Final Project Report

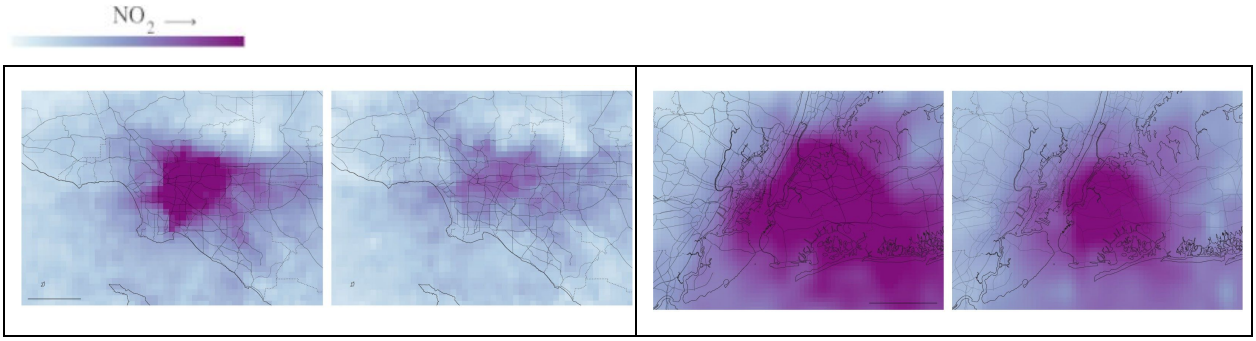
By Evan Coulson and Lauren West
Math189Z
15 May 2020

Research Question

Is expected air quality for the next year higher than models informed with pre COVID-19 data? In other words, will the new air quality numbers generate a prediction with better air quality for the future?

Motivation

Seeing clearer skies across the globe during the era of COVID-19, we decided it would be worth looking into the changes in pollution in the United States. The New York Times has reported "Traffic and Pollution Plummet," during shutdown, prompting us to quantify the implications of current trends (NY Times, 2020).



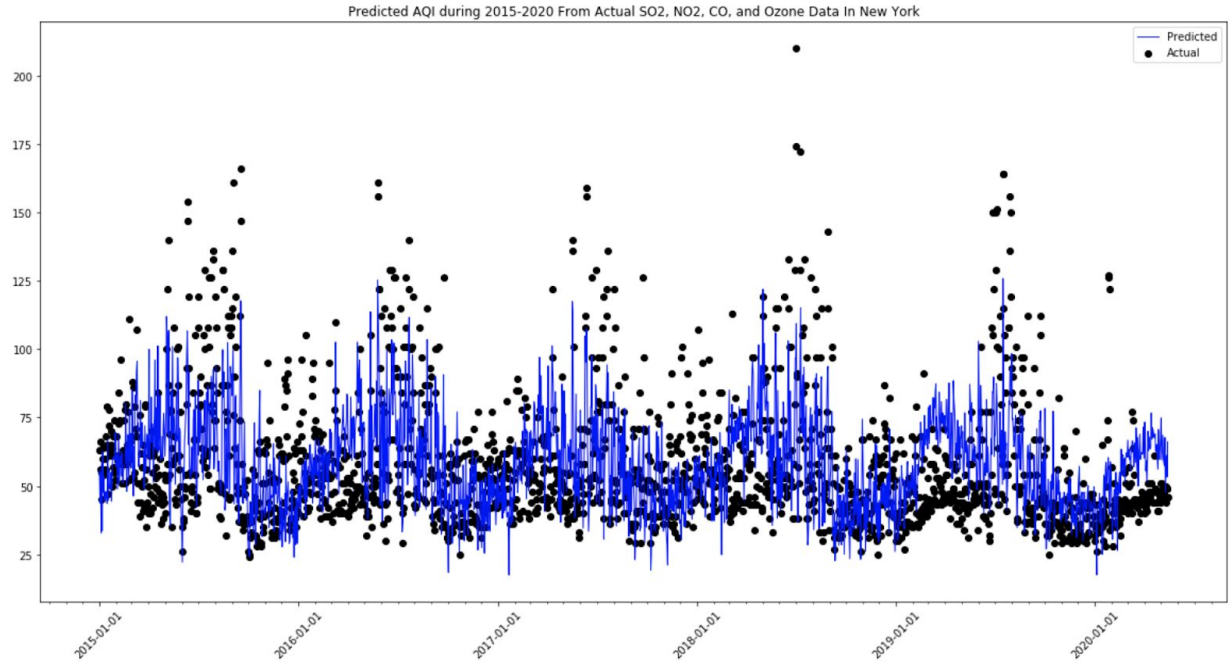
NY Times, March 2020. Before and After: nitrogen dioxide levels in Los Angeles (left) and New York (right).

The decreased pollution for the COVID-19 era could motivate positive changes to be maintained if future predictions hold that pollution levels should continue decreasing if it follows the current trends. We set out to discover whether future predictions for air quality seem to sustain positive changes in air quality. We also, out of intrigue, set out to discover what the current and the future pollution would look like if this COVID-19 era decrease never happened.

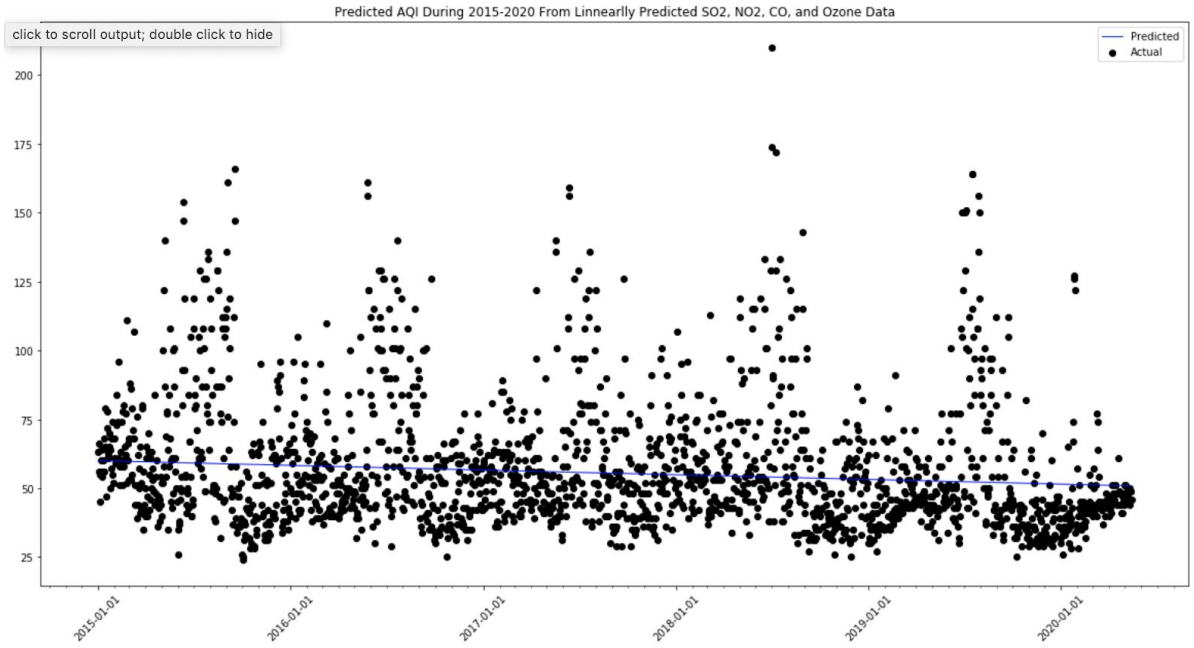
Prior Research

First, we created a linear regression on past data to predict the Air Quality Index, and our predictions mirrored the actual values rather well. So, we decided we could trust linear

regression to predict the Air Quality Index in our Final Project. These are our Preliminary Results:



However, when it came to running our individual pollutants in a linear regression, the results were far from reality due to the cyclical nature of pollutant levels from season to season. Still, trends were showing a decrease in air pollution, especially in the COVID-19 era, confirming our belief that this area of research was worth exploring further. However, we needed a different method than linear regression to predict our pollutants.



So, in order to make predictions for pollutant levels, we saw it fit to instead use a Recurrent Neural Network (RNN), as we will discuss in the next section.

Methods

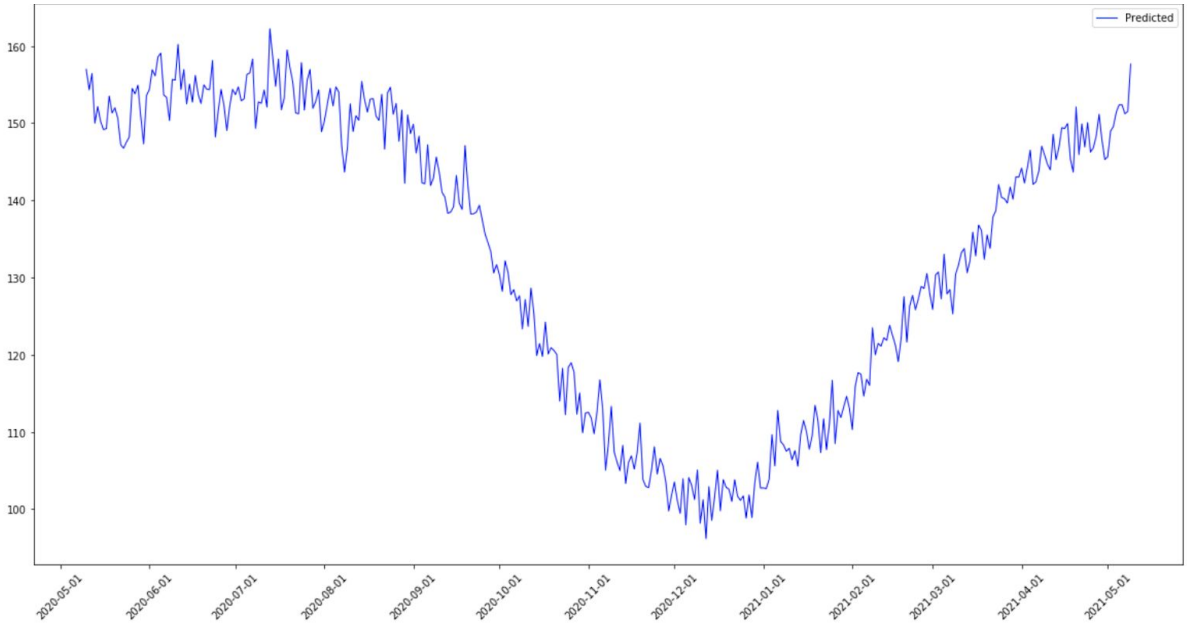
We decided to inspect two major metropolitans of the U.S.: Los Angeles and New York with our RNN predicting the future of the four pollutants that help determine AQI. The pollutants we tracked were nitrogen dioxide, sulfur dioxide, carbon monoxide, and ozone. We obtained this data from the United States Environmental Protection Agency. After reading and cleaning the data, we generated the training set. We did this with the AQI and the four different pollutants' data going back to the year 1980. We designed our RNNs to use long short term memory layers and dense nodes to attempt to predict the future concentrations of the particles. After training our RNNs, we were able to create a multi-step plot using our 40 years of training data to predict 1 year in the future. We will discuss our results in the next section.

Results

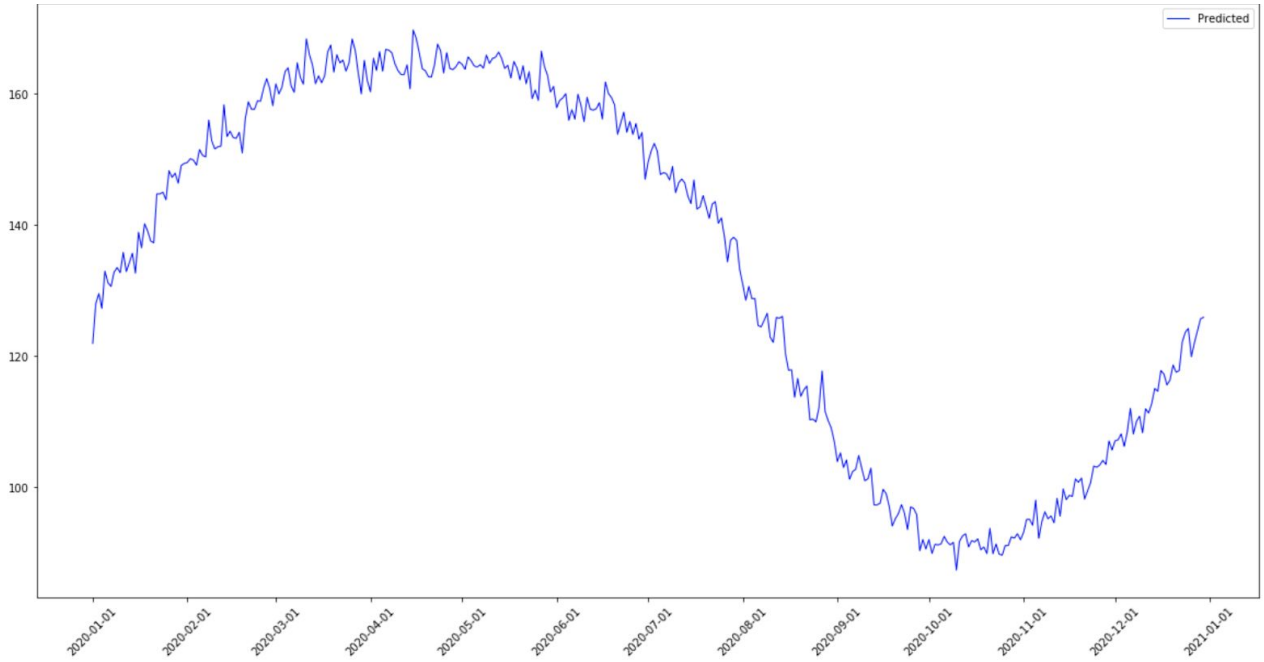
In this section, we define our models to be as after COVID-19 or before COVID-19. The distinction between these two definitions is that we consider after COVID-19 to be the most recent data we could obtain about air quality, which was May 10th 2020. The before COVID-19 era refers to data collected up until January 1st of 2020. Our neural networks then predicted the next year's worth of data, May 10th 2020 to May 9th 20201, and January 1st of 2020 to December 31st of 2020. The graphs of our AQI's display these dates, however, the pollutant graphs are not as obvious as they are in timesteps on the X-axis that represent how many days worth of data the model is predicting from. So, comparing the red components of the before and after graphs, it is important to keep in mind that the zero will be January 1st for before, and May 10th for after. Thus, the graphs show negative values representing the days since the date of prediction by the model. The graph then predicts 365 timesteps (days) into the future, which is the red trend displayed on the pollutant graphs. Also note that there will be no true future for the predicted pollutants as they are predictions.

Los Angeles

Using a linear regression, we were able to make future Air Quality Index Predictions excluding and including our COVID-19 era predictions.



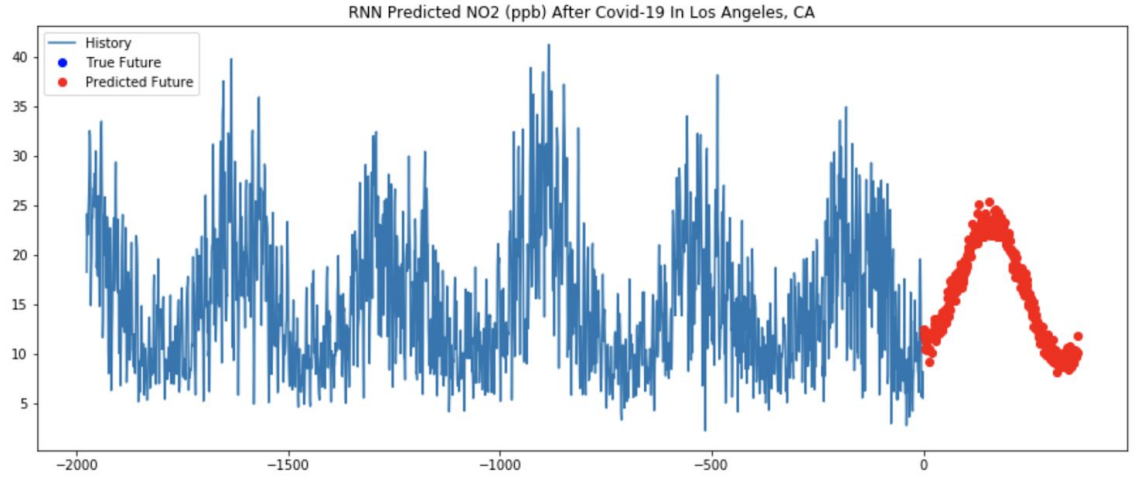
Predicted AQI from January 1st, 2020 to December 31st, 2020 (After-COVID-19 era data).

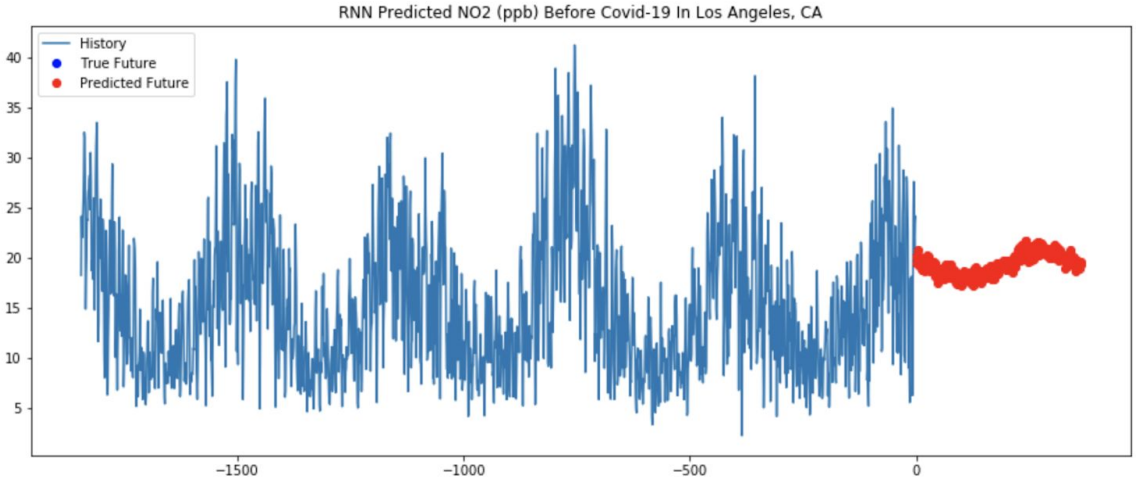


Predicted AQI from January 1st, 2020 to December 31st, 2020 (Before-COVID-19 era data).

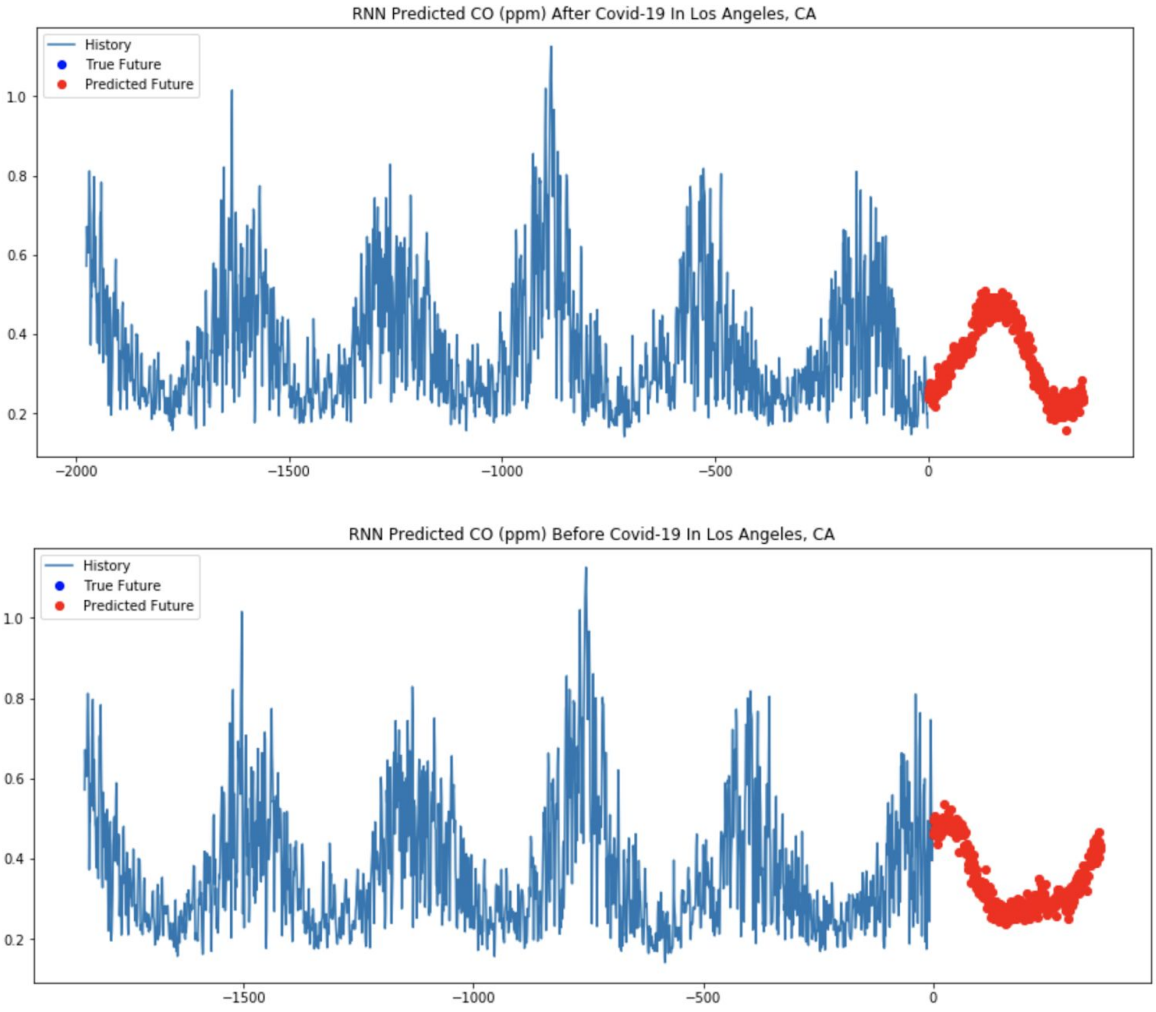
Unsurprisingly, our plot including COVID-19 data shows a slightly more environmentally favorable future in terms of AQI. The peak in the plot excluding COVID-19 data hit a higher peak AQI.

We used an RNN to predict future NO_2, CO, SO_2, and Ozone levels including and excluding COVID-19 era data. Here are our results below.

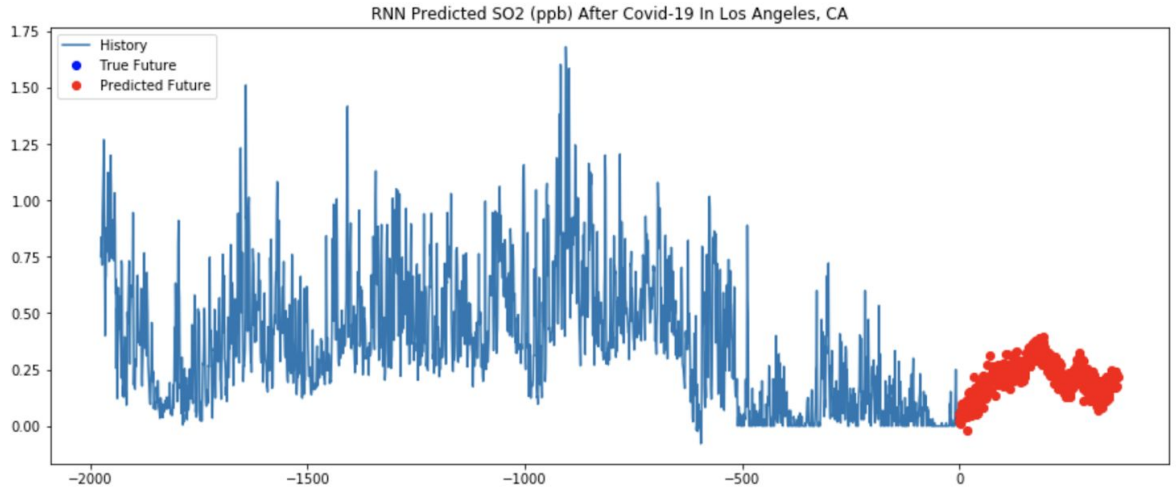


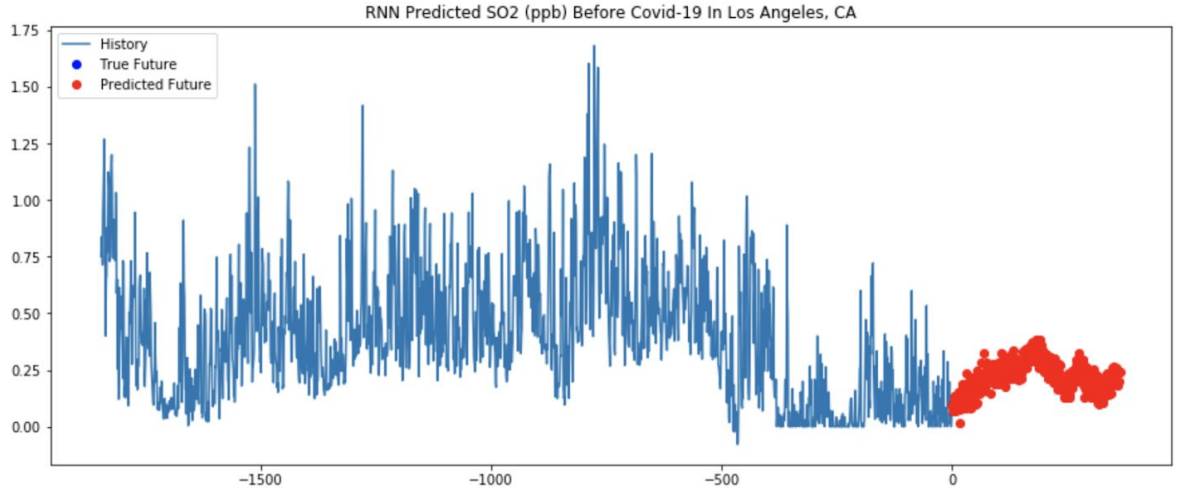


The NO_2 before COVID-19 was predicted to have much less variation and remain at a high level for the majority of the year, while the NO2 prediction after COVID-19 shows that there is more variation in the NO2 data and that it will drop well below what is predicted before COVID-19.

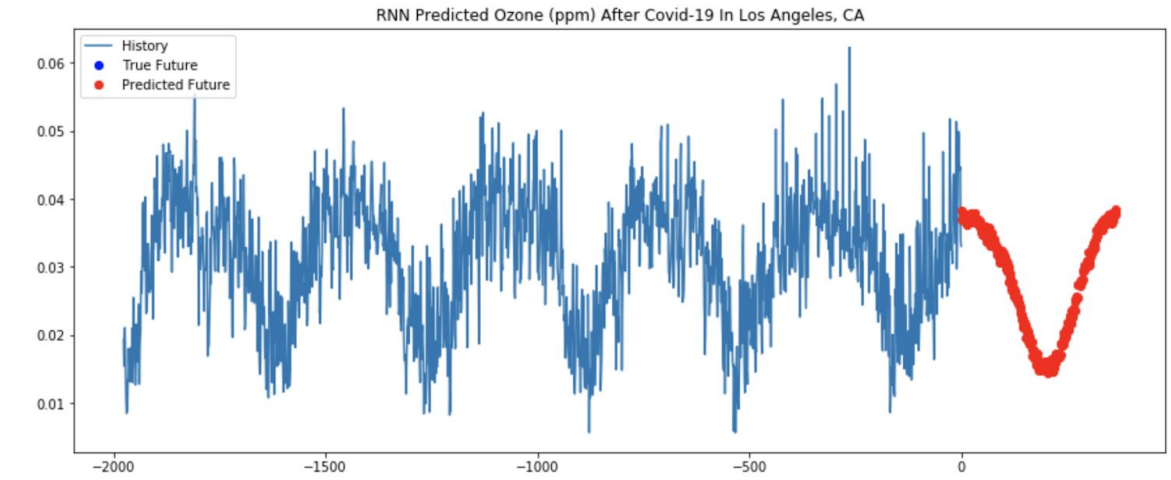


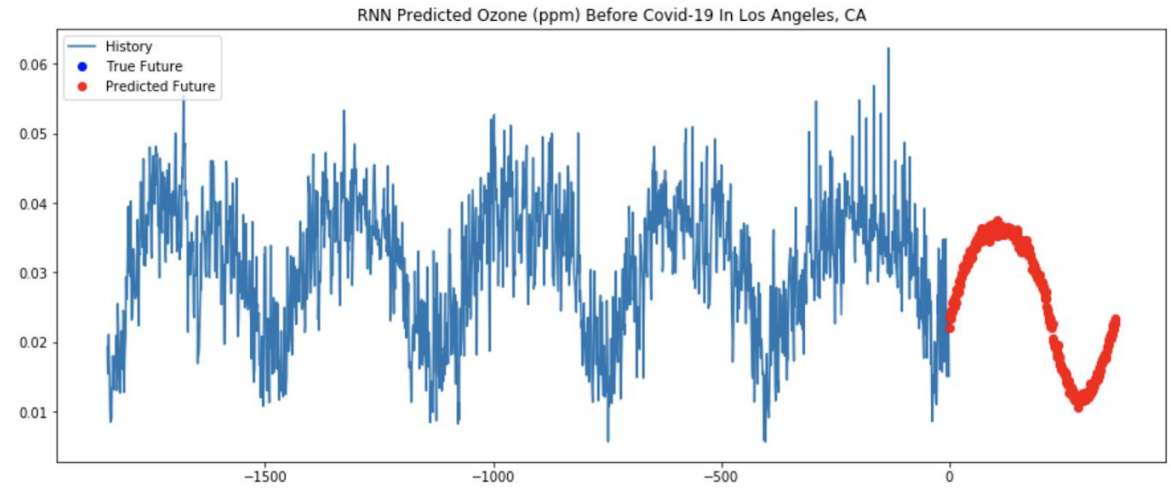
The CO before COVID-19 looks relatively similar to each other regardless of whether the COVID-19 data is taken into account for the prediction. However, the after COVID-19 era does seem to have a slightly lower peak.





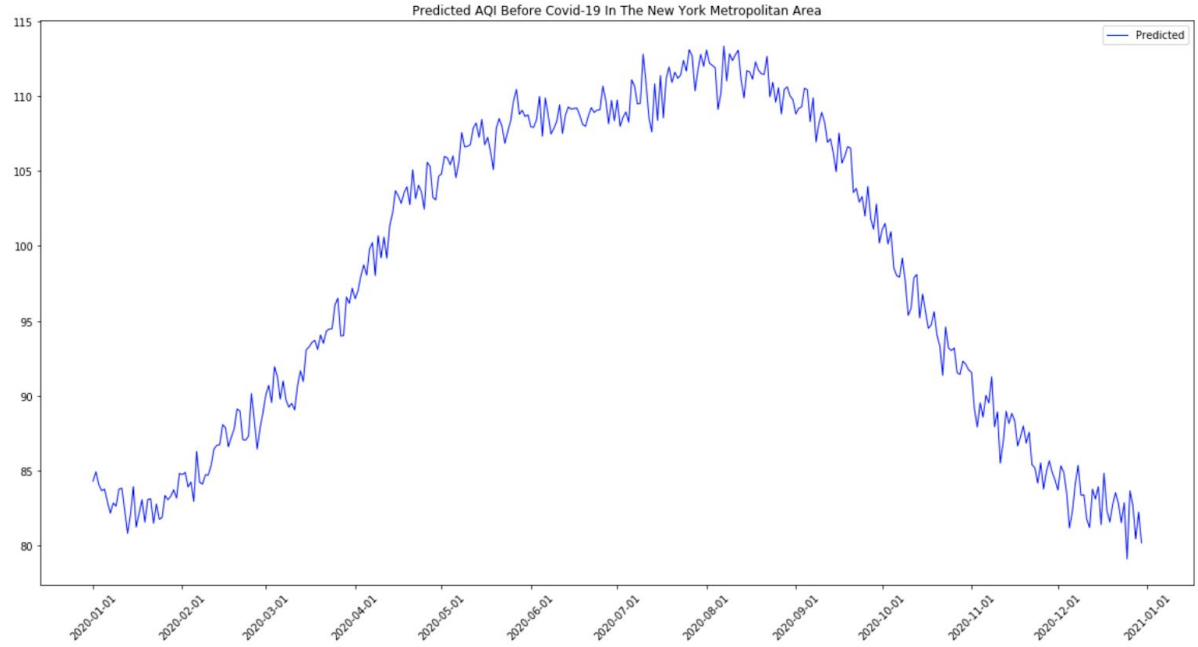
The SO2 pollution is predicted to be lower in the before COVID-19 era than in the after COVID-19 era.



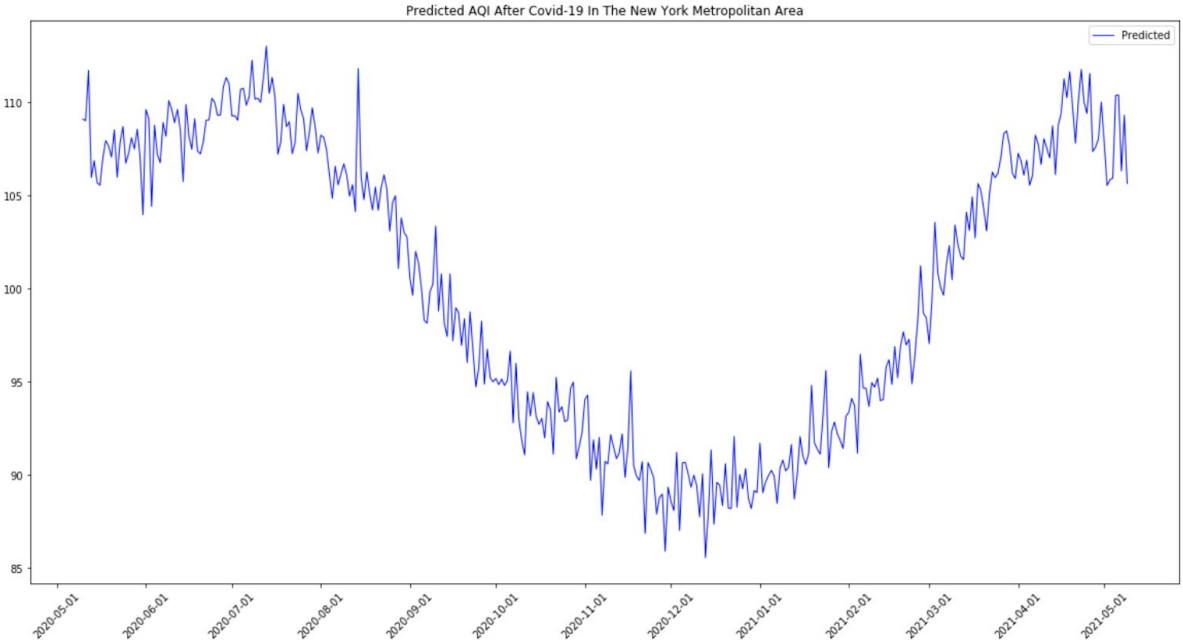


The Ozone before COVID 19 looks relatively similar to each other regardless of whether the COVID-19 data is taken into account for the prediction.

New York



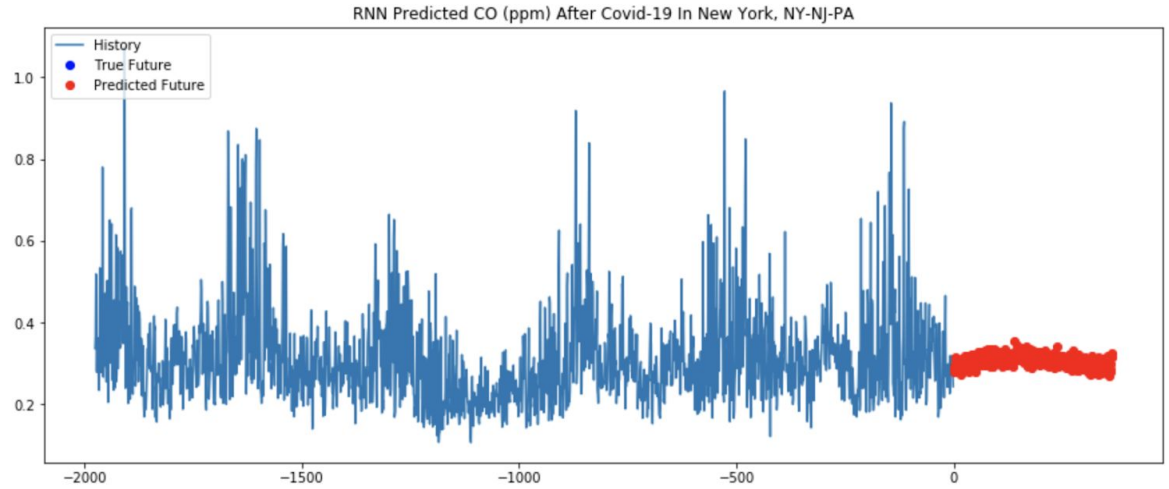
Predicted AQI from January 1st, 2020 to December 31st, 2020 (Before-COVID-19 era).

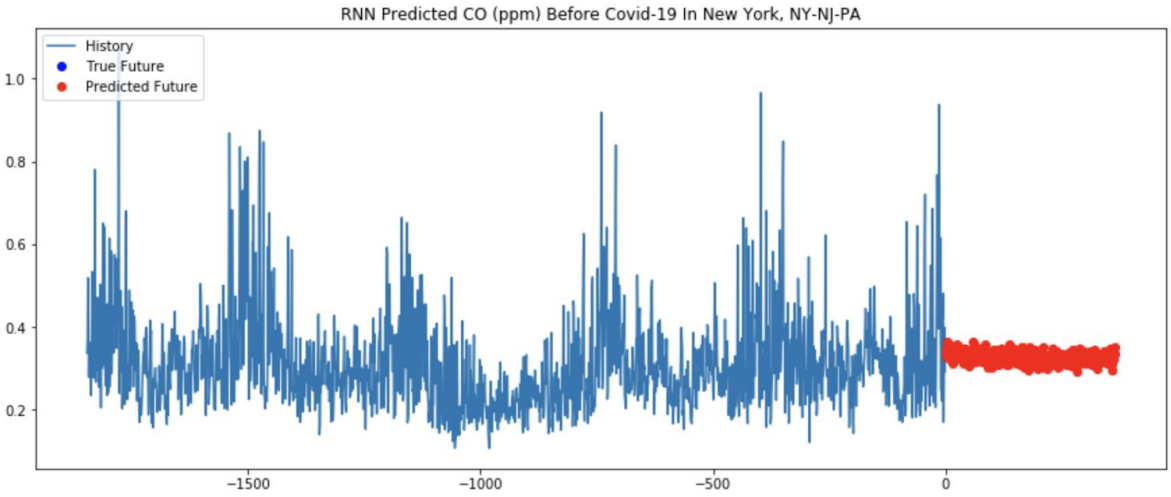


Predicted AQI from May 10th, 2020 to May 9th, 2021 (After-COVID-19 era).

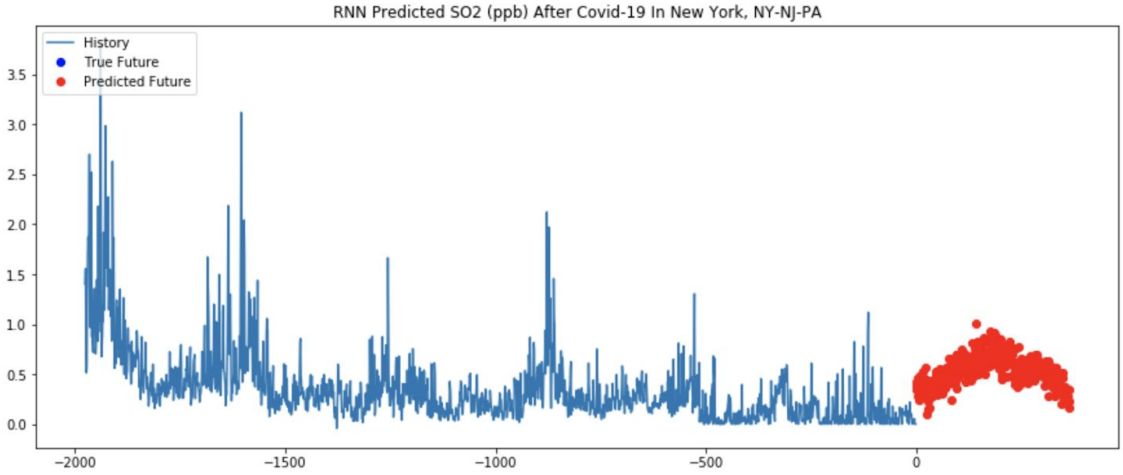
Unsurprisingly, our plot including COVID-19 data shows a more environmentally favorable future in terms of AQI. The peak in the plot excluding COVID-19 data hit a higher peak AQI.

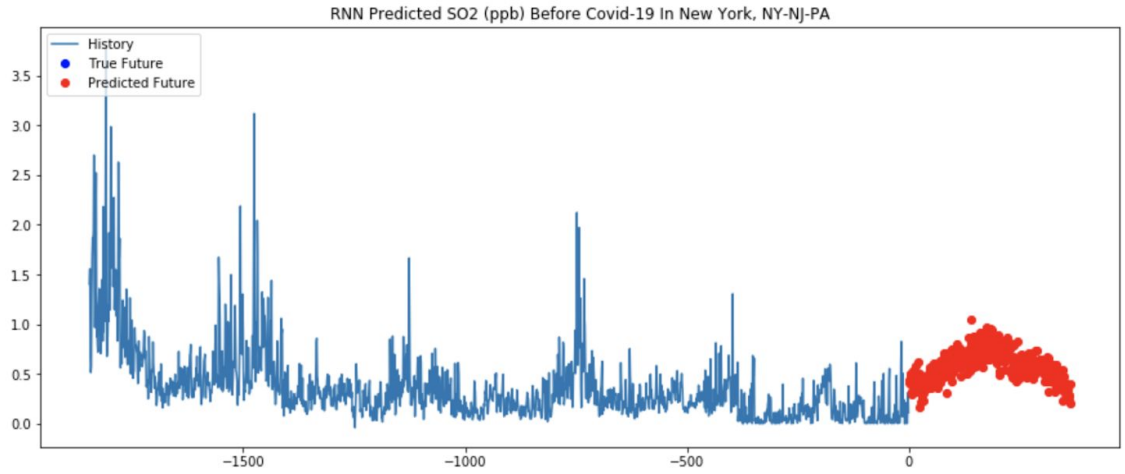
Here are the New York results for the four pollutants:



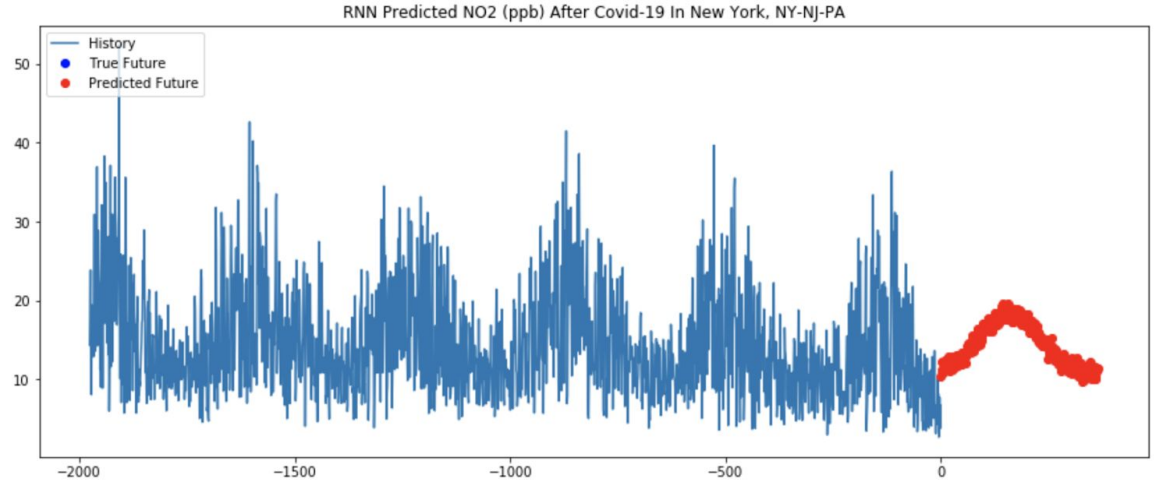


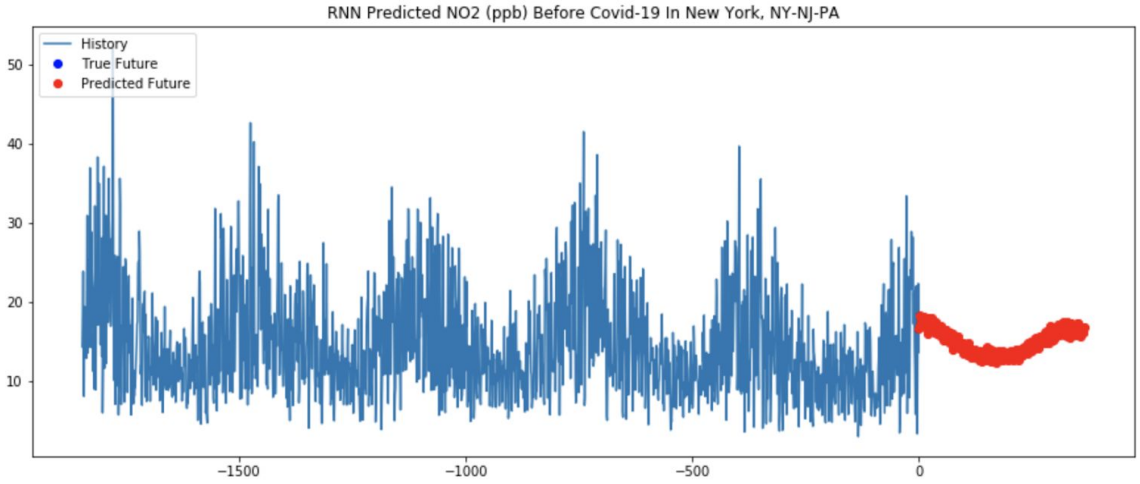
The predicted amount of carbon monoxide in the New York metropolitan area looks similar to what would have occurred without COVID-19.



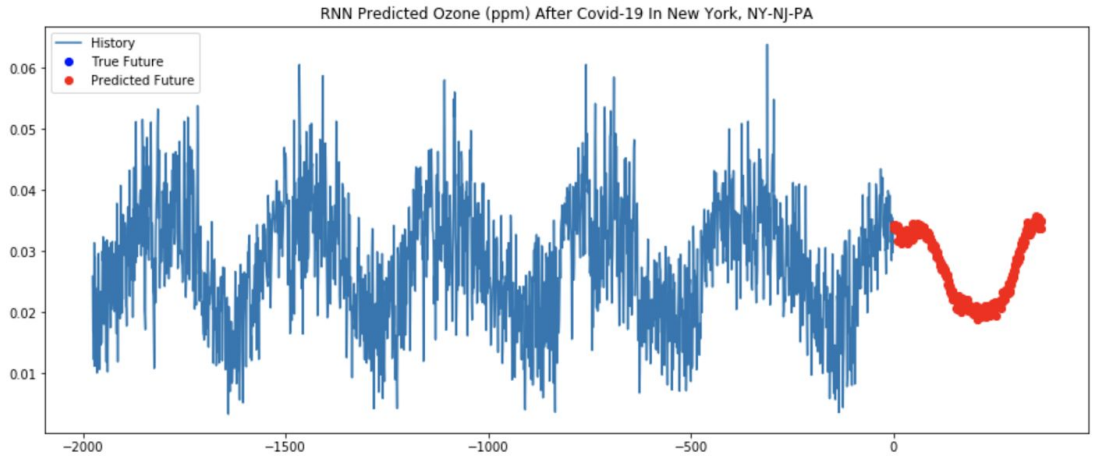


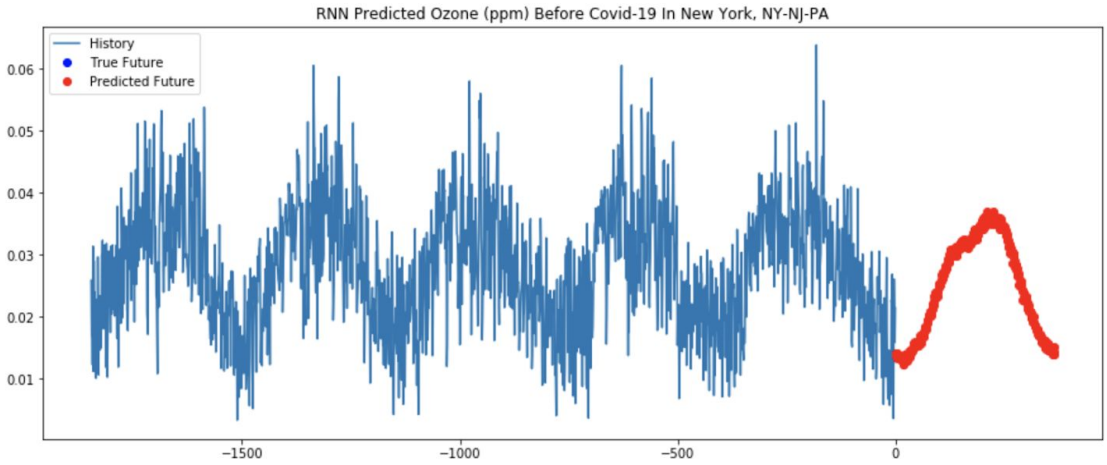
The levels of SO2 look very similar regardless of COVID-19.





The trend of NO2 in New York looks similar, however, there is less variation in the before COVID-19 prediction.





The Ozone level in the New York metropolitan area looks to have a lower ozone level than expected which could be due to COVID-19 era pollution changes.

Discussion

Before our analysis by RNN, we had expected for the predictions based on historic data with COVID-19 era data to show a downward trend for the pollutants and AQI. We also expected in our data trained before COVID-19, for future predictions to be relatively higher for our pollutants and AQI. Seeing these things proven in our future predictions, we suppose that our results are somewhat reasonable. The linear regression and the RNN when applied to real data fit it well, so we also believe that our predictions for air quality and pollutants will be credible, but not perfect, as they are just predictions. Our RNNs were most likely not well designed or perfectly trained since we have very little experience with neural networks.

Viewing our results in an environmental context, the locations we studied see a slight improvement in AQI. Future trends for our air pollution prove somewhat more hopeful, looking at our comparison of future air quality predictions. Hopefully, our predictions due to the

undertraining of our RNN tell a more cautionary story and could be improved by combining them together into a multivariate RNN instead of four univariate RNNs.

References

Popovich, B. P. and N. (2020, March 22). Traffic and Pollution Plummet as U.S. Cities Shut Down for Coronavirus, NY Times. Retrieved May 15, 2020, from www.nytimes.com/interactive/2020/03/22/climate/coronavirus-usa-traffic.html.

AirData Website: File Download page. (n.d.). Retrieved May 15, 2020, from aqs.epa.gov/aqsweb/documents/data_api.html.