JOURNAL OF
EMERGING INVESTIGATORS

Article

The relationship between meteorological factors and air quality in Washington D.C. metropolitan area

Yanling Lin[1], Yixi Kennedy[2]
[1] Thomas Jefferson High School for Science and Technology, Alexandria, Virginia
[2] Polygence Research Program, Menlo Park, California

SUMMARY

Air pollution can lead to substantial health issues, contributing to 10 percent of deaths worldwide. Past research has shown that meteorological factors influence particulate matter concentrations and, consequently, air quality. We hypothesized that stronger wind speed and higher precipitation improve air quality by dispersing or removing pollutants while higher temperature worsens air quality by increasing ozone and fine particle pollution. We used air quality data from the Environmental Protection Agency (EPA) and meteorological factor data from the National Oceanic and Atmospheric Administration (NOAA) in the Washington D.C. metropolitan area in 2020-2024 to analyze associations between specific weather conditions (wind speed, precipitation, and temperature) and air quality metrics, including fine particulate matter (PM2.5) and Air Quality Index (AQI). We found that wind speed was negatively associated with PM2.5 and AQI (p<0.01). After controlling for seasonality, we found a 1% increase in temperature was associated with 0.56% increase in PM2.5 and 0.51% increase in AQI (R^2=0.301 and R^2=0.294, respectively). Furthermore, our results suggested that while summer was associated with better air quality than winter, a 1% increase in temperature during summer was associated with a 1.97% increase in PM2.5 and 1.75% increase in AQI (R^2=0.301 and R^2=0.294, respectively); during winter, a 1% increase in temperature was associated with 1.55% decrease in PM2.5 and 1.43% decrease in AQI (R^2=0.301 and R^2=0.294, respectively). Our study contributes to improving the understanding of the role of meteorological factors, particularly temperature, in air quality. This is especially important as heat waves become more frequent with global warming.

INTRODUCTION

Poor air quality, which is associated with particle pollution, is a growing concern to public health and, thus, has become a particular focus of scientific studies (1-3). Particulate matter (PM) is often a by-product of burning wood or fossil fuels emitted from factories, power plants, diesel- and gasoline-powered vehicles and equipment, or wildfires (1). A high level of airborne PM is associated with an increase of the risk of health problems (4). PM2.5 is a type of particulate matter, a mixture of solid particles and liquid droplets suspended in the air, that is 2.5 micrometers or less in diameter (5). These tiny particles penetrate deep into the respiratory system and damage health.

The Air Quality Index (AQI) includes six pollutants: PM2.5, PM10, carbon monoxide, sulfur dioxide, nitrogen dioxide, and ground-level ozone. The AQI scale ranges from 0 to 500 split into six categories. Zero indicates perfect air quality, while a score of 500 indicates a severe threat to public health. AQI is typically interpreted as follows: 0–50 is good air quality, 51–100 is moderate, 101–150 is unhealthy for sensitive individuals, 151–200 is unhealthy, 201–300 is very unhealthy, and 301–500 is classified as hazardous. For example, PM2.5 penetrates the respiratory system into the lungs and generates a range of diseases such as respiratory diseases, cardiovascular diseases, reproductive diseases, neurological disorders, and cancers (6). When people, especially infants and youth, are exposed to PM2.5 pollution for a prolonged period, their health and wellbeing may suffer significantly (7). According to a World Health Organization (WHO) study, outdoor and indoor air pollution combined is associated with 6.7 million premature deaths annually (8). Furthermore, according to a 2019 World Bank estimate, exposure to polluted air led to health costs of $8.1 trillion, or 6.1 percent of the total gross production worldwide (9). Air pollution increases mortality from respiratory diseases, while severe air pollution increases automobile accidents, reduces urbanites' happiness, and causes major economic damage (2).

Weather conditions are known to have an impact on air quality affecting pollutant concentrations and the amounts of particulates (3). Meteorological factors are closely related to the formation and removal of particles through condensation, coagulation, and diffusion mechanisms (10). In the United States, up to 50% of daily variations in PM2.5 can be explained by weather conditions including wind speed, temperature, and precipitation (11, 12). Typically, higher wind speeds can help disperse the particles, potentially contributing to lower PM2.5 levels. Additionally, precipitation can remove atmospheric particulate matter and reduce PM2.5 concentration. However, higher heat traps the pollutants and leads to worse air quality (13). Other weather conditions, including humidity, air pressure, wind direction, and solar radiation also affect PM2.5 concentration (14). Meanwhile, the associations between many weather conditions and air quality often vary in different seasons, with worse pollution in winter (15).

We hypothesized that three factors, wind speed, precipitation, and temperature, affect the air quality in the Washington D.C. metropolitan area. More specifically, we hypothesized that 1) higher wind speed lowers PM2.5 concentration and AQI, 2) more precipitation reduces PM2.5 concentration and AQI, and 3) higher temperature increases PM2.5 concentration and AQI, especially in summer when temperature is already high. We used the daily records from 2020-2024 of PM2.5 concentration and AQI from the

https://doi.org/10.59720/25-084

Environmental Protection Agency (EPA) and wind speed, temperature, and precipitation from the National Oceanic and Atmospheric Administration (NOAA) recorded near Reagan Washington National Airport (16-17). We analyzed the impact between daily weather conditions and AQI using multivariate linear regressions. Our results showed that wind speed was negatively associated with PM2.5 and with AQI, while temperature was positively associated with PM2.5 and AQI with seasonal variation. Precipitation was not significantly associated with PM2.5 and AQI. Our study contributes to the understanding of how meteorological factors influence air quality in the Washington D.C. metropolitan area.

RESULTS

To characterize the relationship between changes in PM2.5 and AQI and temperature, we examined the daily mean levels of PM2.5 and AQI from January 2020 to July 2024 to assess the temporal variations. The mean levels of PM2.5 and AQI were 7.5 and 37.4, respectively. The levels of PM2.5 and the AQI were relatively consistent throughout the period of this study, with exceptions during June 2023 when wildfire smoke in Canada and New Jersey spread through the Washington D.C. region (**Figure 1**).

Wind speed, temperature, and precipitation showed clear patterns of variation from 2020-2024. Wind speed varied with a general declining trend between 2020-2024 (**Figure 2**). Similarly, precipitation followed a slight declining trend with the most recent years witnessing lower levels of precipitation on rainy days (**Figure 3**). As expected, temperature had a clear seasonal pattern, with higher PM2.5 and AQI during summer and lower during winter (**Figure 4**).

The correlation between the weather conditions and air quality variables were as expected. PM2.5 concentration was positively correlated with AQI, with a high coefficient value of 0.90 (**Figure 5**). Wind speed and precipitation were both negatively correlated with PM2.5, with coefficients of -0.25 and -0.06, respectively (**Figure 5**). Coefficient values range from -1 (perfect negative correlation) to +1 (perfect positive correlation), so a coefficient value of 0.90 is a strong positive correlation, while a coefficient value of -0.25 is a weak negative correlation, and -0.06 indicates negligible correlation. Meanwhile, all three temperature measures - daily average, minimum, and maximum temperature – were positively correlated with PM2.5, with coefficients of 0.13, 0.11, and 0.15, respectively. As expected, the three measures of daily temperature were highly correlated with each other, with coefficients of 0.94, 0.97 and 0.98, respectively (**Figure 5**).

To capture the potential non-linear relations and skewed distribution of variables, logarithmic fits were tested to examine the relationship between weather conditions and the two measures of air quality, PM2.5 and AQI, respectively (**Table 1-2**). The regression results showed that wind speed had a significant negative relationship with PM2.5 concentration and AQI ($p < 0.01$), while the link between precipitation or temperature and PM2.5 or AQI was not significant ($p > 0.10$) (**Table 1-2**). The negative impact of wind speed on PM2.5 was consistent with the general observation that higher wind speed helped disperse pollutants and improve air quality in Washington D.C. metropolitan area.

To capture the possible effect of the strong seasonal variation of temperature, we introduced four seasonal dummy variables – spring (March-May), summer (June-August), fall (September-November), and winter (December-February) – as well as their interactions with temperature into the regression (**Table 1-2**). More specifically, after controlling for the effects of precipitation temperatures, seasonality, and interactive terms between temperature and seasonality, a 1% increase in wind speed was associated with 0.49% decrease in PM2.5 and 0.45% decrease in AQI ($p < 0.01$ for both, and $R^2 = 0.301$ and 0.294, respectively). Precipitation was also negatively associated with the PM2.5 level and AQI, but not in a statistically significant manner. Temperature was positively associated with PM2.5 and AQI, and it was statistically significant when seasonality and the interactions between temperature with seasonality were accounted for. After controlling for other variables, higher temperature were typically associated with higher PM2.5 concentration, and this was particularly the case during summer when temperature was already high. A 1% increase in temperature was associated with 0.56% increase in PM2.5 and 0.51% increase in AQI ($p < 0.05$ and $p < 0.10$, and $R^2 = 0.301$ and 0.294, respectively); during summer, a 1% increase in temperature

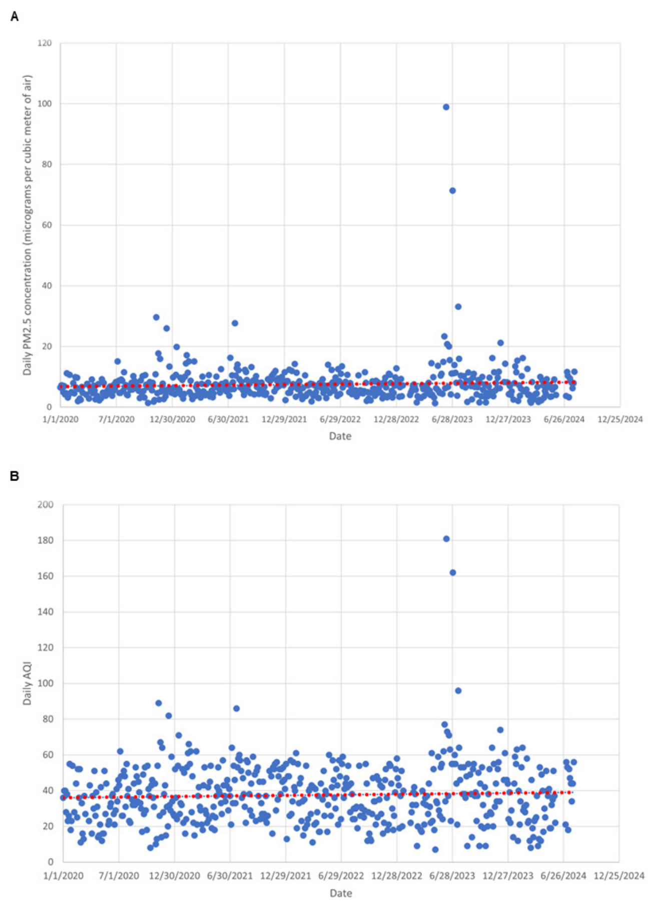

Figure 1: Air pollution increased slightly with outliers during periods with external shocks. Changes in air quality over time (n=517). Air quality measured by daily A) PM2.5 concentration (micrograms per cubic meter), and B) AQI. Both datasets are fitted with linear trend line and were recorded at the Aurora Hills Visitors Center. Source: EPA (16).

https://doi.org/10.59720/25-084

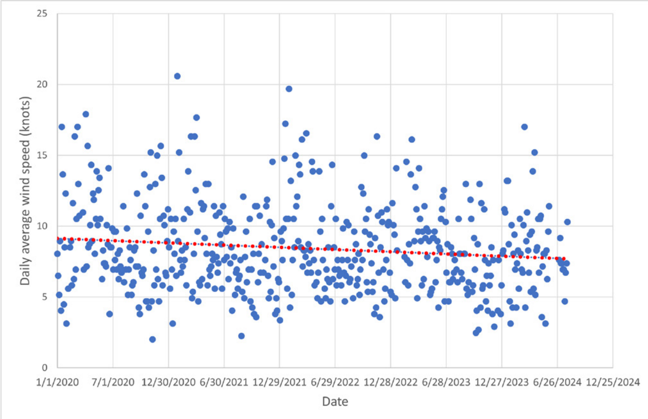

Figure 2: Wind speed declined. Daily wind speed, recorded at the Washington National Reagan Airport, with a fitted linear trend line over time (n=517). Source: NOAA (17).

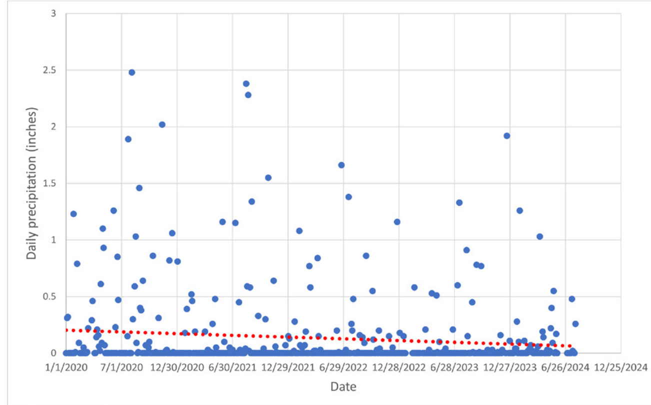

Figure 3: Precipitation declined. Daily precipitation, recorded at the Washington National Reagan Airport, with a fitted linear trend line over time (n=517). Source: NOAA (17).

was associated with 1.97% increase in PM2.5 and 1.75% increase in AQI (p < 0.05 and p < 0.01, and R^2 = 0.301 and 0.294, respectively), while during winter, a 1% increase in temperature was associated with 1.55% decrease in PM2.5 and 1.43% decrease in AQI (p < 0.01 for both, and R^2 = 0.301 and 0.294, respectively). After controlling for the seasonal effect, the positive relationship between temperature and PM2.5 or AQI became significant (p < 0.10 and p < 0.05, respectively).

DISCUSSION

There is extensive literature examining the impact of meteorological factors on air quality (18-20). Many studies, particularly those covering large countries such as China, India, and the United States, reflected on the regional variations in addition to seasonality. For example, in China, a regional analysis of Jiangsu, Anhui, Shandong, and Henan using geographical detectors found that average temperature was the primary factor influencing PM2.5 concentrations (18). Furthermore, a study of urban and rural centers in India, including Delhi, Bengaluru, Ahmedabad, Kolkata, and Gadanki, found that meteorological factors were key drivers of PM2.5 levels, with substantial regional and seasonal variation (19). In the U.S., a study in California found that temperature, extreme drought, and fire-burned area was positively associated with both PM2.5 and ozone levels, while wind speed had a negative association, showing distinct spatial and seasonal patterns (20).

The association between wind speed and air quality was significant but that between precipitation and air quality was not. Because wind continuously disperses pollutants over wide areas, it more effectively reduces concentrations—including fine particles like PM2.5—while rain primarily removes larger particles and has only a limited effect on finer pollutants. However, on some occasions, strong wind might bring in pollutants such as from wildfire depending on the direction, as the distribution of fine particulates is heavily influenced by the direction of incoming wind.

Summer (after controlling for the effect of temperature, wind speed, and precipitation and its interaction with temperature) tended to be associated with better air quality than winter. Part of the reasons could be related to the increased heating use in winter days and stable atmospheric structure and cold temperatures prevented diffusion, so the pollutants "stayed in place" (21).

Our results suggested higher temperatures were associated with poorer air quality. Summer was associated with lower PM2.5 and AQI (better air quality) while winter was associated with higher PM2.5 and AQI (worse air quality). The interactive terms between temperature and seasonal dummy for summer (June-August) were positive and those between temperature and seasonal dummy for winter were negative, which suggested the damaging effect on air quality of higher temperature was particularly strong during summer, others being equal. Temperature impacts air quality more strongly in summer than in winter because high heat and sunlight accelerate photochemical ozone production. Heat acted as a chemical catalyst and converted existing elements in the air into ozone and other harmful byproducts (22). For example, in the U.S., summer heat waves can raise the risk of severe ozone pollution by up to seven times, whereas winter inversions may double particulate pollution risk, though with a less pronounced effect (23). Extreme heat waves, such as over 100°F, were often associated with worsened air quality

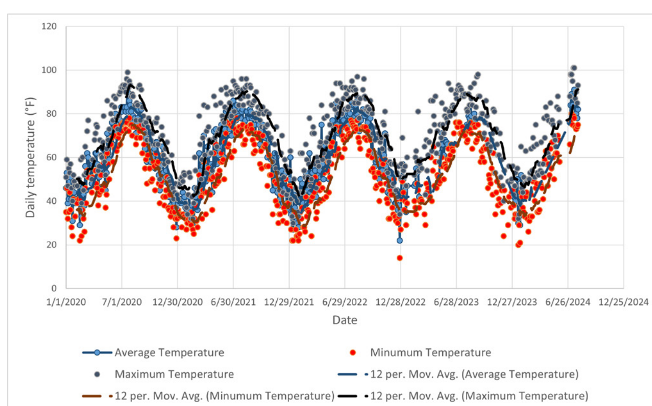

Figure 4: Temperature oscillated, high in summers and low in winters. Daily max, min, and average temperature recorded at the Washington National Reagan Airport (n=517). Source: NOAA (17).

JOURNAL OF
EMERGING INVESTIGATORS
https://doi.org/10.59720/25-084

(24). One notable example of such shocks occurred in June 2023, when wildfire smoke in Canada and New Jersey produced widespread ozone across the Mid-Atlantic and Great Lakes regions, and as it spread southward, it led to poor air quality in Washington D.C. metropolitan area, where PM2.5 concentration skyrocketed to 98.9; and AQI to 181, indicating unhealthy conditions (25).

The results must be interpreted with caution as the analysis was subject to several limitations. The R^2 value of 0.301 and 0.294 for both regressions on PM2.5 and AQI were relatively low, which suggested the share of the variation of PM2.5 and AQI explained by the meteorological factors and seasonal dummy variables as expressed by the regressions were relatively low, which likely reflects the limited set of

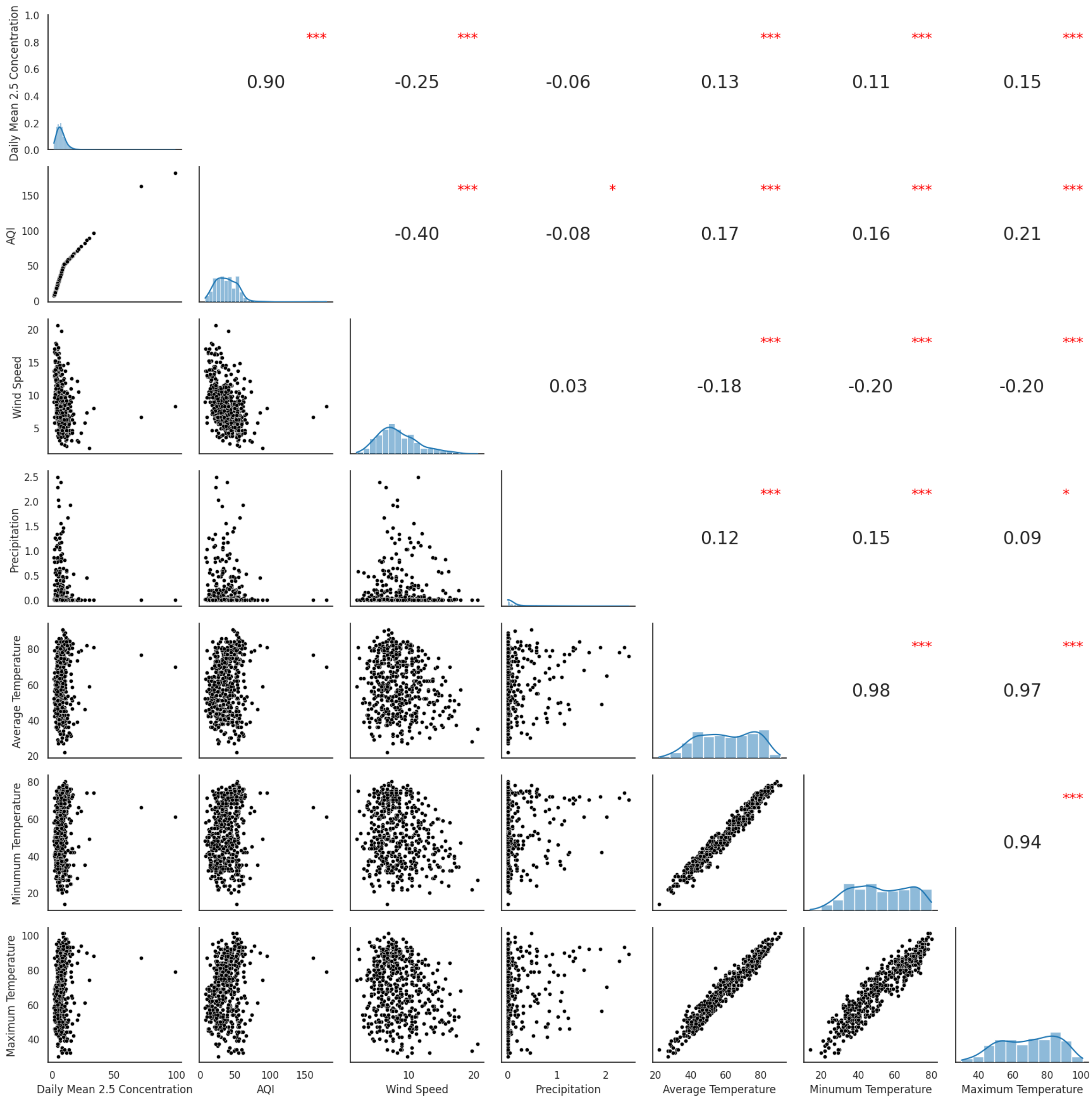

Figure 5: Meteorological factors, including wind speed, precipitation, and temperatures, correlate with air quality, measured by PM2.5 concentration and AQI. Daily average temperature was calculated as the mean of maximum and minimum temperatures. Each chart in the bottom half of the figure shows the relationship between the variable in each row and column. A positive slope indicates a positive correlation between the two variables, and a negative slope indicates a negative correlation between the two variables. For example, the top left chart shows a positive correlation between PM 2.5 concentration and AQI. As PM 2.5 concentration increases, AQI increases. The charts on the diagonal show the histogram distribution of each variable. The top first row shows the correlation between each variable and PM 2.5 concentration. Asterisks indicate statistical significance of the Pearson correlation: *p < 0.10, **p < 0.05, and ***p < 0.01.

https://doi.org/10.59720/25-084

meteorological variables, lack of interaction effects, and the use of only linear, real-time relationships. In other words, other factors not captured by the regressions were likely to play an important role in determining air quality. Specifically, this was related to several reasons. First, the analysis included only three meteorological factors – wind speed, precipitation, and temperature. Due to data constraints, the analysis did not include other meteorological factors, such as humidity, air pressure, wind direction, and solar radiation, although they also likely affect air quality. Studies show that higher humidity, higher air pressure, and lower temperatures are often associated with poorer air quality (26-29). During pollution episodes, low temperature, reduced solar radiation, and high humidity can reinforce one another, further elevating PM2.5 concentrations (26). For example, high humidity and air pressure can trap more PM in the air causing them to accumulate, strong wind can blow in pollutants from afar or disperse them, and solar radiation can influence the rate of the chemical reactions that produce the smog from other pollutants (27). Second, our analysis did not account for how the different meteorological factors interact and how their interactions affect air quality. For example, wind speed had greater impacts on PM2.5 concentration when humidity was low and solar radiation had more influence when humidity was high (28). Third, we tested only the linear regressions with real time effect while some meteorological factors can influence PM2.5 or AQI in a non-linear manner or with delayed effects. For example, the effects of precipitation, wind, and humidity on air quality may be non-linear, delayed, and lasting for several days (29).

In addition, the analysis showed that several spikes of PM2.5 and AQI were associated with the spread of wildfire smoke, factors exogenous to the Washington D.C. metropolitan area. An increase in temperature had a significant negative effect on air quality during summer months.

The feedback mechanisms between air pollution and climate change are complex. For example, one study estimates that anthropogenic climate change contributed, on average, 49% to smoke-related PM2.5 concentrations in the western United States between 1997 and 2020 and accounted for 58% of the observed increase from 2010 to 2020 (30). The negative effect of meteorological factors on air quality can be non-linear and substantially increase after certain breaking points, resulting in high economic and human costs (31). As climate change and global warming alter meteorological conditions, weather patterns change, affecting how often extreme weather events occur, how severe they are, and how long they last. This suggests more attention should be focused on the effect on air quality of temperature and its interaction with other factors (32).

Poor air quality has a harmful impact on health. A clear understanding of how the meteorological factors influence air quality is important to make informed decisions to mitigate the negative impact and protect the vulnerable population. Our study, using publicly available daily air quality data from the EPA and meteorology data from the NOAA near Reagan National Airport in the Washington D.C. metropolitan area, showed wind speed and precipitation reduce PM2.5 concentration and AQI (16-17). After controlling for seasonality and interactions, our results indicated an increase in temperature resulted in a significant worsening of air quality, particularly during summer. This helps enhance understanding of how meteorological factors, particularly temperature, impact air quality, which is particularly crucial as heat waves intensify with global warming.

Future research should focus on expanding the data coverage and incorporating more meteorological factors into the analysis to comprehensively test the effect of each factor and their interactions. Future research should go beyond linear regressions and test different forms, including for example quadratic relationships to determine the turning points and lagged models to identify the delay and duration of the lasting effect of meteorological factors.

MATERIALS AND METHODS

Data

Two main sources of data were used: air quality data measured by PM2.5 concentration and AQI from the EPA; and weather conditions, including wind speed, temperature, and precipitation, from the NOAA (16-17). The data covered the most recent five years, from January 1, 2020, through July 31, 2024. Wind speed was measured in knots as nautical miles per hour, temperature in degrees Fahrenheit, and precipitation

	log(PM2.5) (1)	log(PM2.5) (2)	log(PM2.5) (3)	log(PM2.5) (4)	log(PM2.5) (5)
log(Wind speed)	-0.56***			-0.54***	-0.49***
	(-5.58)			(-5.33)	(-4.82)
log(Precipitation)		-0.18		-0.13	-0.16
		(-1.33)		(-1.04)	(-1.29)
log(Temperature)			0.13	0.01	0.56**
			(0.88)	(0.039)	(1.95)
Spring					-0.98
					(-0.51)
Summer					-7.91*
					(-1.92)
Fall					1.82
					(0.82)
Winter					6.87***
					(4.09)
Spring*log(Temperature)					0.43
					(0.92)
Summer* log(Temperature)					1.97**
					(2.09)
Fall* log(Temperature)					-0.28
					(-0.53)
Winter* log(Temperature)					-1.55***
					(-3.70)
Constant	2.99***	1.87***	1.28**	2.98***	-0.20
	(13.99)	(32.28)	(2.12)	(4.62)	(-0.16)
R^2	0.183	0.013	0.006	0.189	0.301

Table 1: Weather conditions, including wind speed, precipitation, and temperature, influence PM2.5 concentration. Multivariate linear regressions were generated with x=wind speed, precipitation, temperature, seasonal dummies, and / or interactions between temperature and seasonal dummies, and y=PM2.5 concentration. Regression (1) estimated log(PM2.5) as a function of log(wind speed), with a constant term; Regression (2) estimated log(PM2.5) as a function of log(precipitation), with a constant term; Regression (3) estimated log(PM2.5) as a function of log(temperature), with a constant term; Regression (4) estimated log(PM2.5) as a function of log(wind speed), log(precipitation), and log(temperature), with a constant term; and Regression (5) estimated log(PM2.5) as a function of log(wind speed), log(precipitation), and log(temperature), as well as the four seasonal dummies and the interaction terms between log(temperature) and each seasonal dummy, with a constant term. The number of observations for all these regressions is 141. The constant term indicates the expected value of Y when X is zero. Coefficients are indicated outside of parentheses and Student's *t*-values in parentheses with $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$. For example, in Regression (1), the estimated coefficient of log(wind speed) on log(PM2.5) is -0.56 ($t = -5.58$), indicating significance at the 1% level.

https://doi.org/10.59720/25-084

	log(AQI) (1)	log(AQI) (2)	log(AQI) (3)	log(AQI) (4)	log(AQI) (5)
log(Wind speed)	-0.51***			-0.50***	-0.45***
	(-5.53)			(-5.25)	(-4.76)
log(Precipitation)		-0.16		-0.12	-0.14
		(-1.30)		(-1.04)	(-1.27)
log(Average Temperature)			0.14	0.03	0.51*
			(1.03)	(0.21)	(1.92)
Spring					-0.55
					(-0.31)
Summer					-6.64*
					(-1.74)
Fall					1.81
					(0.89)
Winter					6.68***
					(4.30)
Spring*log(Temperature)					0.40
					(0.93)
Summer* log(Temperature)					1.75**
					(2.01)
Fall* log(Temperature)					-0.20
					(-0.42)
Winter* log(Temperature)					-1.43***
					(-3.68)
Constant	4.58***	3.55***	2.93***	4.48***	1.30
	(23.25)	(66.71)	(5.27)	(7.52)	(1.11)
R^2	0.180	0.012	0.008	0.187	0.294

Table 2: Weather conditions, including wind speed, precipitation, and temperature, influence AQI. Multivariate linear regressions were generated with x=wind speed, precipitation, temperature, seasonal dummies, and / or interactions between temperature and seasonal dummies, and y=AQI. Regression (1) estimated log(AQI) as a function of log(wind speed), with a constant term; Regression (2) estimated log(AQI) as a function of log(precipitation), with a constant term; Regression (3) estimated log(AQI) as a function of log(temperature), with a constant term; Regression (4) estimated log(AQI) as a function of log(wind speed), log(precipitation), and log(temperature), with a constant term; and Regression (5) estimated log(AQI) as a function of log(wind speed), log(precipitation), and log(temperature), as well as the four seasonal dummies and the interaction terms between log(temperature) and each seasonal dummy, with a constant term. The number of observations for all these regressions is 141. The constant term indicates the expected value of Y when X is zero. Coefficients are indicated outside of parentheses and Student's *t*-values in parentheses with $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$. For example, in Regression (1), the estimated coefficient of log(wind speed) on log(PM2.5) is -0.51 (t = -5.53), indicating significance at the 1% level.

in inches. Air quality was measured by PM2.5 concentration and AQI, with higher values indicating lower air quality. The PM2.5 concentration was measured in micrograms per cubic meter.

The PM2.5 data was recorded at the Aurora Hills Visitors Center (site latitude: 38.8577; site longitude: -77.0592) every three days. The weather conditions were recorded at Washington National Reagan Airport (site latitude: 38.8472; site longitude: -77.0345) daily, using the method of R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC (United States EPA) (16). The direct distance between the two sites is 1.02 miles, and the Aurora Hills Visitors Center is the location with air quality information closest to the Washington National Reagan Airport.

To examine this impact of local weather conditions on air quality in the Washington D.C. metropolitan area near Washington National Reagan Airport, we merged the weather condition data with the air quality data during the period of January 2020 through July 2024 and constructed a dataset of 517 observations with all five variables: wind speed, precipitation, temperature, PM2.5, and AQI. As a result, data was available every three days, excluding dates with weather data but no corresponding PM2.5 information. From the weather condition data set, several days only had maximum temperature and minimum temperature but not average temperature. We calculated the daily temperature as the mean of the maximum temperature and minimum temperature in order to maximize the number of observations in the analysis. For the precipitation variable, the days with zero precipitation were dropped from the analysis after transforming to the logarithmic forms. We excluded observations with incomplete information of at least one of these five variables or with zero precipitation, and 141 observations remained for analysis.

Methodology

All data analysis was conducted using Excel and Python (version 3.13). Linear regressions in logarithmic forms were performed for all independent variables x_i*={wind speed, precipitation, temperature}* and dependent variables *y={PM2.5, AQI}*, first for each independent variable then for all three independent variables, as specified in the equations below:

$$y = a + b_i \times x_i \text{ (Eqn. 1)}$$

$$y = a + \sum b_i \times x_i \text{ (Eqn. 2)}$$

Seasonal dummies for spring (March – May), summer (June – August), fall (September – November), and winter (December – February) along their interactive terms with temperature were introduced to control for the effect of seasonal variation, using the equation below:

$$y = a + \sum b_i \times x_i + \sum c_i \times seasonal\ dummy_i + \sum d_i \times seasonal\ dummy_i \times temperature \text{ (Eqn. 3)}$$

a is the constant, b_i stands for the coefficient of the independent variables (including wind speed, precipitation, and temperature), c_i is the coefficient of the seasonal dummies, and d_i is the coefficient of the interactive terms. We performed a Student's *t*-test to test for significance and used the *R*-squared value to predict the proportion of the variation in the dependent variable attributable to the independent variables as specified in the regressions.

ACKNOWLEDGMENTS

The authors sincerely thank Ms. Olya Rachtchinine for her guidance and support on this research.

Received: March 2, 2025
Accepted: September 6, 2025
Published: July 22, 2026

REFERENCES

1. "Health Impact of Air Pollution." *American Lung Association*. https://www.lung.org/research/sota/health-risks. Accessed 3 January 2025.
2. Liu, Yansui, *et al*. "Exploring the Relationship between Air Pollution and Meteorological Conditions in China under Environmental Governance." *Scientific Reports*, vol. 10, no. 1, 3 Sept. 2020, https://doi.org/10.1038/s41598-020-71338-7.
3. "How Weather Affects Air Quality." University Corporation for Atmospheric Research Center for Science Educa-

https://doi.org/10.59720/25-084

tion, https://scied.ucar.edu/learning-zone/air-quality/how-weather-affects-air-quality. Accessed 3 January 2025.
4. "Research on Health Effects from Air Pollution." *United States Environmental Protection Agency*, https://www.epa.gov/air-research/research-health-effects-air-pollution. Accessed 3 January 2025.
5. "Particulate Matter (PM) Basics." *United States Environmental Protection Agency*, https://www.epa.gov/pm-pollution/particulate-matter-pm-basics. Accessed 6 January 2026.
6. Manisalidis, Ioannis, *et al*. "Environmental and Health Impacts of Air Pollution: A Review." *Frontiers in Public Health*, vol. 8, 20 Feb. 2020, https://doi.org/10.3389/fpubh.2020.00014.
7. Tian, Li, *et al*. "Spatiotemporal Changes in PM2.5 and Their Relationships with Land-Use and People in Hangzhou." *International Journal of Environmental Research and Public Health*, vol. 15, no. 10, 8 Oct. 2018, p. 2192, https://doi.org/10.3390/ijerph15102192.
8. "Ambient (outdoor) Air Pollution." World Health Organization, 24 Oct. 2024, https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health. Accessed 3 January 2025.
9. Awe, Yewande Aramide, *et al*. *The Global Health Cost of PM2.5 Air Pollution : A Case for Action Beyond 2021 (English).* International Development in Focus Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/455211643691938459
10. Hua, Yang, *et al*. "Investigating the Impact of Regional Transport on PM2.5 Formation Using Vertical Observation during APEC 2014 Summit in Beijing." *Atmospheric Chemistry and Physics*, vol. 16, no. 24, 15 Dec. 2016, pp. 15451-60, https://doi.org/10.5194/acp-16-15451-2016.
11. Tai, Amos P.K, *et al*. "Correlations between Fine Particulate Matter (PM2.5) and Meteorological Variables in the United States: Implications for the Sensitivity of PM2.5 to Climate Change." *Atmospheric Environment,* vol. 44, no. 32, Oct. 2010, pp. 3976-84, https://doi.org/10.1016/j.atmosenv.2010.06.060.
12. Chen, Ziyue, *et al*. "Influence of Meteorological Conditions on PM2.5 Concentrations across China: A Review of Methodology and Mechanism." *Environment International,* vol. 139, June 2020, p. 105558, https://doi.org/10.1016/j.envint.2020.105558.
13. Wang, Jianhua, and Susumu Ogawa. "Effects of Meteorological Conditions on PM2.5 Concentrations in Nagasaki, Japan." *International Journal of Environmental Research and Public Health*, vol. 12, no. 8, 3 Aug. 2015, pp. 9089-101, https://doi.org/10.3390/ijerph120809089.
14. Nguyen, Giang Tran Huong, *et al*. "An Exploration of Meteorological Effects on PM2.5 Air Quality in Several Provinces and Cities in Vietnam." *Journal of Environmental Sciences*, vol. 145, Nov. 2024, pp. 139-51, https://doi.org/10.1016/j.jes.2023.07.020.
15. Chen, Tao, *et al*. "Spatial and Temporal Variations of PM2.5 and Its Relation to Meteorological Factors in the Urban Area of Nanjing, China." *International Journal of Environmental Research and Public Health*, vol. 13, no. 9, 16 Sept. 2016, p. 921, https://doi.org/10.3390/ijerph13090921.
16. United States Environmental Protection Agency (EPA) Air Quality Data. https://www.epa.gov/outdoor-air-quality-data/download-daily-data. Accessed 3 January 2025.
17. National Centers for Environmental Information National Oceanic and Atmospheric Association (NOAA) Climate Data Online. https://www.ncdc.noaa.gov/cdo-web/. Accessed 3 January 2025.
18. Xia, Xiaoli, *et al*. "Study on the Spatio-temporal Characteristics and Driving Factors of Pm2.5 in the Inter-provincial Border Region of Eastern China (Jiangsu, Anhui, Shandong, Henan) from 2022 to 2024." *Atmosphere*, vol. 16, no. 8, 2025, p. 895, https://doi.org/10.3390/atmos16080895.
19. Krishnaveni, A. Sai, *et al*. "Spatial, Temporal Features and Influence of Meteorology on PM2.5 and O3 Association across Urban and Rural Environments of India." *Atmospheric Environment: X*, vol. 22, Apr. 2024, p. 100265, https://doi.org/10.1016/j.aeaoa.2024.100265.
20. Pan, Jianan, *et al*. "Spatio-Temporal Joint Analysis of PM2.5 and Ozone in California with INLA." *ArXiv.org*, 2024, https://arxiv.org/abs/2404.14446.
21. Wang, Ju, *et al*. "Impact Analysis of Meteorological Variables on PM2.5 Pollution in the Most Polluted Cities in China." *Heliyon*, vol. 9, no. 7, July 2023, p. e17609, https://doi.org/10.1016/j.heliyon.2023.e17609.
22. IQAir Staff Writers. "How Heat Waves Make Air Quality Worse." *IQAir*, https://www.iqair.com/us/newsroom/how-heat-waves-make-air-quality-worse. Accessed 3 January 2025.
23. Hou, Pei, and Shiliang Wu. "Long-term Changes in Extreme Air Pollution Meteorology and the Implications for Air Quality." *Scientific Reports*, vol. 6, no. 1, 2016, https://doi.org/10.1038/srep23792.
24. Borunda, Alejandra. "Extreme Heat Makes Air Quality Worse–that's Bad for Health." *NPR*, 6 Sept. 2023, https://www.npr.org/2023/09/06/1198007014/extreme-heat-makes-air-quality-worse-thats-bad-for-health. Accessed 3 January 2025.
25. Sforza, Lauren. "16 States, DC Are under Air Quality Alerts from Canada's Wildfire Smoke." *The Hill*, 28 June 2023, https://thehill.com/homenews/state-watch/4072340-these-17-states-are-under-air-quality-alerts-from-canadas-wildfire-smoke/. Accessed 3 January 2025.
26. Nguyen, Giang Tran Huong, *et al*. "An Exploration of Meteorological Effects on PM2.5 Air Quality in Several Provinces and Cities in Vietnam." *Journal of Environmental Sciences*, vol. 145, Nov. 2024, pp. 139-51, https://doi.org/10.1016/j.jes.2023.07.020.
27. Kirešová, Simona, and Milan Guzan. "Determining the Correlation between Particulate Matter PM10 and Meteorological Factors." *Eng*, vol. 3, no. 3, 2022, pp. 343–63. https://doi.org/10.3390/eng3030025.
28. Yang, Zhou, *et al*. "Nonlinear and Lagged Meteorological Effects on Daily Levels of Ambient PM2.5 and O3: Evidence from 284 Chinese Cities." *Journal of Cleaner Production*, vol. 278, Jan. 2021, p. 123931, https://doi.org/10.1016/j.jclepro.2020.123931.
29. United States Environmental Protection Agency (EPA), "Basics of Climate Change." https://www.epa.gov/climatechange-science/basics-climate-change. Accessed 3

https://doi.org/10.59720/25-084

January 2025.
30. Feng, Xu, *et al*. "Large Role of Anthropogenic Climate Change in Driving Smoke Exposure across the Western United States from 1992 to 2020." *ArXiv,* 4 Dec. 2024, https://doi.org/10.48550/arxiv.2412.03733.
31. Bu, Qiaoli, *et al*. "The Modulation of Meteorological Parameters on Surface PM2.5 and O3 Concentrations in Guangzhou, China." *Aerosol and Air Quality Research*, vol. 21, 2021, p. 200084. https://doi.org/10.4209/aaqr.2020.03.0084.
32. United States Environmental Protection Agency (EPA), "Sampling Methods for PM2.5 Speciation Parameters." https://aqs.epa.gov/aqsweb/documents/codetables/methods_speciation.html. Accessed 3 January 2025.

Appendix

Data is from EPA and NOAA (16-17). The following code merges the data and runs all the regressions relevant to this paper. Explanations on specifics of codes are commented alongside the code below.

```
#import necessary packages
import pandas as pd
import numpy as np
#combine air quality data
df = pd.read_csv("air_quality/Air Quality 2020.csv")
df = pd.concat([pd.read_csv(f"air_quality/Air Quality 202{i}.csv") for i in range(0, 5)])
df.to_csv('air_quality_data.csv')
#combine weather data
names = []
for i in range(0, 5):
    names.append(f'weather_data/202{i}-Jan-June.csv')
    names.append(f'weather_data/202{i}-Jul-Dec.csv')
df = pd.concat([pd.read_csv(name) for name in names])
df.to_csv('weather_data.csv')
#filter for Reagan National Airport
def is_reagan(x):
    return 'WASHINGTON REAGAN NATIONAL AIRPORT, VA US' in x
airport_filter = df.loc[df['NAME'].apply(is_reagan)]
airport_filter.to_csv('weather_data_airport.csv')

#read and format data
wd = pd.read_csv('weather_data_airport.csv')
wd = wd.drop(wd.columns[[0]], axis = 1)

aqd = pd.read_csv('air_quality_data.csv')
aqd = aqd.drop(aqd.columns[[0]], axis = 1)

aqd.set_index(['Local Site Name'], inplace = True, drop = False)
wd.set_index(['NAME'], inplace = True, drop = False)

#set up connected locations
connected_sites = [('Aurora Hills Visitors Center', 'WASHINGTON REAGAN NATIONAL
AIRPORT, VA US')]

#format date into datetime objects
import datetime
from datetime import date, timedelta

def daterange(start_date: date, end_date: date):
    days = int((end_date - start_date).days)
    for n in range(days):
        yield start_date + timedelta(n)
```

```
start_date = date(2020, 1, 1)
end_date = date(2024, 12, 31)
for single_date in daterange(start_date, end_date):
    print(single_date.strftime("%Y-%m-%d"))
    break

wd['DATE'] = wd['DATE'].map(lambda x : datetime.datetime.strptime(x, '%m/%d/%Y').date())
aqd['Date'] = aqd['Date'].map(lambda x : datetime.datetime.strptime(x, '%m/%d/%Y').date())
aqd.loc[aqd['Date'] == date(2020, 1, 1)]
aqd.loc[(aqd['Date'] == single_date) & (aqd['Local Site Name'] == 'MathScience Innovation
Center')].iloc[0]

#create final output with air quality and weather matched by date
output = pd.DataFrame({"Air Quality Site Name":[],
                    "Weather Data Site Name":[],
                    "Air Quality Site Latitude":[],
                    "Air Quality Site Longitude":[],
                    "Weather Site Latitude":[],
                    "Weather Site Longitude":[],
                    "Date":[],
                    "POC":[],
                    "Daily Mean 2.5 Concentration":[],
                    "AQI":[],
                    "Wind Speed":[],
                    "Precipitation":[],
                    "Average Temperature":[],
                    "Minumum Temperature":[],
                    "Maximum Temperature":[]
                    })
for cs in connected_sites:
    print(cs)
    aqd_site = cs[0]
    weather_site = cs[1]
    for single_date in daterange(start_date, end_date):
        added_list = [
            aqd_site,
            weather_site,
            aqd.loc[aqd_site]['Site Latitude'].iloc[0],
            aqd.loc[aqd_site]['Site Longitude'].iloc[0],
            wd.loc[weather_site]['LATITUDE'].iloc[0],
            wd.loc[weather_site]['LONGITUDE'].iloc[0],
            single_date.strftime('%m-%d-%Y'),
        ]
        aqd_found = False
        filtered = aqd.loc[(aqd['Date'] == single_date) & (aqd['Local Site Name'] == aqd_site)]
        if len(filtered) > 0:
            aqd_found = True
            filtered = filtered.iloc[0]
            added_list.append(filtered['POC'])
            added_list.append(filtered['Daily Mean PM2.5 Concentration'])
            added_list.append(filtered['Daily AQI Value'])
```

```
        else:
            for i in range(3): added_list.append('')
        wd_found = False
        filtered = wd.loc[(wd['DATE'] == single_date) & (wd['NAME'] == weather_site)]
        if len(filtered) > 0:
            wd_found = True
            filtered = filtered.iloc[0]
            added_list.append(filtered['AWND'])
            added_list.append(filtered['PRCP'])
            added_list.append(filtered['TAVG'])
            added_list.append(filtered['TMIN'])
            added_list.append(filtered['TMAX'])
        else:
            for i in range(5): added_list.append('')
        if aqd_found or wd_found:
            output.loc[len(output)] = added_list
output.to_csv("combined_air_quality_weather.csv")

datapts = pd.read_csv("combined_air_quality_weather.csv")
#create new variables with log of variables
datapts["log-avg-temp"]=np.log(datapts["Average Temperature"])
datapts["log-min-temp"]=np.log(datapts["Minumum Temperature"])
datapts["log-max-temp"]=np.log(datapts["Maximum Temperature"])
datapts["log-wind"]=np.log(datapts["Wind Speed"])
datapts["log-precip"]=np.log(datapts["Precipitation"])
datapts["log-pm"]=np.log(datapts["Daily Mean 2.5 Concentration"])
datapts["gen-avg-temp"]=(datapts['Minumum Temperature'] + datapts['Maximum Temperature'])
/ 2
datapts["gen-log"]=np.log(datapts["gen-avg-temp"])
datapts["log-aqi"]=np.log(datapts["AQI"])

datapts.to_csv("WashingtonReagan-AuroraHills-withLogs.csv")

#read data file
df = pd.read_csv("WashingtonReagan-AuroraHills-withLogs.csv")
df = df.drop(df.columns[[0, 1]], axis=1)

#create seasonal dummies
def is_winter(date):
    date = datetime.datetime.strptime(date, '%m-%d-%Y').date()
    return date.month in [12, 1, 2]
def is_spring(date):
    date = datetime.datetime.strptime(date, '%m-%d-%Y').date()
    return date.month in [3, 4, 5]
def is_summer(date):
    date = datetime.datetime.strptime(date, '%m-%d-%Y').date()
    return date.month in [6, 7, 8]
def is_fall(date):
    date = datetime.datetime.strptime(date, '%m-%d-%Y').date()
    return date.month in [9, 10, 11]
```

```
#create columns for each seasonal dummy
df['is_winter'] = df['Date'].map(is_winter)
df['is_spring'] = df['Date'].map(is_spring)
df['is_summer'] = df['Date'].map(is_summer)
df['is_fall'] = df['Date'].map(is_fall)

df.to_csv('WashingtonReagan-AuroraHills-Log-dummies.csv')

#import packages
import matplotlib.pyplot as plt
import statsmodels.api as sm
datapts = pd.read_csv('WashingtonReagan-AuroraHills-Log-dummies.csv')

#drop data for days without rain
without_zero = datapts.replace(0, np.nan)
without_zero = without_zero.dropna(how='any', axis=0)
seasons = ['winter', 'spring', 'summer', 'fall']
for s in seasons:
    without_zero[f'is_{s}'] = without_zero[f'is_{s}'].map(int)

#create variables in new dataframe
without_zero["gen-log-winter"]=(without_zero['gen-log'] * without_zero["is_winter"])
without_zero["gen-log-spring"]=(without_zero['gen-log'] * without_zero["is_spring"])
without_zero["gen-log-summer"]=(without_zero['gen-log'] * without_zero["is_summer"])
without_zero["gen-log-fall"]=(without_zero['gen-log'] * without_zero["is_fall"])
without_zero["gen-prec"]=(without_zero["Precipitation"]+1)
without_zero["log-gen-prec"]=np.log(without_zero["gen-prec"])

#run all necessary regressions
exog_names = ['log-gen-prec', "log-wind", "gen-log", "is_winter", "is_spring", "is_summer",
"is_fall", "gen-log-winter", "gen-log-spring","gen-log-summer","gen-log-fall"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-pm'], exog)
results = model.fit()
results.summary()

df = pd.read_csv('allVarsPM.csv')
df.columns = df.columns.str.replace(' ','')
df.to_csv('allVarsPM.csv')

exog_names = ['log-gen-prec', "log-wind", "gen-log", "is_winter", "is_spring", "is_summer",
"is_fall", "gen-log-winter", "gen-log-spring","gen-log-summer","gen-log-fall"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-aqi'], exog)
results_all_aqi = model.fit()
results_all_aqi.summary()

top_all_aqi = results_all_aqi.summary().tables[0]
LRresult_all_aqi = (results_all_aqi.summary().tables[1])
with open('allVarsAQI.csv', 'w') as f:
    f.write(top_all_aqi.as_csv() + '\n' + LRresult_all_aqi.as_csv())
```

```
df = pd.read_csv('allVarsAQI.csv')
df.columns = df.columns.str.replace(' ','')
df.to_csv('allVarsAQI.csv')

exog_names = ['log-gen-prec', "log-wind", "gen-log", "is_winter", "is_spring", "is_summer",
"is_fall"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-pm'], exog)
results_plainseason_pm = model.fit()
results_plainseason_pm.summary()

top_plainseason_pm = results_plainseason_pm.summary().tables[0]
LRresult_plainseason_pm = (results_plainseason_pm.summary().tables[1])
with open('plainSeasonPM.csv', 'w') as f:
   f.write(top_plainseason_pm.as_csv() + '\n' + LRresult_plainseason_pm.as_csv())
df_plainseason_pm = pd.read_csv('plainSeasonPM.csv')
df_plainseason_pm.columns = df_plainseason_pm.columns.str.replace(' ','')
df_plainseason_pm.to_csv('plainSeasonPM.csv')

exog_names = ['log-gen-prec', "log-wind", "gen-log", "is_winter", "is_spring", "is_summer",
"is_fall"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-aqi'], exog)
results_plainseason_aqi = model.fit()
results_plainseason_aqi.summary()

top_plainseason_aqi = results_plainseason_aqi.summary().tables[0]
LRresult_plainseason_aqi = (results_plainseason_aqi.summary().tables[1])
with open('plainSeasonAQI.csv', 'w') as f:
   f.write(top_plainseason_aqi.as_csv() + '\n' + LRresult_plainseason_aqi.as_csv())
df_plainseason_aqi = pd.read_csv('plainSeasonPM.csv')
df_plainseason_aqi.columns = df_plainseason_aqi.columns.str.replace(' ','')
df_plainseason_aqi.to_csv('plainSeasonAQI.csv')

exog_names = ['log-gen-prec', "log-wind", "gen-log"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-pm'], exog)
results_triple_PM = model.fit()
results_triple_PM.summary()

top_triple_PM = results_triple_PM.summary().tables[0]
LRresult_triple_PM = (results_triple_PM.summary().tables[1])
with open('triplePM.csv', 'w') as f:
   f.write(top_triple_PM.as_csv() + '\n' + LRresult_triple_PM.as_csv())
df_triple_PM = pd.read_csv('triplePM.csv')
df_triple_PM.columns = df_triple_PM.columns.str.replace(' ','')
df_triple_PM.to_csv('triplePM.csv')

exog_names = ['log-gen-prec', "log-wind", "gen-log"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-aqi'], exog)
```

```
results_triple_AQI = model.fit()
results_triple_AQI.summary()

top_triple_AQI = results_triple_AQI.summary().tables[0]
LRresult_triple_AQI = (results_triple_AQI.summary().tables[1])
with open('tripleAQI.csv', 'w') as f:
   f.write(top_triple_AQI.as_csv() + '\n' + LRresult_triple_AQI.as_csv())
df_triple_AQI = pd.read_csv('tripleAQI.csv')
df_triple_AQI.columns = df_triple_AQI.columns.str.replace(' ','')
df_triple_AQI.to_csv('tripleAQI.csv')

exog_names = ['log-gen-prec']
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-pm'], exog)
results_prec = model.fit()
results_prec.summary()

top_prec = results_prec.summary().tables[0]
LRresult_prec = (results_prec.summary().tables[1])
with open('precPM.csv', 'w') as f:
   f.write(top_prec.as_csv() + '\n' + LRresult_prec.as_csv())
df_prec = pd.read_csv('precPM.csv')
df_prec.columns = df_prec.columns.str.replace(' ','')
df_prec.to_csv('precPM.csv')

exog_names = ['log-gen-prec']
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-aqi'], exog)
results_prec_aqi = model.fit()
results_prec_aqi.summary()

top_prec_aqi = results_prec_aqi.summary().tables[0]
LRresult_prec_aqi = (results_prec_aqi.summary().tables[1])
with open('precAQI.csv', 'w') as f:
   f.write(top_prec_aqi.as_csv() + '\n' + LRresult_prec_aqi.as_csv())
df_prec_aqi = pd.read_csv('precAQI.csv')
df_prec_aqi.columns = df_prec_aqi.columns.str.replace(' ','')
df_prec_aqi.to_csv('precAQI.csv')

exog_names = ["log-wind"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-pm'], exog)
results_wind_pm = model.fit()
results_wind_pm.summary()

top_wind_pm = results_wind_pm.summary().tables[0]
LRresult_wind_pm = (results_wind_pm.summary().tables[1])
with open('windPM.csv', 'w') as f:
   f.write(top_wind_pm.as_csv() + '\n' + LRresult_wind_pm.as_csv())
df_wind_pm = pd.read_csv('windPM.csv')
df_wind_pm.columns = df_wind_pm.columns.str.replace(' ','')
```

```
df_wind_pm.to_csv('windPM.csv')

exog_names = ["log-wind"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-aqi'], exog)
results_wind_aqi = model.fit()
results_wind_aqi.summary()

top_wind_aqi = results_wind_aqi.summary().tables[0]
LRresult_wind_aqi = (results_wind_aqi.summary().tables[1])
with open('windAQI.csv', 'w') as f:
   f.write(top_wind_aqi.as_csv() + '\n' + LRresult_wind_aqi.as_csv())
df_wind_aqi = pd.read_csv('windAQI.csv')
df_wind_aqi.columns = df_wind_aqi.columns.str.replace(' ','')
df_wind_aqi.to_csv('windAQI.csv')

exog_names = ["gen-log"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-pm'], exog)
results_gen_pm = model.fit()
results_gen_pm.summary()

top_gen_pm = results_gen_pm.summary().tables[0]
LRresult_gen_pm = (results_gen_pm.summary().tables[1])
with open('tempPM.csv', 'w') as f:
   f.write(top_gen_pm.as_csv() + '\n' + LRresult_gen_pm.as_csv())
df_gen_pm = pd.read_csv('tempPM.csv')
df_gen_pm.columns = df_gen_pm.columns.str.replace(' ','')
df_gen_pm.to_csv('tempPM.csv')

exog_names = ["gen-log"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-aqi'], exog)
results_gen_aqi = model.fit()
results_gen_aqi.summary()

top_gen_aqi = results_gen_aqi.summary().tables[0]
LRresult_gen_aqi = (results_gen_aqi.summary().tables[1])
with open('tempAQI.csv', 'w') as f:
   f.write(top_gen_aqi.as_csv() + '\n' + LRresult_gen_aqi.as_csv())
df_gen_aqi = pd.read_csv('tempAQI.csv')
df_gen_aqi.columns = df_gen_aqi.columns.str.replace(' ','')
df_gen_aqi.to_csv('tempAQI.csv')

exog_names = ['log-gen-prec', "log-wind", "gen-log", "gen-log-winter", "gen-log-spring","gen-log-
summer","gen-log-fall"]
exog = sm.add_constant(without_zero[exog_names])
model = sm.OLS(without_zero['log-pm'], exog)
results_all_pm = model.fit()
results_all_pm.summary()
```