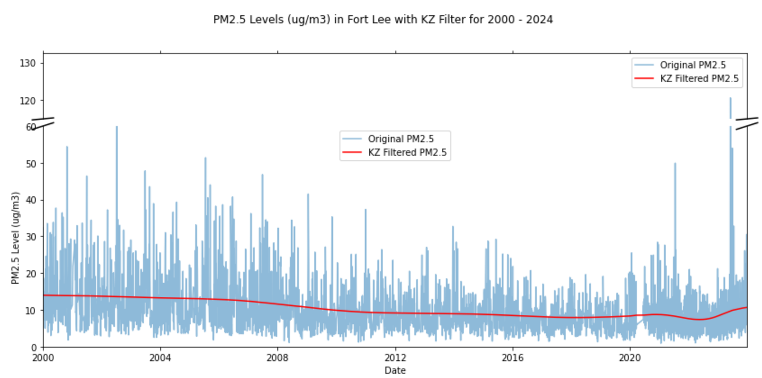

Figure 1. Long-term Trends of $PM_{2.5}$ 1-yr rolling average using KZ-filter Technique.

This graph showed the recent large spike in $PM_{2.5}$ that was observed in the air quality monitoring stations in New Jersey. Its sources can be attributed to long range transport (LRT), specifically the Canadian Wildfires.

Previously, Shuang Gao et al., (2023) was able to utilize the KZ filter analysis method to evaluate the contribution of meteorology and emissions to $PM_{2.5}$ reduction during 2018–2020 in China. Results showed that from 2018 to 2020, $PM_{2.5}$ concentrations in China decreased by 14%, with annual average concentrations of 37.19, 35.28, and 31.94 μg/m³, respectively. In 2023, Ismail Sezen et al. would also use KZ filter to quantify the long-term meteorological and emission impacts on air quality within coastal and inland cities of Turkey. They gathered PM_{10} data from 2010-2020 and KZ filter was implemented to decompose the data into its temporal components. The analysis in this study revealed that out of twelve major cities in Turkey, only three had PM_{10} levels at or below the WHO's 24-hour Air Quality Guideline (AQG) level of 45 μg/m³. Overall, during the selected period, ten out of the twelve cities showed a long-term decrease in PM_{10} concentrations, ranging from -39.6 μg/m³ to -1.2 μg/m³. The KZ filter analysis is a key method when quantifying long term meteorological and emission data.

KZ filter is useful for witnessing the long-term changes of pollutant levels in a certain area. However, the Candian Wildfires were only able to be linked with the large jump in $PM_{2.5}$ level due to scale of the event. When determining factors contributing to pollutants in a location, other methods must be used to find the source receptor relationship.

Vol. 2025 (12) 964 – 970
ISSN 2688-3651 [online]

2.2 HYSPLIT4 Back-Trajectory Modeling

"HYSPLIT4 (Hybrid Single-Particle Lagrangian Integrated Trajectory) models air parcel movement, calculating distance and direction (Draxler, 1999; Stein et al., 2015)."

$$dX(t) = V(X(t),\, t)dt\; + B(t)dW(t)$$

Equation 2. HYSPLIT4 representative particle dispersion questions where X(t), V(X(t), t), B(t), and dW(t) represent the position of the particle at time (t), the mean wind velocity at the position of the particle, the turbulence parameter, and a Wiener process which represents the random nature of turbulent motion, respectively.

Using HYSPLIT4 model, air quality can be predicted in terms of plume dispersion and air trajectory and thus be used to develop emission reduction programs for metropolitan cities and the states that have several non-attainment counties. In addition, HYSPLIT4 model has been widely used for source apportionment research due to its ability to calculate forward and backward trajectories.

2.3 Back-trajectory Endpoint Analysis

Inserting the data of high $PM_{2.5}$ days into HYSPLIT4, back-trajectories that display all the pathways of these large masses of $PM_{2.5}$ can be generated. The hundreds of spaghetti plots, dating years back, created by HYSPLIT4 can be plotted altogether with Python, distinguishing the $PM_{2.5}$ concentration as well the cardinal/ordinal direction of the back trajectories. This analysis method allows the user to source apportion clusters of $PM_{2.5}$, tracing them back to their origins. It also shows the $PM_{2.5}$ concentration of each back-trajectory so that the user can understand which trajectories have the most highly concentrated masses of $PM_{2.5}$ entering the source receptor site.

Back-trajectory analysis can be useful as policymakers and environmental agencies can use back-trajectory analysis to develop strategies for controlling and reducing air pollution. By knowing the sources and pathways of pollutants, targeted interventions can be designed. The results of the back-trajectory analysis can also be further analyzed and refined for better visualization and detailing.

2.4 Cluster Analysis

Cluster analysis is one of the methods that builds upon back-trajectory results. Utilizing the same back-trajectory endpoints that were produced by the back-trajectory plots, the average $\mu g/m^3$ of $PM_{2.5}$ and percent contribution from each direction can be calculated. These calculations are all done on Python. The averaged back-trajectories are then plotted and labelled with their corresponding averaged values on Python. This helps us to identify what cardinal or ordinal direction has the largest long-range transport impact on Upper New Jersey.

Cluster analysis has previously been utilized in cases like Daniel Choi et al., (2024) when they aimed to decrease the lack of information and analysis spanning across multiple administrative regions in South Korea. They used k-means clustering on NCEP FNL reanalysis data from 2016 to 2020 to identify mesoscale wind field patterns associated with high-$PM_{2.5}$ episodes in all 17 administrative regions of South Korea. Their statistical analysis produced region-specific wind fields for each region and identified eight representative patterns, four for Long-Range Transport (LRT) and four for Short-Term Local (STL) modes, applicable across all regions. These patterns were consistent with prior studies.

2.5 Potential Source Contribution Function Analysis

PSCF utilizes the back trajectories produced by HYSPLIT4 and overlays a grid on top of the spaghetti plots. As shown in Equation 3, it utilizes endpoints with a high $PM_{2.5}$ concentration in one grid cell and divides it by the number of endpoints within that same grid cell. PSCF value indicates the source contribution at a certain cell to the $PM_{2.5}$ concentration at the receptor site. This process is then repeated for every other grid cell mapped.

Additionally, by utilizing a Down-weight Function, the reduction of influence of data points that are less reliable is possible, thereby improving the statistical robustness and thus PSCF analysis accuracy. This is done by removing

Vol. 2025 (12) 964 – 970
ISSN 2688-3651 [online]

all grid cells with data that contain, in this study specifically, less than 20 endpoints that had a high concentration of $PM_{2.5}$ (>12µg/m^3) as seen in Table 1. The rest of the values are then divided into three parts, each part containing more highly concentrated endpoints than the last. These values were then input into Python to develop a heat map that utilizes the values to show the various hotspots in Northern New Jersey.

$$PSCF = \frac{P_{ij}}{T_{ij}}$$

Equation 3. Down weight function calculation of PSCF values for each grid (i,j) where P_{ij} = No. Of endpoints > 12 µg/m3 at grid (i,j) and T_{ij} = Total No. Of endpoints at grid (i,j).

PSCF has been used in the past (Songyan Du et al., 2007) where it was used to identify geographical sources of gas-phase polychlorinated biphenyls (PCBs) in the atmosphere of Camden, NJ. With two models and positive matrix factorization, four factors were identified. Utilizing PSCF, factors 1 and 4 showed that the source was from local sources due to the lack of distinct source regions, while factor 2 and 3 revealed that the PCBS sources were Philadelphia and southern New Jersey/southern Philadelphia respectively.

Table 1. Down-weight function: To minimize the artifacts from small values of total endpoints (n_{ij}), PSCF values were down weighted with weight function (W) when n_{ij} was less than average n_{ij}.

Down-weight Function W(n_{ij})	1.0	$50 < n_{ij}$
	0.7	$30 < n_{ij} \leq 50$
	0.3	$20 < n_{ij} \leq 30$
	0.0	$N_{ij} \leq 20$

While the back trajectory and cluster analysis graphs alone show long trajectories that are beneficial when looking at the impact of multiple states and even adjacent countries, it becomes unreliable for local or medium source contribution analysis. This is where PSCF shines and with the help of Down-weight Function, it can be used to discover the effects of large producers of concentrated pollutants like certain power plants, airports, shipyards, etc. which impact the source-receptor site.

2.6 Weighted PSCF

Weighted PSCF uses a contour plot to map $PM_{2.5}$ severity across regions. It categorizes areas as Low (0.01–0.265), Moderate (0.265–0.52), High (0.52–0.775), or Very High (0.775–1.03), with hotspots requiring urgent policy intervention.

This is the tool that bridges the connection between scientists and policy makers as it allows policy makers to see which geographic locations must be focused. Furthermore, policymakers and environmental agencies can utilize WPSCF results to implement targeted air pollution control measures. For example, if a specific industrial zone is identified as a dominant $PM_{2.5}$ contributor, stricter emission regulations or technological upgrades may be recommended to mitigate its impact.

2.7 Interconnection Between Tools

This study employed three primary analytical tools to address distinct aspects of the research question. The KZ filter was, as a temporal analysis, used to isolate and analyze long-term temporal patterns in $PM_{2.5}$ concentrations by removing short-term variability. After long-term temporal analysis, back-trajectory spatial analysis generated spaghetti plots to trace the long-range transport pathways of $PM_{2.5}$ from eight sectors. Next, to investigate mid and local contributions on a grid system, PSCF analysis was applied to identify potential geographic source regions contributing to elevated $PM_{2.5}$ levels for the past three years. Together, these tools provided a multi-dimensional understanding of pollution sources and patterns.

3. Results

3.1 HYSPLIT4 Back-trajectory Analysis

As seen on Figure 2, HYSPLIT4 plots for 2023, the North, Northwest, West, and Southwest exhibited concentrations above 'Unhealthy for Sensitive Groups', while South, East, Southeast, and Northeast mostly had 'good' to 'moderate' levels. Back-trajectory plots from 2020-2022 showed similar trajectories to the 2023 plot: only containing slightly less influence and concentration from north and northwest directions due to the Canadian Wildfire's strong influence on the 2023 plot.

Vol. 2025 (12) 964 – 970
ISSN 2688-3651 [online]

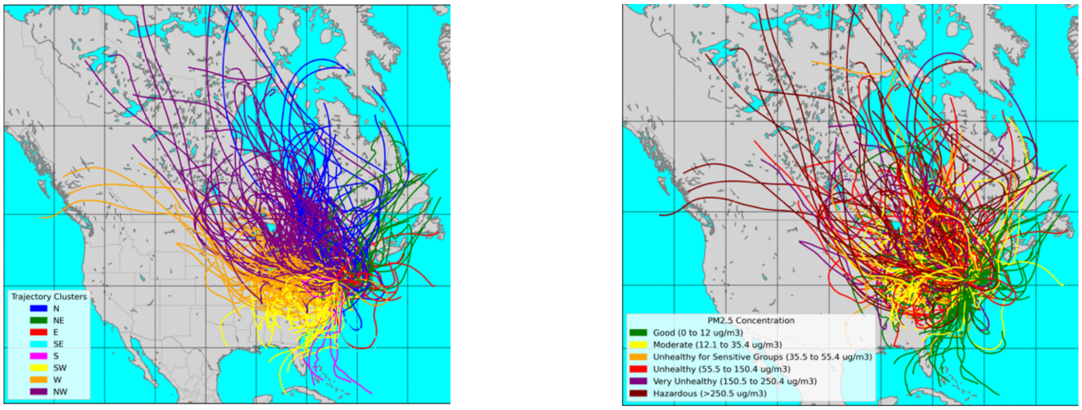

Figure 2. The back-trajectory analysis results (2023). Plots show back trajectories clusters classified by their cardinal or ordinal directions (left) or back-trajectories classified by their $PM_{2.5}$ concentration (right).

3.2 Cluster Analysis

Utilizing cluster analysis, we can see that in Figure 3, the North and Northwest were responsible for approximately 39.17% of high $PM_{2.5}$ days above 12 $\mu g/m^3$ of $PM_{2.5}$ (Long Range Transports). On the other hand, Northeast, East, and Southeast was responsible for 12.49% of high $PM_{2.5}$ days. South and Southwest were responsible for 19.43% of high $PM_{2.5}$ days (including most of the contribution from local sources).

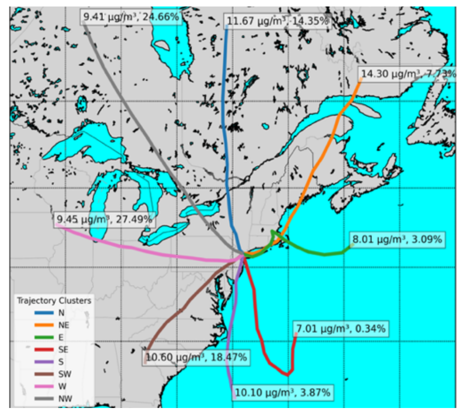

Figure 3. Cluster Analysis (2020-2023). Plot shows the percent contribution of $PM_{2.5}$ coming from each cardinal and ordinal direction. It also averages out each trajectory from a specific cardinal ordinal direction, which helps identify which areas most $PM_{2.5}$ comes from.

3.3 Potential Source Contribution Function Analysis

In Figure 4, PSCF values above 0.5 (south and southwest) indicated contributions from local sources like Newark and New York City, and mid-range sources in Pennsylvania, Virginia, and Maryland. Values below 0.5 (west, northwest, north, and northeast) pointed to the urban Midwest areas and Canadian wildfires.

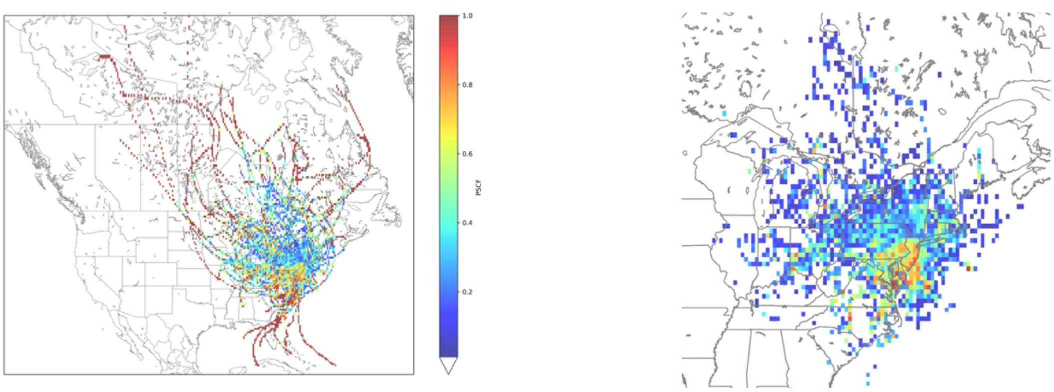

Figure 4. The PSCF analysis results (2020-2023). Plots showed the back-trajectories overlayed on grid cells that gave concentration of $PM_{2.5}$ in specific locations. Plot on the right is zoomed in and utilized the downweight function to remove less reliable data.

The Weighted Potential Source Contribution Function contour plot (Figure 5) once again created clarity on where the hot spots that contribute $PM_{2.5}$ are. It showed local and mid-range impacts of states like North Carolina, Virginia, and Maryland: the southern states.

Vol. 2025 (12) 964 – 970
ISSN 2688-3651 [online]

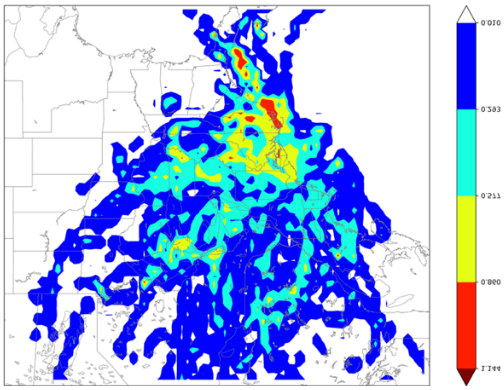

Figure 5. Kriging Analysis with further enhanced visibility to show hotspot concentrations of $PM_{2.5}$.

4. Discussion

This study was able to identify long, mid, and local range source contributors of $PM_{2.5}$. In Long-range transport PSCF and back-trajectory cluster analysis suggested that the north and northwest shared a collective $PM_{2.5}$ contribution of 39.17%, and the west accounted for 28.91% of $PM_{2.5}$ contribution, both of which are associated with critical emission sources in adjacent states. The north cluster was characterized by air parcel passing over the critical wildfire smoke transport, major urban areas, and industrialized areas in New York, U.S. and Ontario and Quebec, Canada while the northwest cluster displayed air mass traveling over major urban areas and highly industrialized areas through Pennsylvania, Illinois, Michigan, Western Ohio. To evaluate temporal characteristics of wildfire smoke transport, KZ filter analysis was carried out and the results exhibited that the recent elevation of annual rolling average was markedly attributable to the intermittent Canadian wildfire events.

Within mid-range transports, PSCF analyses indicated that the urban cities and highly industrialized areas in Delaware, Maryland, Pennsylvania, and Virginia, had a large impact on $PM_{2.5}$ concentrations observed in Northern New Jersey. Figure 4 and 5 clarified that mid-range transports, particularly near Virginia, held the largest contribution of $PM_{2.5}$ within Fort Lee, NJ.

Using PSCF analysis, in the vicinity of the Northern New Jersey, there were elevated $PM_{2.5}$ concentrations predominantly originating from NYC urban areas, industrial facilities, airports, and shipyards contributing significantly to $PM_{2.5}$ level at Fort Lee, NJ. Figure 4 and 5 were, again, also able to clarify that local range sources contributed roughly the second largest concentration of $PM_{2.5}$ to Fort Lee, NJ. Calculating significance of local, mid-range, and long range sources can be further studied to precisely quantify percent contributions from one another.

5. Limitations

One major limitation when utilizing these analysis methods is considering the rural and suburban areas that do not have nearby monitoring sites. While these findings utilized Fort Lee's weather monitoring site, which was perfectly situated for analyzing the source-receptor relationship of $PM_{2.5}$ in upper New Jersey, many locations in the U.S. do not have such assets creating lesser accurate analysis results depending on proximity to the closest monitoring sites. This research also only used $PM_{2.5}$ mass and thus investigated only geographical contributions. To create further understanding of pollutants, chemical speciation must be used. Chemical speciation allows researchers to break down pollutants and identify what the exact chemical compounds/elements exist within them to link them to specific sources such as power plants, transportation emissions, steel smelters, or agricultural burnings etc.

6. Conclusion

This research lays down the foundation for further $PM_{2.5}$ source contribution research not only in New Jersey, but for the rest of the U.S. It brings forth immense possibilities such as guiding zoning laws, allowing for specific locations to be regulated, and in combination with chemical speciation, teaches policy makers where and what is polluting locations like Fort Lee. Understanding the damage that pollutants cause is important, but likewise preventing its spread is right up there alongside. By producing efficient and direct policies, we can ensure a safer and cleaner home.

References:

Burke, M., et al. (2023, September 20). The contribution of wildfire to PM2.5 trends in the USA. *Nature*, 622(7984), 761–766. https://doi.org/10.1038/s41586-023-06522-6.

Vol. 2025 (12) 964 – 970
ISSN 2688-3651 [online]

Choi, D., et al. (2024, September 24). Mesoscale wind field patterns conducive to the high-PM2.5 episodes over South Korea: Cluster analysis, *Atmospheric Environment*, 333, 120653, ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2024.120653.

Download Daily Data. (2024, May). United States Environmental Protection Agency. Retrieved December 29, 2024 from https://www.epa.gov/outdoor-air-quality-data/download-daily-data.

Draxler, R. R., (1999). HYSPLIT4 user's guide. NOAA Tech. Memo. ERL ARL-230, NOAA Air Resources Laboratory, Silver Spring, MD.

Du, S. & Rodenburg L., A. (2007, December). Source identification of atmospheric PCBs in Philadelphia/Camden using positive matrix factorization followed by the potential source contribution function, *Atmospheric Environment*, 41(38), 8596-8608, ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2007.07.042.

Gao, S., et al. (2023, November). Meteorological influences on PM2.5 variation in China using a hybrid model of machine learning and the Kolmogorov-Zurbenko filter, *Atmospheric Pollution Research*, 14(11), 101905, ISSN 1309-1042, https://doi.org/10.1016/j.apr.2023.101905.

Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. National Centers for Environmental Information (NCEI), NOAA. Retrieved December 29, 2024 from https://www.ncei.noaa.gov/access/search/data-search/daily-summaries.

HYSPLIT4 Back-trajectory Modeling National Ocean and Atmospheric Association. Retrieved from https://www.ready.noaa.gov/hypub-bin/trajtype.pl?runtype=archive.

Particulate Matter (PM) Basics. (2024, June). United States Environmental Protection Agency. Retrieved December 29, 2024 from https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM.

Sezen, I., et al. (2023, June). Application of Kolmogorov-Zurbenko filter to quantify the long-term meteorological and emission impacts on air quality, *Atmospheric Research*, 288, 106714, ISSN 0169-8095, https://doi.org/10.1016/j.atmosres.2023.106714.

Stein, A. F., et al. (2015, December 1). NOAA's HYSPLIT atmospheric transport and dispersion modeling system, *Bulletin of the American Meteorological Society*, 96, 2059-2077, http://dx.doi.org/10.1175/BAMS-D-14-00110.1.