The Effect of Air Pollution on COVID-19 Cases

Nathan Dai

Abstract

Air pollution’s growing problems have plagued the world since Europe’s Industrial Revolution. This study evaluated the association between air pollution, weather, and the intensity of COVID-19. During the time frame between April 1, 2020 and May 31, 2020, daily new cases, air pollution data, and meteorological data were obtained. A multiple linear regression was used to determine whether air pollution (PM2.5, PM10, CO, NO2, O3, and SO2) and weather (mean temperature, humidity, air pressure, and wind speed) could predict COVID-19 cases. The individual predictors were examined, and it was determined that PM2.5,O3, humidity, and pressure were significant predictors of COVID-19 new cases.

Keywords: COVID-19, air pollution, multiple linear regression

# Introduction

The COVID-19 pandemic took the world by surprise; its sudden spread and massive scope caught many off-guard. The novel coronavirus first appeared in Wuhan, China, then spread rapidly throughout the world. Although a similar virus reached multiple countries, its effect varied from region to region. In addition to the various responses from the government, the differences in the COVID-19 cases may be attributed to the variations in air quality.

A further understanding of air pollution personally concerns me because I have seen the effects of air pollution. When I went to Beijing, the thick layer of pollution in the air gave everything a dark and gloomy feel. Only the lights were visible on distant buildings, and a faint odor constantly lingered in the air. Air pollution severely harms human health. It causes a variety of risks such as increased infant mortality, increased prevalence of respiratory symptoms, and decreased lung function (Bates, 1995). Every time I see the yellowish gray cloud and hear my breath on my masked face, I think about how detrimental air pollution is to health. This leads to the question: How does air pollution affect the intensity of COVID-19? In my research, I investigated the effects of air pollution on the number of COVID-19 cases.

# Methods

## Study Area

This study included 88 counties across the United States. These 88 counties accounted for 77% of all confirmed cases as of April 1, 2020. For each county, air pollution, meteorological, and COVID-19 confirmed cases data was gathered. Under the air pollution category, PM2.5, PM10, CO, NO2, O3, and SO2 data was collected. For the meteorological data, 4 climatic measurements were taken: mean temperature, relative humidity, air pressure, and wind speed.

## Data Collection

The meteorological data was collected from the World Weather Online API.

## Statistical methods

I performed a multivariate regression to study the dependence of new COVID-19 cases on PM2.5, O3, SO2, NO2, PM10, CO, temperature, humidity, wind speed, pressure, UV index and the existing confirmed COVID-19 cases.

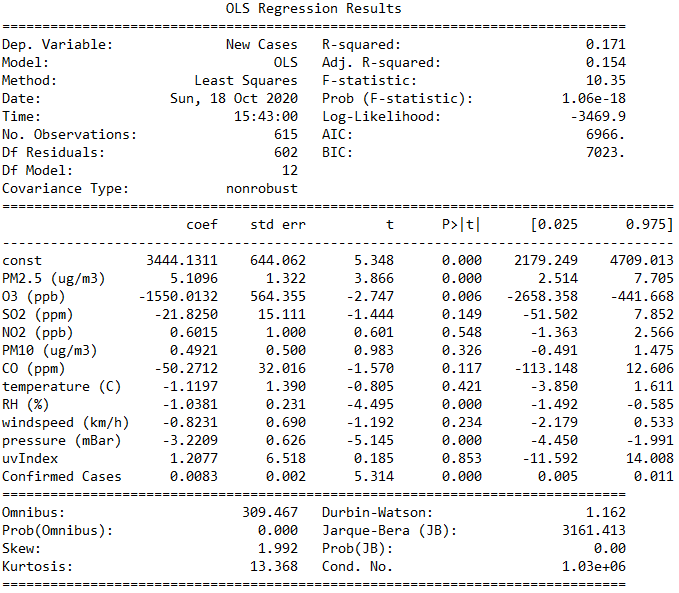
We expected that many of the pollutants, meteorological data and the existing confirmed COVID-19 cases would impact the number of new cases, since they would all influence how fast the virus spread. However, as shown in Table 1, there are strong correlations among the pollutants, such as between CO and NO2 (*r*  = 0.5344, *p* =1.2517e-207 ) or between PM10 and O3 (*r* = 0.5526, *p* = 1.5683e-49). Similarly, there are strong correlations between several meteorological parameters, such as between temperature and UV index (*r* = 0.9437, *p* = 0). Therefore, we much be careful when we select independent variables to avoid strong correlations among them.

Table : Correlation matrix of independent variables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | PM2.5 (ug/m3) | O3 (ppb) | SO2 (ppm) | NO2 (ppb) | PM10 (ug/m3) | CO (ppm) |
| PM2.5 (ug/m3) | 1 |  |  |  |  |  |
| O3 (ppb) | -0.0124 | 1 |  |  |  |  |
| SO2 (ppm) | 0.2119 | -0.0317 | 1 |  |  |  |
| NO2 (ppb) | 0.3686 | -0.1171 | 0.3857 | 1 |  |  |
| PM10 (ug/m3) | 0.4883 | 0.5526 | 0.2245 | 0.0615 | 1 |  |
| CO (ppm) | 0.3844 | -0.1564 | 0.196 | 0.5344 | 0.0702 | 1 |
| temperature (C) | 0.2639 | 0.036 | 0.2002 | 0.1706 | 0.2661 | 0.4395 |
| RH (%) | 0.043 | -0.354 | -0.1236 | -0.1684 | -0.1714 | -0.0094 |
| windspeed (km/h) | -0.1621 | 0.1446 | -0.0657 | -0.2652 | -0.2421 | -0.201 |
| pressure (mBar) | 0.0644 | -0.1136 | -0.0127 | 0.13 | -0.0319 | 0.0617 |
| uvIndex | 0.2078 | 0.0987 | 0.1645 | 0.1503 | 0.3298 | 0.412 |
| Confirmed Cases | -0.0193 | 0.0345 | 0.1327 | 0.0594 | 0.3326 | 0.0469 |
|  |  |  |  |  |  |  |
|  | temp (C) | RH (%) | windspeed (km/h) | pressure (mBar) | uvIndex | Confirmed Cases |
| PM2.5 (ug/m3) |  |  |  |  |  |  |
| O3 (ppb) |  |  |  |  |  |  |
| SO2 (ppm) |  |  |  |  |  |  |
| NO2 (ppb) |  |  |  |  |  |  |
| PM10 (ug/m3) |  |  |  |  |  |  |
| CO (ppm) |  |  |  |  |  |  |
| temp (C) | 1 |  |  |  |  |  |
| RH (%) | -0.1479 | 1 |  |  |  |  |
| windspeed (km/h) | -0.2092 | 0.0459 | 1 |  |  |  |
| pressure (mBar) | 0.0718 | -0.0951 | -0.3154 | 1 |  |  |
| uvIndex | 0.9437 | -0.2387 | -0.1897 | 0.128 | 1 |  |
| Confirmed Cases | 0.0182 | -0.0226 | 0.116 | 0.0737 | 0.0484 | 1 |

Table 2 shows the summary of linear regression between new COVID-19 cases and all the pollutants, meteorological data and existing confirmed COVID-19 cases. The following predictors were found to be significant: PM2.5, relative humidity, pressure, and the existing confirmed COVID-19 cases. However, this does not necessarily indicate that these are the only good predictors because of the correlation between these independent variables and other independent variables as we will show later. The model explained 15.4% of the variance in the predictors. As we expected, the result is poor. The p-values of several independent variables are too large.

Table 2 Summary of linear regression between new COVID-19 cases and multiple pollutants, meteorological data and existing confirmed COVID-19 cases



After careful selection, we have only kept PM2.5, SO2, temperature, relative humidity, pressure, existing confirmed cases as independent variables, since they generate the most satisfactory result of regression, and there is little correlation among these predictors. The summary of the linear regression is shown in Table 3. All p-values are statistically significant. The adjusted R-squared has improved to 0.600. The partial residual plots for each independent variable are plotted in Figure 1 through Figure 6 to help visualize the result of the regression.

A picture containing diagram

Description automatically generated

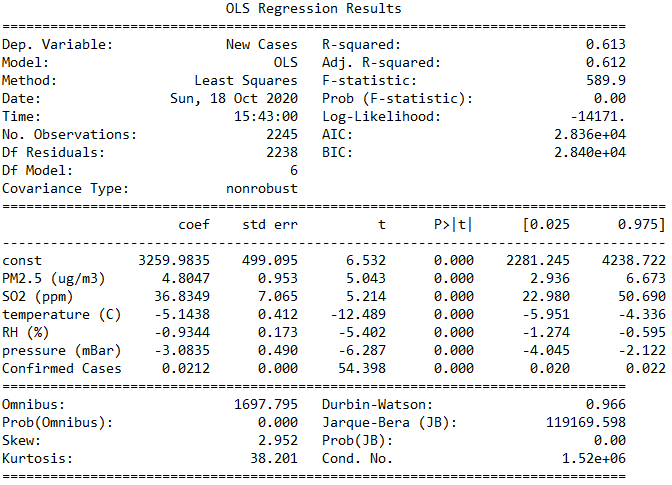
Figure Residual plots for PM2.5, NO2, RH, temperature, pressure and confirmed COVID-19 cases

It is interesting to observe from Table 3 the impact of the independent variables on the new cases of COVID-19. For every 1 ug/m3 increase in PM2.5, the daily new COVID-19 cases increase by 4.8047. For every 1 part per million (ppm) increase in SO2, the daily new COVID-19 cases increase by 36.8349. Although there might be other factors (government policies, demographic, etc.) that are not considered in our study, it can be reasonably concluded from our analysis that pollution, among other factors, accelerates the spread of the virus. The mean value of PM2.5 in our collected data in the US is 6.7 ug/m3, and the mean value of SO2 in our collected data in the US is 0.4542 ppm. In many metropolitan areas around the world, the pollution level can be as high as 30x of our average level. Therefore, clear air is important to control the spread of COVID-19 and other similar respiratory diseases.

Also, from Table 3 we can see that for every 1% increase in humidity, the daily new COVID-19 cases decrease by 0.9344. For every 1C increase in temperature, the daily new COVID-19 cases decrease by 5.1438. For every 1 millibar increase in air pressure, the daily new COVID-19 cases decrease by 3.0835. Therefore, cold temperature, low humidity and low pressure accelerate the spread of COVID-19. This is also consistent with the seasonality of COVID-19.

Finally, Table 3 also shows that for every person increase in the existing confirmed case, the daily new cases increase by 0.0212. Equivalently, for every 47 existing cases, there is 1 new case. If this rate of spread persists, the number of confirmed COVID-19 cases would grow at the rate of , where *C0* is the initial number of cases, and *d* is the number of days. It can be inferred that expected to double every 33 days.

Table 3 Summary of linear regression between new COVID-19 cases and selected pollutants, meteorological data and existing confirmed COVID-19 cases



# Conclusion

We have collected multiple pollutants and meteorological data from numerous US counties where COVID-19 are widely spread and studied the correlation between these data and the daily new COVID-19 cases. We have concluded that air pollution can accelerate the spread of COVID-19, and weather also has various impact on the disease. We also expect the total number of cases to double every 33 days unless proper control is put in place to contain the spread.