An Analysis of Big Data to Curate, Analyze and Visualize the Patterns of Multiple Atmospheric Measurements in NYC to Characterize Urban Air Quality Patterns

CUNY Summer Undergraduate Research Program

ADVANCED SCIENCE RESEARCH CENTER
THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Farhan Tawsif Chowdhury and Ricardo Toledo-Crow
CUNY/ GC Advance Science Research Center, New York, NY, USA

Abstract

Greenhouse gas (GHG) emissions and air particles have been increasing at an alarming rate for the past few decades. Since 80.7% of the population lives in urban areas across the United States (U.S. Census Bureau, 2016), atmospheric particulate matter and greenhouse gases are hazardous environmental threat to the health of urban populations overall and specifically to New York City residents. Although federal, state and local regulations have reduced emissions from transportation, off-road, and stationary sources, the City's air quality still fails to meet Federal standards mostly due to ozone and fine Particulate Matter (PM2.5) levels. Studying the patterns of various components in the atmosphere will lead us to better understand some of the insights of NYC urban air quality and its spatial and temporal patterns.

My research is to compile, compare and correlate data from sensors measuring PM2.5, CO2, 03, CO, NO, NO2, temperature, pressure, relative humidity, wind speed and wind direction. A large volume of data in multiple and different datasets has been generated in near real time for several years (~1GB/year/node x 100 nodes ~ 100 GB). For data quality and consistency, we performed down sampling of the data when appropriate, checking ranges and scales and filtering outliers/noises. The analysis consisted in correlating the different datasets through comparison plotting and linear regressions. Similar efforts can be found on the BEACO2N, Purple Air, and Zentra sites whose data is included in this analysis. This analysis will lead to further measurements to determine the major sources and sinks of atmospheric components.

Background

Data Sources:

- The Berkeley Atmospheric CO2 Observation Network (BEACO2N) is a network of 100+ environmental sensor nodes deployed on school rooftops and other sites around the San Francisco Bay Area, New York City, and Houston. Each node contains low-cost commercial sensors for CO2, CO, NOx, O3, PM2.5, TP&RH, and transmits the data to a publicly accessible website.
- We collected data from a BEACO2N node on the rooftop of ASRC building together with data from two Purple Air PM2.5 sensors and an Atmos 41 weather stations and anemometers.
- These networks have open access to the data collected worldwide that can be integrated to this study in the future.

Objectives

- To gather information(data) about greenhouse gases and air pollutants and analyze them to better understand the air quality and pollution patterns over the ASRC building in Harlem, New York City.

Approach

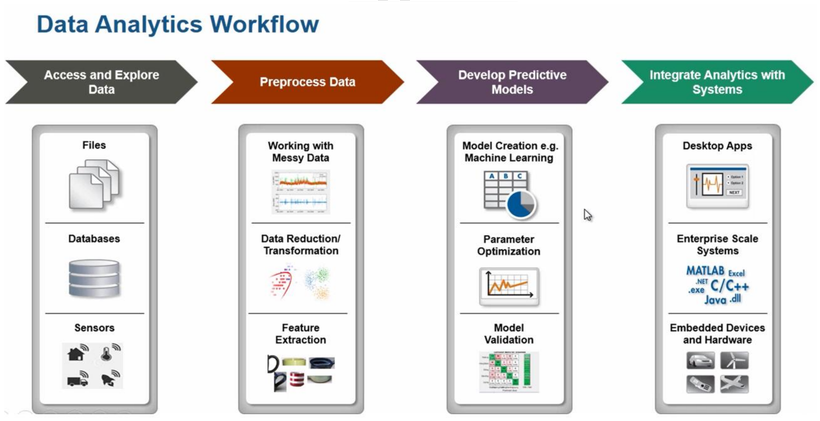


Fig.: Data analytics work flow in general (MathWorks)

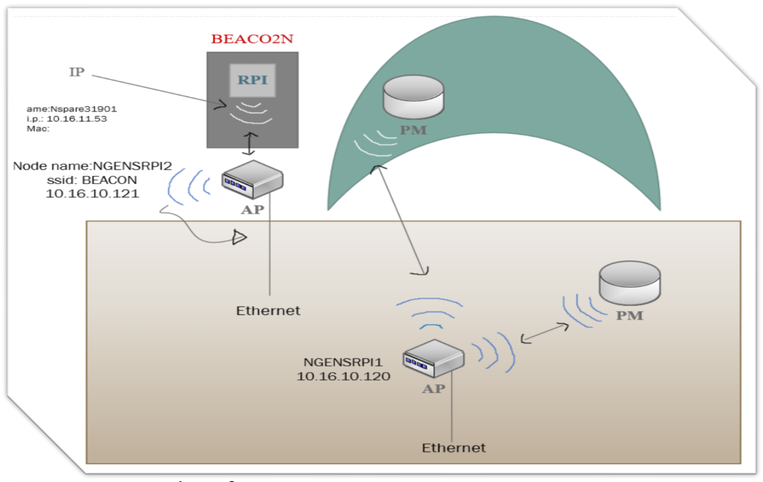
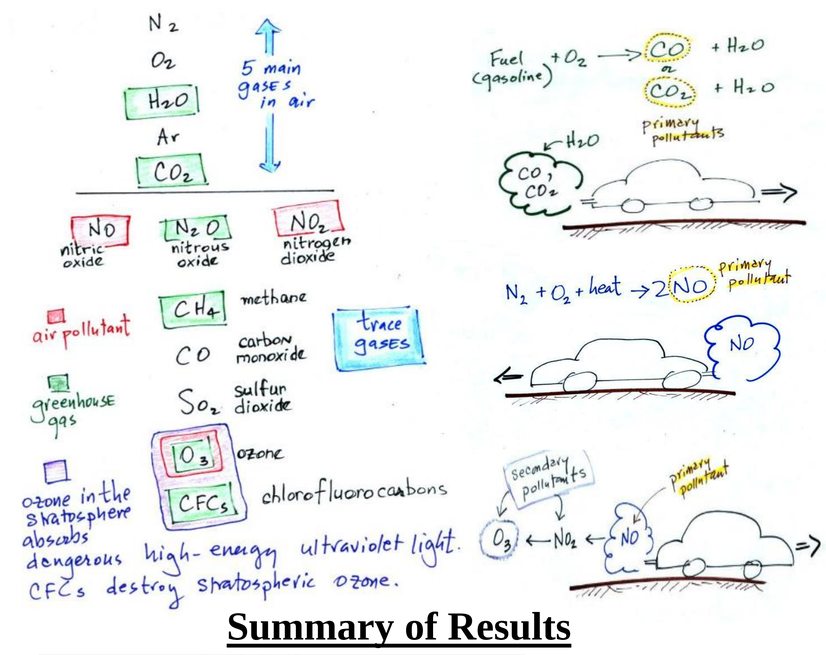


Fig.: Accessing data from sensors

Study of Pollutants[2]



Summary of Results

a)

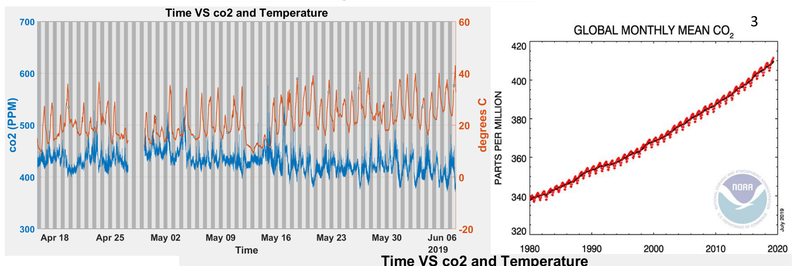
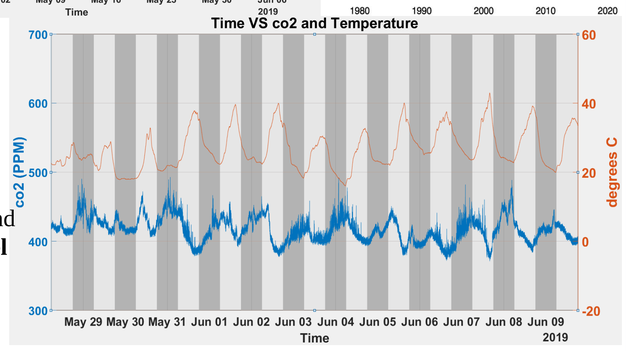
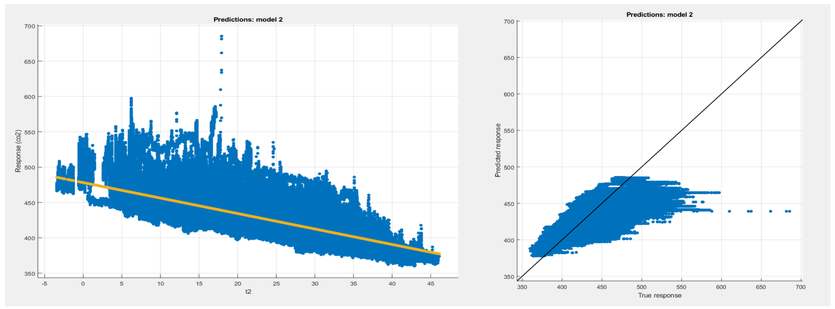


Fig a) **correlation** between **Co2 and Temperature** in different time dur--ation over the year. The variation of Co2 at **day and night**. And the **prediction model** & **linear regression** of co2 as a response.





Results: RMSE= 16.963, R-Squared= 0.53

Summary of Results(continues)

b)

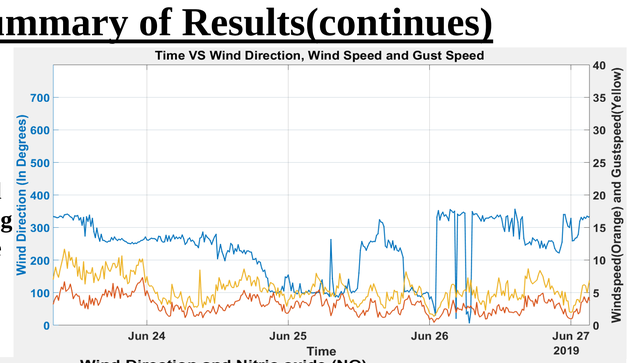
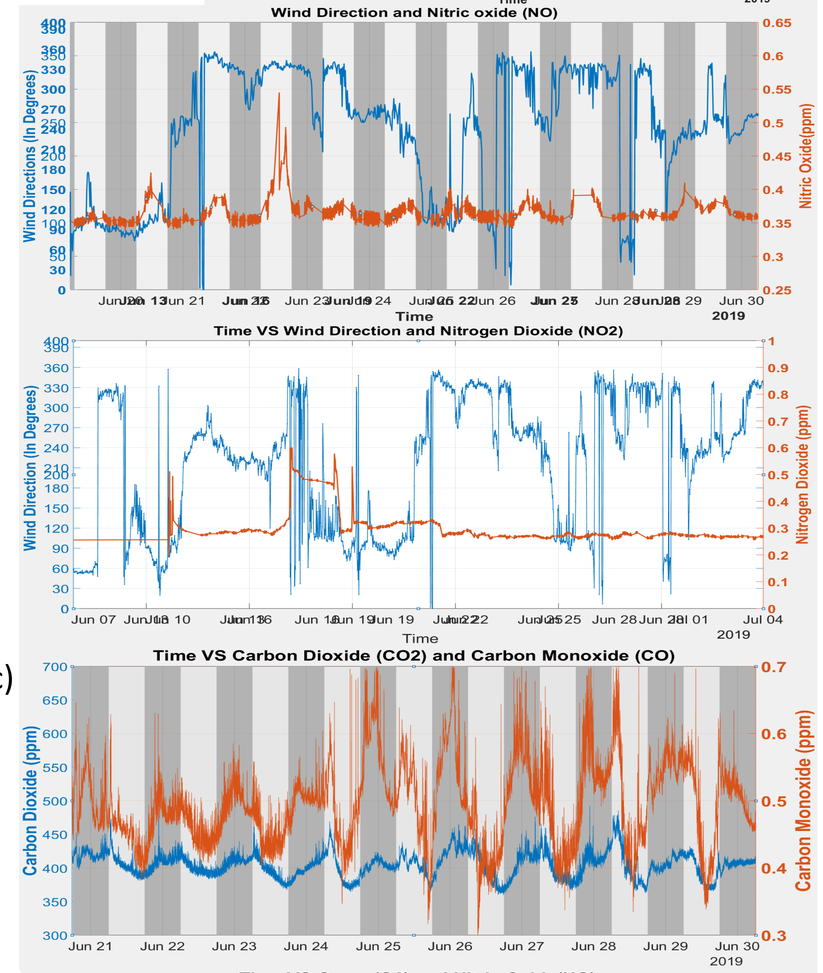
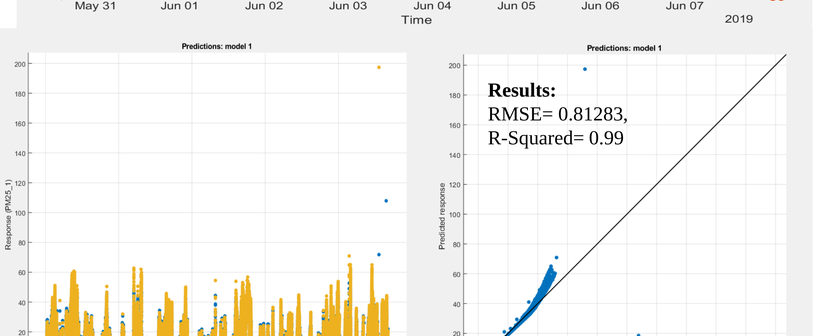


Fig b) illustrates that the measured **NO and NO2 might be coming** from to ASRC By the measurement of **wind direction**.



c)



d)

e)

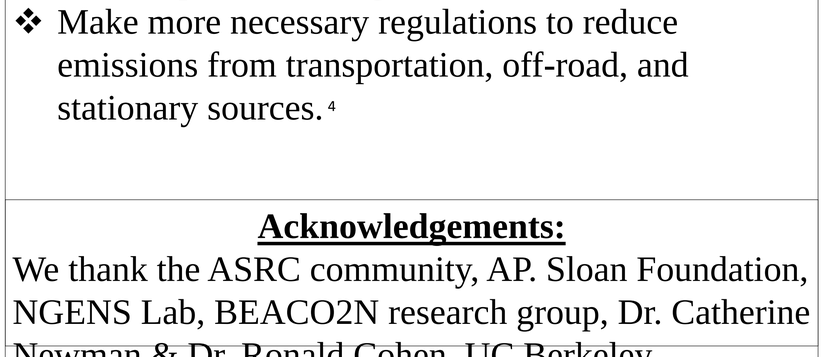


Fig e) shows that, the graph of **Particular Matter (PM 2.5) and temperature** over the time period and the **prediction** & **simple linear regression** of PM 2.5 form our data set.

Take Home Massages:

- Need to plant a lot of green lives.
- Make more necessary regulations to reduce emissions from transportation, off-road, and stationary sources.[4]

References:

[1] BEACO2N Website: http://beacon.berkeley.edu/about/
[2] University of Arizona, HAS
http://www.atmo.arizona.edu/students/courselinks/fall12/atmo170a1s1/coming_up/week_1/lect2_pollutants_pt1.html
[3] NOAA: https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html
[4] NYC CCPA

Acknowledgements:

We thank the ASRC community, AP. Sloan Foundation, NGENS Lab, BEACO2N research group, Dr. Catherine Newman & Dr. Ronald Cohen, UC Berkeley.