Final Report

For this project, I plotted my lab's personal monitor (APT sensor) air pollution data and beta attenuation monitor (BAM) air pollution data. Personal monitors are used to measure the concentration of air pollution over space and time. Stationary sensors cannot measure the concentration of air pollution over space and time because they are too large to be transported. I chose this particular topic of study because my research involves this field of study, and I am curious how accurate current personal monitors are in comparison to the stationary sensors to provide us with better insight on improving the fabrication of personal monitors. My lab works on improving upon personal monitors and informing fabrication labs on the limitations of the personal monitors when conducting field work.

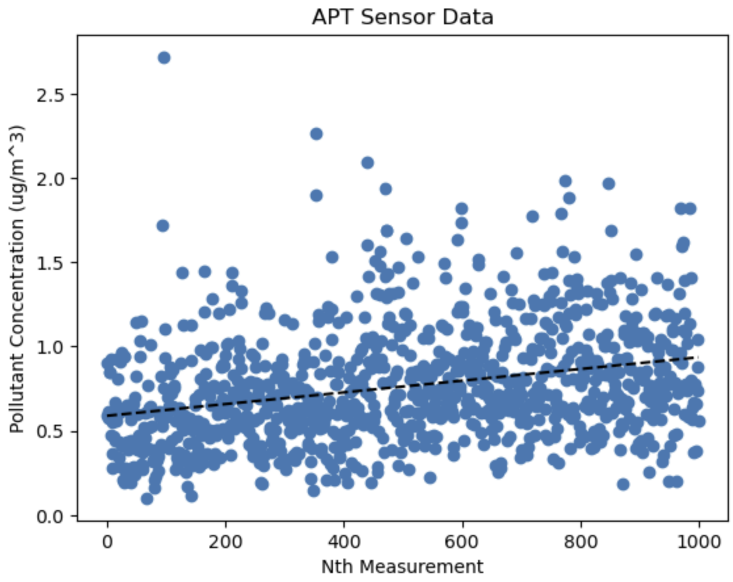
Both the APT sensor and BAM sensor measure particulate matter concentrations in ug/m^3. The BAM is a stationary sensor and is considered the 'gold standard' since it is a federally regulated method of measuring air pollutants. I used this project to determine how the APT sensor compares to the BAM sensor in performance. I graphed both the APT sensor's Particulate Matter concentrations and the BAM sensor's Particulate Matter concentrations over time. I found each sensor's corresponding linear regression metrics, such as the slope and y-intercept of the linear regression line. Since the slopes and y-intercepts were drastically different from one another, I can conclude that the APT sensor and BAM sensor measurements did not match at all.

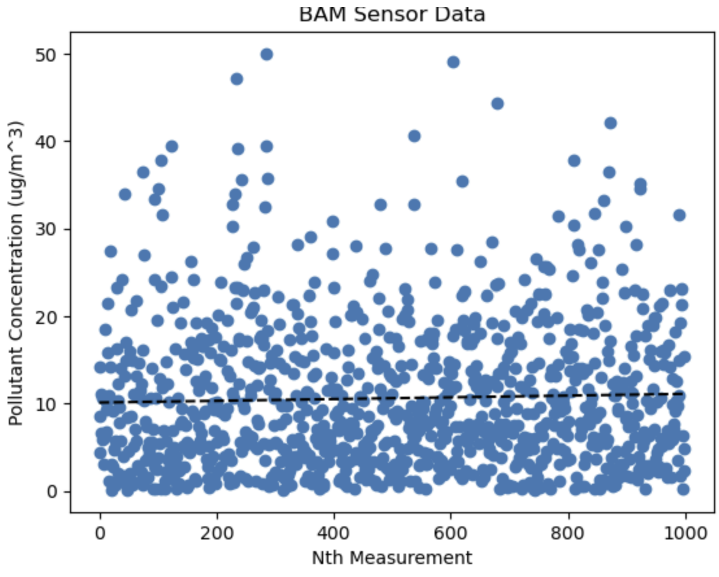
I achieved the quality filter criteria by conducting boolean indexing. This removed any non-detect values that were indicated by the number '99999'. This process also removed any instances of negative pollutant concentrations. I conducted a fit with error by implementing curve fit on my dataset, which provided me with resulting linear regression metrics.

This project was rather difficult for me because the dataset had many errors. I noticed the APT sensor dataset was showing particulate matter concentrations of max 2.5 ug/m^3 whereas the BAM dataset was showing particulate matter concentrations of max 50 ug/m^3. This was quite strange because 50 is abnormally high and can result in death when exposed at such high concentrations. I knew that this was wrong, so I checked the dataset and noticed that the dataset showed concentrations from the year 2000. This is incorrect because our lab did not deploy any sensors in 2000, we only deployed these sensors in 2023. Thus, I knew that the dataset had many incorrect values. I did not have time for this project, but, in the future, I will filter out my data so that any concentration values associated with the year 2000 are removed from the dataset.

Another challenge I came across was running the curve fit part of my code. It was giving me problems because it kept telling me the dimensions of the x and y inputs were not the same. After testing a few things out, I realized that the problem came down

to the x values being interpreted as a tuple for some reason. I had to go to office hours to figure this out. After some time, I also realized that the times don't correspond exactly on the dataset causing discrepancies in the respective scatter plot linear regressions. This is because the APT sensor collects measurements every 15 seconds whereas the BAM sensor collects measurements every 5 seconds. To extend upon this project, I will be binning the times so that they correspond exactly to each other. I unfortunately ran out of time to fix this error hence why my plots show Nth Measurement as the x-axis label instead of times. Unfortunately, my current plots don't tell me much because the Nth measurement is not all that important. For the sake of completing the project though, I left it as is. I will most definitely be addressing this in the following week in an effort to determine the accuracy of the personal monitor sensors. The plots are below:





It was very difficult for me to use these datasets. This is because the datasets required a lot of editing and cleanup. I also had problems remembering how to conduct

certain steps, and I struggled troubleshooting some of the errors I experienced that were associated with linear algebra understanding. Overall though, I learned a lot during this project. I was able to gain a better understanding of the capabilities of the Pandas and Numpy libraries, gained a better understanding of how to use linear algebra concepts in my work, and learned more about how to use the concepts learned in class for data analysis. I also realized how much patience coding requires. I got very very frustrated very frequently, and I would want to give up many times. I am hoping to learn to get better at coding over time, so that I don't start having a random panic attack when things are not working for the 100th time. I think this class gave me a lot of insight on how to improve upon my coding skills but to also think about data analysis more critically. I am excited to add coding to my toolbox to apply to my field of research in environmental engineering in the future.