The Effects of COVID-19 Lockdown on Air Pollution in LA County

SANDRA VILLAMAR

1 ABSTRACT

For our project we decided to look at two important problems in the world right now: the COVID-19 pandemic and air pollution. We specifically focused on LA County, to figure out whether the COVID-19 lockdown regulations had contributed to an improvement in air quality. We started by visualizing and comparing the AQI levels of the pollutants pm2.5, O3, NO2, and CO, from the years 2018, 2019, and 2020. We found that three out of the four pollutants had a significant drop in 2020. We also analyzed weather trends in order to see if there was a major fluctuation which could have contributed to the significant drop in pollution in 2020. All in all, after considering possible confounding variables and performing multiple analyses to compare the pollution level between different years, we came to the conclusion that the regulations imposed by the COVID-19 lockdown is most likely the reason for the improvement in air quality in LA County.

2 INTRODUCTION

We are interested in answering the following questions: Has the air quality—in particular, the level of CO, NO2, O3, or pm2.5—changed in Los Angeles County since the COVID-19 lockdown regulations were in place? How did it change? Can we conclude that the COVID-19 lockdown is a reasonable guess for this change? Or are there other factors such as weather that may have contributed to a change in the air pollution?

We hypothesize that all four pollutants have significantly reduced in level in Los Angeles County due to the COVID-19 lockdown. We believe this because activities such as transportation and manufacturing have notably decreased due to the restrictions of the lockdown. These are all known sources that contribute to air pollution.

3 BACKGROUND

We decided to choose a question related to COVID-19 because we felt that this pandemic is by far the most prevalent factor that has significantly changed our daily lives in the past months. In addition, we wanted to incorporate our other interest with environment-related topics. Therefore, we discussed in what ways the COVID-19 lockdown has affected our normal activities and noticed that the lockdown could have an impact on the level of air pollution. We wanted to see if there is a change in air pollution and if so, how significant this change is. We specifically chose Los Angeles County because we wanted to choose an area in California that is known to have poor air quality as well as an area with a big population, since our findings could then be relevant to a greater audience.

This question is important to ask because improving air quality has proven to be a prominent topic that has consistently shown up in nationwide as well as worldwide discussions. In the United States alone, "the cost of poor air quality to the U.S. from air pollution related illness has been estimated at $150 billion per year [1]." This is because poor air quality leads to several respiratory and cardiovascular health problems, affecting individuals as well as entire communities. Moreover, sensitive ecosystems can and have been damaged as well. That being said, if the lockdown has indeed contributed to an improvement in air quality, then LA County (as well as other counties) could continue to restrict certain activities that are possible in a remote setting. On the other hand, if the lockdown has not contributed to a significant change in air pollution, then we will know to focus on other possible sources in search for finding a solution for better air quality.

While searching for our topic, we came across a data science project that studied the change in air pollution in Lombardy, Italy, comparing current COVID-19 lockdown dates to previous years [2]. The data scientists explored air quality data from NO2, Benzene, SO2, pm2.5, and pm10, across the years 2018, 2019, and 2020. They discovered a significant drop in some pollutants and no change in others. They also looked at other natural factors that are known to affect air quality such as wind, temperature, and precipitation, to see how they changed in relation to the decrease of some pollutant levels. They considered these weather factors because sunshine and heat cause pollutants to undergo chemical reactions faster, contributing to the development of smog. On the other hand, rain and wind help wash away pollutants or spread them out from a particular area, which typically results in improved air quality [3]. Nonetheless, after this analysis, they concluded that the intensity of human activity is the only factor that significantly changed throughout the years and is hence a reasonable guess to why the air quality improved in Lombardy, Italy.

Another study posted to the World Economic Forum found that some of the world's most polluted cities have seen a large improvement in air quality [4]. In order to show this, the scientists plotted levels of various pollutants in India, China, and South Korea. They also displayed satellite images of these locations taken before and during the COVID-19 lockdown. Both of these visualizations demonstrated how drastic the change in air quality was in these regions, after just a few weeks of lockdown regulations.

4 DATA SETS

4.1 Description of Data

We used two data sets to conduct our analysis. The first data set [5] contains information on the weather in downtown Los Angeles. It contains the station ID and name, the date the data was collected, and the average wind speed, precipitation, and maximum and minimum temperature for that day. The station and name are nominal data, date is ordinal data, and the rest of the data are numerical. We will be using this data set to see if there were any significant weather changes that could correlate to a possible change in pollutant levels during the COVID-19 lockdown.

The second data set [6] contains information on the pollutant levels for each day in downtown Los Angeles. It contains the date the data was collected as well as the AQI (air quality index) level for the pollutants pm2.5, pm10, O3, NO2, SO2, and CO. Date is ordinal data and the pollutant levels are numerical data. We will use this

data set to find out if there is a relationship between air pollution and lockdown regulations.

4.2 Pre-Processing the Data

In order to prepare our data sets for the analysis, we take a few steps to clean and organize them. First, we rename and drop unnecessary columns. Then, we fix the format of the date column and filtered the time frame of the data to January 2018 to May 2020. Additionally, we check for missing entries and notice there are only a handful of missing entries at random, so we decide that the missing values will not greatly perturb our results or contribute to any substantial biases. We replace these missing values by the integer zero (since no other entry in the data set has a zero entry), and make note of these missing values when we continue on to our visualizations and analyses.

5 DATA ANALYSIS AND VISUALIZATION

5.1 Pollutant Levels Over Time

We create line graphs in order to visualize the pollutant levels over the first phase lockdown time period (March 19th to May 8th). We use this time period for 2018, 2019, and 2020, so that we can clearly compare and see any possible differences in pollution between the years.

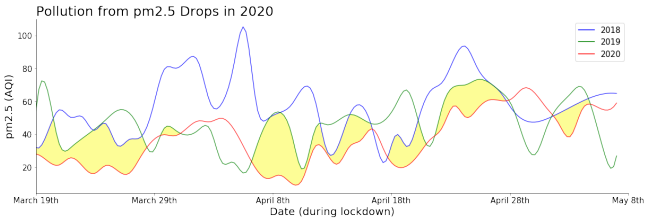

Fig. 1. Pollution from pm2.5 Drops in 2020

From this line graph, we can see that for most of the lockdown period, pollution from pm2.5 has dropped relative to the previous years, 2018 and 2019. The only major gap we see from this graph is right before April 8th. However, since we have three missing dates (April 5th, April 6th, and April 7th), the spline interpolation may not have accurately drawn the overall trend. Hence, the gap may be smaller in reality.

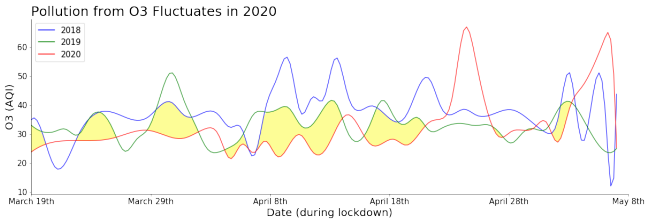

Fig. 2. Pollution from O3 Fluctuates in 2020

From this graph, it is not quite clear if the drop in pollution from O3 is significant enough to conclude that the lockdown regulations in 2020 had an impact. From the start of lockdown until about mid-April, most of the area is highlighted yellow, meaning that pollution from O3 had dropped relative to previous years. However, at the end we see two major spikes. One reason for this is that summer was nearing and O3 is known to have a positive correlation with temperature.

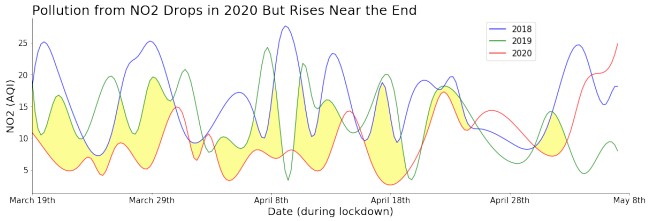

Fig. 3. Pollution from NO2 Drops in 2020 But Rises Near the End

NO2 emissions are primarily due to road traffic and other fossil fuel combustion processes. Because May 9th was the start of the California phase one lockdown plan, in which certain businesses and factories were allowed to re-open, this could explain why NO2 rose in the last portion of this graph. Other than that, the rest of the yellow highlighted area shows that pollution from NO2 was consistently below the previous years.

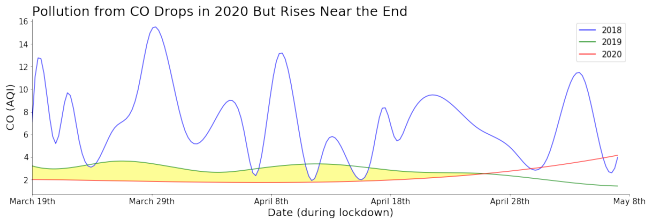

Fig. 4. Pollution from CO Drops in 2020 But Rises Near the End

We can see that for the majority of the lockdown period, CO levels in 2020 were below those in previous years. At the very end, we see something similar to NO2 in that the pollution levels start to rise. Because CO is also primarily emitted by vehicles, the start of the California phase one lockdown plan could explain why CO rose.

5.2 Correlation Between Pollutants and Weather

Because we know from prior research that weather factors affect pollution levels, we would like to see which of those factors have significantly high correlations. This way, we can check if during the 2020 lockdown time period, there was a significant change in those highly correlated weather factors. This is important because a change in weather may be the reason for a change in pollution levels, instead of the lockdown regulations themselves.

The coloring of the heatmap displays high positive correlations in dark red, and high negative correlations in dark blue. Hence, the stronger the correlation, the darker the color. From the heatmap, we can see that some pollutants have a notable correlation to certain weather factors. For example, pm2.5 does not seem to have a strong

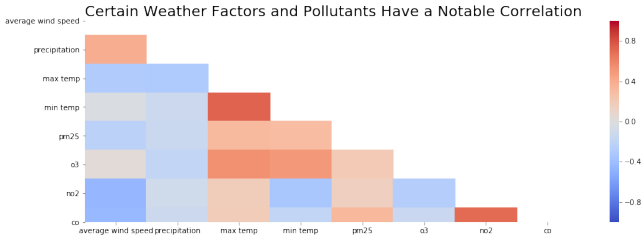

Fig. 5. Certain Weather Factors and Pollutants Have a Notable Correlation

correlation to any of the weather factors, as the highest correlation we see is about .31 with temperature. O3, on the other hand, has a decently strong (positive) correlation to temperature, measuring to about .55. Finally, NO2 and CO have a notable (negative) correlation to wind speed, measuring to about -.45.

5.3 Weather Fluctuation per Pollutant

Now that we have seen which weather factors are highly correlated to which pollutants, we can see if any of these weather factors have had major fluctuations during the 2020 lockdown. If there is indeed any major fluctuation, then this could have been a source for the change in pollution levels, instead of just the lockdown regulations.

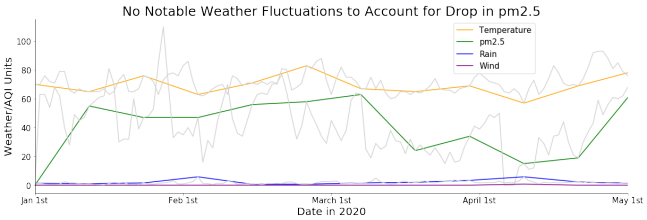

Fig. 6. No Notable Weather Fluctuations to Account for Drop in pm2.5

Looking at the green line, we can see that pm2.5 had a significant drop when lockdown started (March 19th). We also can see that none of the three weather lines show any major change throughout all of 2020. Hence, we can make a reasonable assumption that weather did not play a role in the drop in pollution from pm2.5.

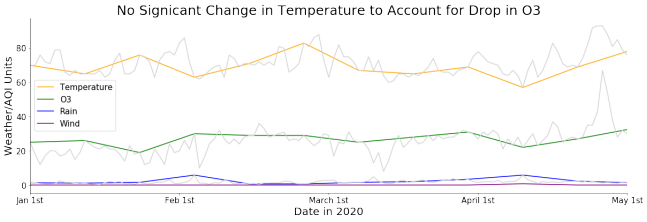

Fig. 7. No Significant Change in Temperature to Account for Drop in O3

From the correlation heatmap earlier, we learned that temperature has a notable positive correlation with O3. Hence, if temperature had gone down during the lockdown period, then this could have been a reason why O3 pollution levels had dropped. However, from the graph we see that the orange line (symbolizing temperature) stays consistent throughout entire 2020. Hence, we can make a reasonable assumption that weather did not play a role in the drop in pollution from O3.

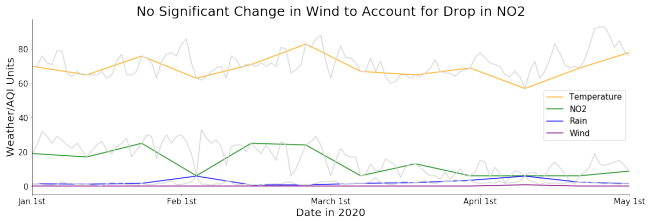

Fig. 8. No Significant Change in Wind to Account for Drop in NO2

From the correlation heatmap earlier, we learned that wind speed has a notable negative correlation with NO2. Hence, if wind speed had gone up during the lockdown period, then this could have been a reason why NO2 pollution levels had dropped. However, from the graph we see that the purple line (symbolizing wind speed) stays consistent throughout entire 2020. Hence, we can make a reasonable assumption that weather did not play a role in the drop in pollution from NO2.

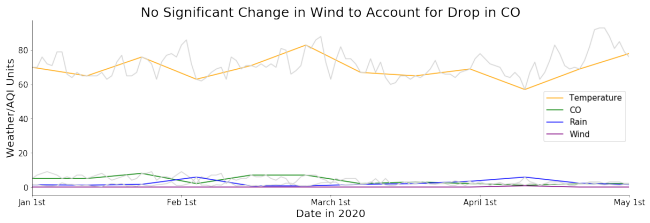

Fig. 9. No Significant Change in Wind to Account for Drop in NO2

Similar to NO2, we also learned that wind speed has a notable negative correlation with CO. We can see that the purple line representing wind speed does not fluctuate while the green line representing CO drops. Hence, we can make a reasonable assumption that weather did not play a role in the drop in pollution from CO.

After looking at all four graphs, we do not see any major fluctuations in temperature, wind, or rain, during 2020. We conclude that there must be some other factor (possibly the lockdown regulations) that caused the drops in pollution levels.

5.4 Permutation Tests

In these permutation tests, we are assessing whether the pollutant in question comes from two different distributions. We want to check if whether or not being in lockdown has affected the level of pollutants recorded. Therefore, we will create an extra boolean column in the data telling whether or not LA county was in lockdown during that date. We are using March 19th as the date that lockdown started. We will compare the data for when LA was in lockdown versus when LA was not. To do this, we will take the average AQI level of each group (lockdown and not lockdown) and use the difference of their

means as the test statistic. When doing this on the actual data, it will be called the observed test statistic.

Next, we will simulate data under the null hypothesis which is that being in lockdown has no effect on the AQI levels of the pollutants. To do this, we will shuffle the boolean lockdown column so that they label random data points. We will then calculate the difference in means of the two groups once again. We will repeat this process 1000 times to accumulate 1000 different simulated test statistics that were formed under the null hypothesis. Finally, we will calculate a p-value by counting the number of times the simulated test statistic was greater than or equal to the observed test statistic (difference in means). We will use a significance level of .05 to determine whether or not the results are statistically significant.

To help visualize this process, we will plot a histogram of the distribution of the test statistic when simulated under the null hypothesis. The red dot on the histogram will represent where the observed test statistic actually lies. We will then use a significance level of .05 to determine whether there is a significant difference between the two groups. The only assumption for the permutation test is that the observations are sampled independently from each other, which we already know is true.

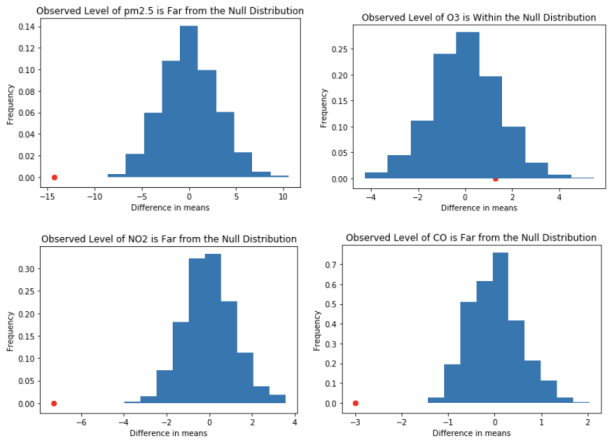

Fig. 10. Permutation Tests

After running the permutation tests on each of the four pollutants, we concluded that pm2.5, NO2, and CO, have significantly different pollution levels during lockdown versus not during lockdown. These results demonstrate that the lockdown regulations indeed had an impact on air pollution in LA County. This could be explained by the many businesses and companies that were forced to shut down, as enforced by the lockdown regulations. As a consequence of this shutdown as well as the discouragement of gatherings, there were less people driving on the road. Therefore, the primary sources that cause pollution had suddenly declined and most likely contributed to the drastic change in pollution levels.

6 ETHICS AND PRIVACY

The data sets we used are from the World Air Quality and the National Centers for Environmental Information project websites. Both host an open data platform allowing anyone to download a csv of air quality indexes and weather measurements of locations around the world. These websites give permission to anyone to utilize the data for projects or research. There are no privacy concerns regarding our data sets as the data are purely scientific, only including pollutant and weather measurements. There is no information on humans, their private data, or any other sensitive subjects. Therefore, there is no need to clean data in a way to preserve privacy.

Regarding the collection of the data, we do not believe there are any human biases as the values are obtained through systematic tests. One potential bias, though, is that the station where the weather data are collected is not the same station where the pollutant data are collected. In addition, stations in different parts of LA County may result in different measurements for both weather and pollutant levels. However, by looking up the exact addresses of both the weather and pollutant data stations, we can see that the distance between them is only 5.6 miles, which is minimal relative to the entire county. Both stations are also located in the central downtown area, which is the most populated area of LA County and hence, the most relevant part of the county to analyze. Therefore, we believe that the potential biases in the station locations will not drastically disrupt our analyses.

Regarding our analyses, there may be a potential bias in comparing the non-lockdown time period with the lockdown time period, as there are other factors such as governmental policies that could have affected the pollution levels in the non-lockdown time period. In this case, the difference between lockdown versus non-lockdown pollution levels would be exaggerated due to a factor other than lockdown regulations. Despite this potential bias, though, we believe that since we are comparing the lockdown time period to two years of non-lockdown time period, we have a large timeframe of comparison in which any unusual activity will most likely be minor relative to the overall pollution trend.

7 CONCLUSION

At the start of this project, we questioned whether the air quality in LA County had changed, and whether the COVID-19 lockdown regulations were a reasonable guess for this change. By visualizing the pollution levels in 2018, 2019, and 2020 through line graphs, we began to get an idea of the overall trend. Then, we studied the correlation between weather factors and the four pollutants, and graphed those factors against each pollutant to see if there were any major fluctuations in weather. We found that there were no notable fluctuations and concluded that there must be another reason for the possible change in pollution levels. Lastly, we ran a permutation test and found that pm2.5, NO2, and CO, all had statistically significant differences in their pollution levels from lockdown versus not in lockdown. Hence, we concluded that pollution from pm2.5, NO2, and CO, all dropped in 2020 and this drop was most likely due to the COVID-19 lockdown regulations. For O3, we did not see a statistically significant difference in pollution levels and hence came to to the conclusion that lockdown regulations most likely did not have an impact on O3 pollution.

As discussed in the Ethics and Privacy section above, some limitations of our analyses include that there may be more confounding

variables (other than weather) that we did not check for. For example, environmental policies or a change in industrial production may have caused differences in pollution levels as well.

All in all, the goal of our project was to see if the drastic measures used to stop the spread of COVID-19 also had the effect of lowering the level of air pollution in LA County. Our data and analyses showed that there was a significant difference in the pollution levels for three out of four pollutants during the time of the lockdown. This is a vital finding not only for LA County but also for other locations around the world. It proves that improving air quality and slowing down climate change is definitely possible. While our society will probably not be able to take as drastic measures as we did during lockdown, this study has identified the activities that make a difference in air pollution. Therefore, we can focus on looking at those activities to cut back or improve on, such as using electric-powered vehicles instead of gas-powered vehicles as well as minimizing the use of fossil fuels in factories.

8 REFERENCES

[1] https://www.weather.gov/sti/stimodeling_airquality_background
[2] https://towardsdatascience.com/is-covid-19-lockdown-cleaning-the-skies-over-milan-42dbba1ec812
[3] https://www.weather.gov/wrn/summer-article-clearing-the-air
[4] https://www.weforum.org/agenda/2020/04/coronavirus-covid19-air-pollution-enviroment-nature-lockdown
[5] https://www.ncdc.noaa.gov/cdo-web/confirmation
[6] https://aqicn.org/city/los-angeles/