Filling in the Gaps: Using Consumer Products to Replace Missing Pollution Data

Aaron Watt

February 2022

Abstract

[abstract here]

1 Introduction

A critical input to good air quality regulation is good air quality measurements. Specifically, the relative efficiency of current pollution regulation hinges on our ability to accurately assess air quality across the country. In the United States, air quality is assessed by the government using a network of monitors that measure levels of ambient air pollution to a high degree of accuracy. The Environmental Protection Agency (EPA) requires these monitors to measure average daily air quality at specific frequencies to ensure enough data is collected for effective regulation.[1] During the days that are required to be measured, the goal is to accurately measure the daily average pollution concentration at the site of the monitor. Statistics of these daily averages, called *design values*, are then used to decide if a region is in or out of compliance with the National Ambient Air Quality Standards (NAAQS).

Though these air quality monitoring stations are regulated by the EPA, they are managed by local and state officials who control when the monitors are on or off. For added flexibility, the EPA allows for some portion of air quality readings to be missing when calculating the design values that determine a region's compliance with the NAAQS. For instance, when

[1] The three main measurement frequencies require measuring daily average air quality every 1, 3 or 6 days.

measuring particulate matter in the air (one of the most common types of pollution), the EPA allows more than 25% of measurements to be missing or omitted (EPA, 2017).[2] In effect, this flexibility means that local managers of monitoring stations can choose up to 25% of readings to omit – readings that would otherwise be used in determination of compliance. Though the measurements of air quality at the site of the monitor are fairly accurate when the monitor is on, omitting some measurements (by turning the monitor off) can bias the daily average and compliance statistics calculated from reported measurements. Additionally, if a region is out of compliance with the standards, the region or state can potentially face large penalties and forced adoption of expensive abatement technology.

The combination of large penalties and the ability to affect compliance statistics indicates misaligned incentives between federal regulators and local officials in charge of monitoring air quality, potentially leading to mismeasured air quality statistics. Indeed, previous research suggests that there is mismeasurement of air quality statistics occurring; Zou (2021) and Mu et al. (2021) provide evidence of strategic behavior in pollution measurement on behalf of local pollution regulators. This paper focuses the size of mismeasurement occurring and the effects this mismeasurement has on determining compliance.

Specifically, I explore the question: is there a bias in reported air quality data and how might this bias affect NAAQS compliance? To explore these issues, I utilize a new dataset of air quality measurements collected from consumer products (PurpleAir sensors). These data help provide an independent groundtruth comparison to air quality reported to the EPA. The most promising new data coming from these consumer products are PM2.5 measurements – the concentration of particles in air that are 2.5 micrometers and below. Specifically, I combine PM2.5 measurements from multiple PurpleAir sensors that are near to federally-regulated monitoring stations to estimate the PM2.5 value at the monitoring station; I use

[2]Design values are used to decide compliance with NAAQS and are statistics of daily averages. In calculating daily averages, the daily average is valid if at least 75% of the hourly readings (18 of 24 hours) are reported and valid. In calculating the design values, the design value is valid if at least 75% of daily averages in each year are reported and valid. Combined, the minimum reporting standard is actually 56-57% of all required hourly PM2.5 readings. This is slightly different for each site depending on their reporting frequency (every 1, 3, or 6 days).

inverse distance weighting to create a weighted average of PurpleAir measurements.[3] This allows me to construct predicted values of PM2.5 at the station during times when the station's readings would be used to calculated NAAQS compliance but when the station was shut down. I first examine how these predicted missing PM2.5 values compare to the reported values, then use the combination of reported and predicted values to generate counterfactual NAAQS compliance statistics.

The NAAQS compliance statistics for PM2.5, called *design values*, are functions of the daily averages reported by air quality monitors. There are two primary design values for PM2.5: the "annual" design value is a three-year average of the daily averages; and "24-hour" design value is a three-year average of the annual 98^{th} percentile of daily averages.[4] Each quarter (3-month period), these two design values are calculated and compared to the NAAQS for PM2.5. If a station's design value is above the standard, then the station (and associated region) is determined to be in *non-attainment* (non-compliance) with the standard for that quarter. Using the reconstructed dataset of PM2.5 (PM2.5 estimates for all hours that would be reported from a given station), I construct counterfactual estimates of the design values that determine if a region is in or out of attainment. I use these counterfactuals to determine which regions would have changed compliance status status if they reported 100% of their PM2.5 measurements – I call these "flipped regions". I also examine how close these flipped regions were to the regulatory threshold and report a measure of the bias related to the station's missing PM2.5 readings.

Though I am examining the affect of pollution data that is missing from a monitor's record (data missing *in time*), there is also the issue of attempting to measure a region's ambient air quality using spatially sparse locations of monitors (you could consider this and issue of data missing *in space*). Previous literature has examined the sparse distribution

[3]Inverse distance weighting has drawbacks: it can apply very large weight to sensors very near to the NAAQS monitor and it does not take into account that some PurpleAir sensors will be better predictors for the NAAQS monitor. See Appendix section 8.1 for an alternative strategy that I plan to implement.

[4]these statistics are discussed more in the Data section. Specific formulas for these statistics are listed in the appendix.

of regulation-grade monitors and the resulting sensitivity of CAA air quality regulation. Grainger et al. (2019) and Grainger and Schreiber (2019) identify a principle-agent problem with the initial spatial placement of sparse pollution monitors; they find evidence that local regulators may be strategically locating their air quality monitors based on pollution, and possibly socioeconomic characteristics. To address the issue of sparse data and fill in the gaps, several authors have used satellite data products to provide finer resolution pollution data (Sullivan and Krupnick 2018, Fowlie et al. 2019). Moving to more time-based issues, Zou (2021) also uses satellite estimates to discuss the issue of strategic behavior in reaction to the timing of pollution monitoring. He provides evidence that some areas have significantly worse air quality on unmonitored days. In related work, Mu et al. (2021) show potential for strategic monitor shutdowns on days of expected high pollution, contributing to air quality data that is missing *in time*.

This paper is most similar to the analysis in Fowlie et al. (2019) where they use PM2.5 estimates generated from satellite data to examine counterfactual compliance status. However, they end their analysis noting that the satellite-based data commonly used in these applications has significant prediction error in some areas; this can cause result in incorrect conclusions about design values. This paper compliments their analysis and that of Mu et al., where I use a different form of ground-truth PM2.5 data to also address missing air quality data *in time*. [5] In contrast to Mu et al. however, I am examining pollution at times when it is missing in the data but required to be reported, whereas their work was on pollution at times that are not required to be reported. While satellite-based PM2.5 estimates have potential for large prediction errors, PurpleAir sensors can be fairly accurate measures of their local air quality[6] and can be averaged over multiple nearby sensors. PurpleAir data also have drawbacks however – the sensors are highly non-uniform in coverage across the

[5]PurpleAir data, and other on-the-ground pollution sensors, also have the potential to examine issues of spatial distributions of monitor networks – work left for future research.

[6]PurpleAir sensors have specifically been shown to be less accurate than regulation-grade monitors at high levels of PM2.5 concentration. However, the EPA has developed a correction technique that result in PurpleAir readings within 5% of co-located EPA monitors. This correction technique is used here and explained in more detail in the appendix.

US and are sensitive to specific placement by the consumer, perhaps leading to hyper-local estimates of air quality.

For these reasons, this analysis should be seen as a compliment to previous works. As consumer sensors become more widespread, we can augment reliable federal air quality measurements with a growing number of auxiliary data points to better understand the shape of mismeasurement in air quality. In this paper, I explore one way of leveraging these data to test for issues with biased reporting of air quality. After predicting missing observations using PurpleAir measurements, I find that XX of XX monitoring stations flipped their NAAQS compliance status in at least one quarter between 2018 and 2020 and that YY had more than one flipped status. On average, the design values for the monitor-quarters that flipped status had a difference of more than AAμg/m^3 and BBμg/m^3 of PM2.5 for the annual and 24-hour design values, respectively. We know from previous research that pollution in non-attainment areas has been decreasing at significantly faster rates since the introduction of the CAA (Currie et al., 2020); combined with my results, this suggests that changes in reporting standards to decrease allowable omitted observations may result in more non-attainment areas and further increases regulatory efficiency.

The remainder of this article is organized as follows. Section 2 briefly reviews the history of air quality standard in the US and some key details of current regulations. Section 3 explains the theoretical framework applied to the problem of measuring pollution at specific points in space. Section 4 then discusses the data used and section 5 describes the empirical framework that will be applied to estimate the missing pollution and resulting policy outcomes. Section 6 reviews the results of the empirical study and discusses the implications. Section 7

2 Background

Amid growing public concern about air quality and pollution, the United States Congress passed the Clean Air Act of 1963 (CAA). Later additions to the CAA, the Clean Air Amendments of 1970, granted the Environmental Protection Agency (EPA) the regulatory authority to create and enforce air quality standards in the US. One major way air quality is regulated is through the National Ambient Air Quality Standards (NAAQS), which set concentration thresholds for a list of different "criteria" pollutants (91st US Congress, 1970). The EPA has since been in charge of setting and updating the NAAQS and require states to submit plans to bring their air quality to within NAAQS limits. An important aspect of enforcing the NAAQS is measuring criteria pollutants across the US by requiring states to install pollution monitoring stations in areas of questionable air quality. Because these monitoring stations are used for potentially costly enforcement, the equipment within each station must abide by specific regulations and are relatively costly to install and run.

Over the last decade, commercially available scientific equipment in measuring various air pollutants has evolved. There is now relatively cheap[7] equipment available to measure particulate matter (one of the criteria pollutants that regulated by the NAAQS). Specifically, the PurpleAir company produces devices that can measure particulate matter that has a diameter of less than 2.5 micrometers (designated as PM2.5).[8] PurpleAir is of particular interest because they have built an opt-out mechanism for end-users to allow their ambient air quality data to be stored in the cloud. They also provide multiple ways for researchers and the general public to use this crowd-sourced air quality data.

This paper is primarily concerned with the minimum reporting requirement. As with

[7] e.g., a PurpleAir outdoor air quality sensor is about $250 to purchase with little upkeep from the end user, compared to roughly $100,000-200,000 to install EPA regulation-grade criteria pollutant monitors and trained staff to upkeep and record measurements. The cost alone is not a good comparison because the EPA monitors use different technology that is known to be more accurate across a wider range of pollution concentrations, have a better sense the sensor error, and measure more pollutants than the PurpleAir monitors. For the purposes of this analysis, PurpleAir monitors should be seen as a compliment to EPA monitors, not a potential replacement.

[8] PurpleAir devices can measure a few other criteria pollutants (namely ozone and PM10) but the comparability of the PM2.5 measurements between PurpleAir and EPA monitors are currently better understood.

many federal regulations, there are many ways that states or emitters can cleverly navigate the rules to emit more than they are meant to according to the spirit of the regulation. One way of navigating the CAA regulations is through the choice of what data to report. The EPA currently requires a minimum threshold of air quality data to be reported – for PM 2.5 75% of daily measurements for

3 Theoretical Framework

Note for Meredith: I simplified the prediction of EPA monitor data from using wind direction to an inverse-distance weighted average. I plan to implement the more complex (but more predictive) model in the next iteration and it is described in the appendix.

3.1 Estimating PM2.5 at the NAAQS Monitor with PurpleAir Sensors

- add description of current IDW prediction
- add equation of current IDW
- add footnote saying "IDW produces fairly poor estimates of PM2.5 at the location of the NAAQS monitor. I plan to implement a more rich prediction model using wind speed and direction – see the Appendix section XXX for the model and data."

3.2 Estimating Design Values

- explain design values and the use of predicted PM2.5 data
- add design value equations
- add design value table (12 and 15 μg/m^3)
- add design value $98^t h$ percentile lookup table from EPA documentation.

4 Data and Descriptive Statistics

4.1 EPA Regulation-grade Monitors

Highlights:

- 388 EPA regulatory-grade monitors used for NAAQS determination for PM2.5.
- I am using 15 monitors in California that take hourly readings every day.
- Future analysis will include the rest of the NAAQS monitors.

There are approximately 388 air quality monitors around the US that are used for NAAQS determination for PM2.5. The EPA refers to these as NAAQS-primary monitors; I will refer to these monitors as *NAAQS monitors*. There are more regulation-grade monitors that meet or approach the regulatory accuracy standards set by the EPA, but these 388 are the monitors that are officially used to calculate the design values that decide attainment status. I say approximately 388 because regulation-grade monitors are occasionally moved in and out of the list that determines NAAQS attainment status, so this number can change. Of the 388 NAAQS-primary monitors in the US, I conduct my preliminary analysis on the 15 monitors in California that take hourly readings every day. Future analysis will include the rest of the NAAQS-primary monitors, which include monitors that only report daily averages (as opposed to hourly averages) and monitors that only report every 2, 3, 6, or 12 days.

4.2 PurpleAir Consumer Sensors

Highlights:

- Goal: predict missing PM2.5 hourly averages from EPA NAAQS-primary monitors using nearby PurpleAir PM2.5 sensors.
- Limited the analysis to include PurpleAir sensors within 5 miles of each NAAQS-primary monitor.

- 16,038 PurpleAir outdoor, publicly-shared PM2.5 sensors in the United States have some accessible data.
- 10,401 sensors in California, Oregon, Nevada, and Arizona.
- 11,205 instances of PurpleAir sensors within 50 miles of the 15 NAAQS-primary monitors (double counting permitted).
- 7,777 unique PurpleAir sensors within 50 miles of 15 NAAQS-primary monitors.
- 592 unique PurpleAir sensors within 5 miles of 15 NAAQS-primary monitors (final sample of PA sensors contributing to estimated EPA monitor values).
- I correct PurpleAir PM2.5 values using EPA's correction equation.
- For each hour that there are valid PurpleAir sensor readings within 5 miles of the EPA monitor, I calculate an inverse-distance weighted average PM2.5 level.
- Future work includes a better PurpleAir prediction for EPA monitor PM2.5 measurements, using wind speed and direction.
- There are some PurpleAir sensors that are biasing the prediction of PM2.5 – seem to be measuring localized pollution that the EPA monitor does not pick up. These would be down-weighted or removed in the future version of the prediction mentioned above.
- I show plots for an example NAAQS monitor in Los Angeles, CA. Plots for other monitors are in the appendix.

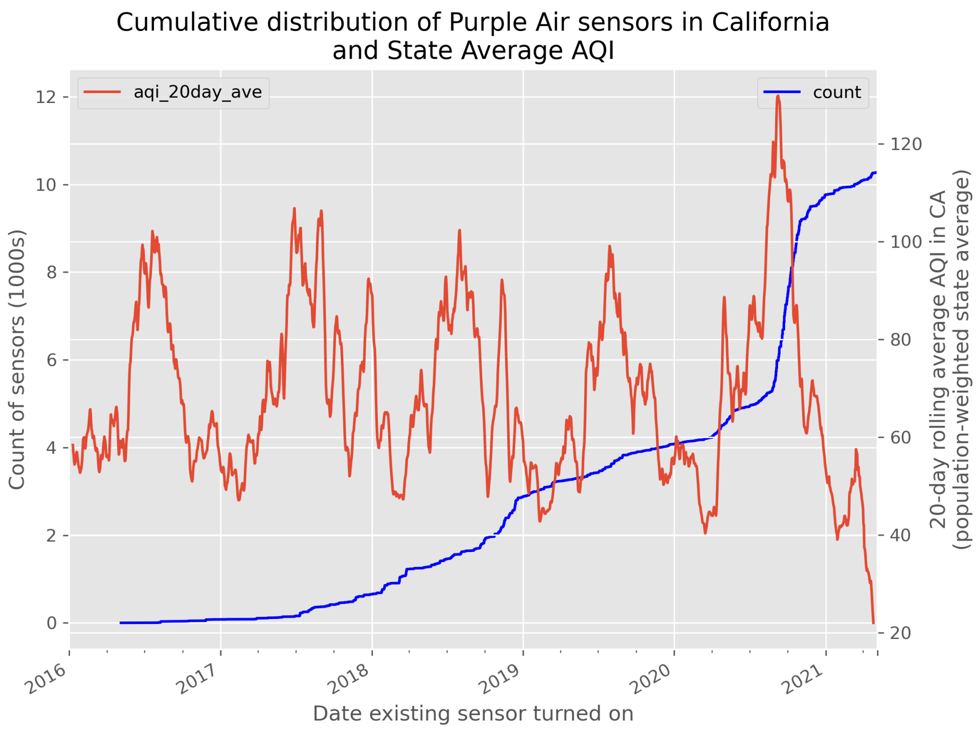

Figure 1: (Red) 20-day rolling average of the Air Quality Index in California. (Blue) Cumulative PurpleAir outdoor sensors posting public PM2.5 data.

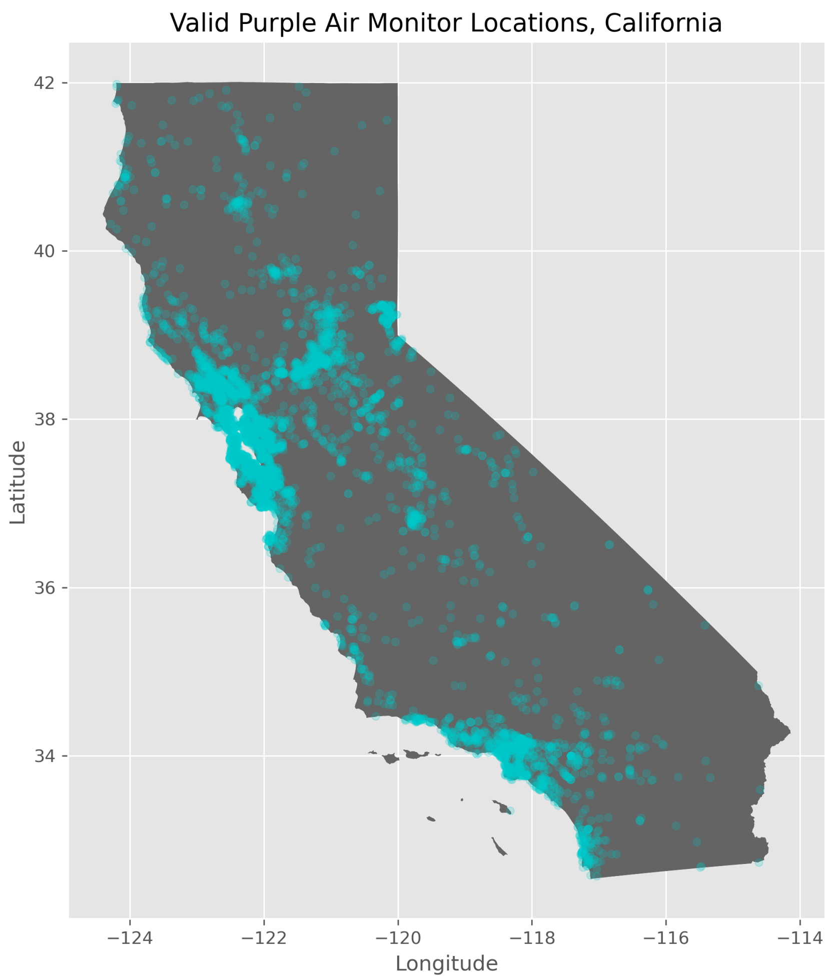

Figure 2: Map of PurpleAir sensors offering public, outdoor PM2.5 measurements. These are sensors that have offered any data in the past, so many are now inactive. The historical data is used in this analysis.

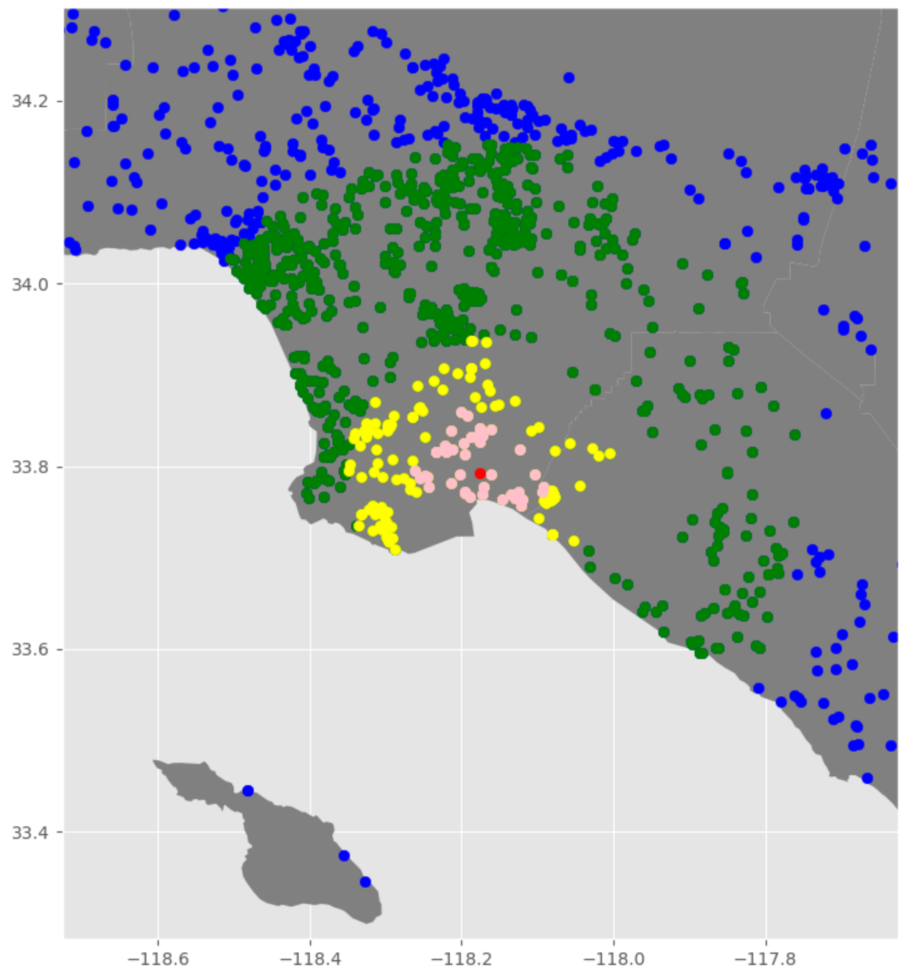

Figure 3: Map of an EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii. This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers).

As a measure of predictive power, I should report the R^2 value of a regression of NAAQS monitor PM2.5 on the PurpleAir PM2.5. I also need to fix the legend in Fig. 5.

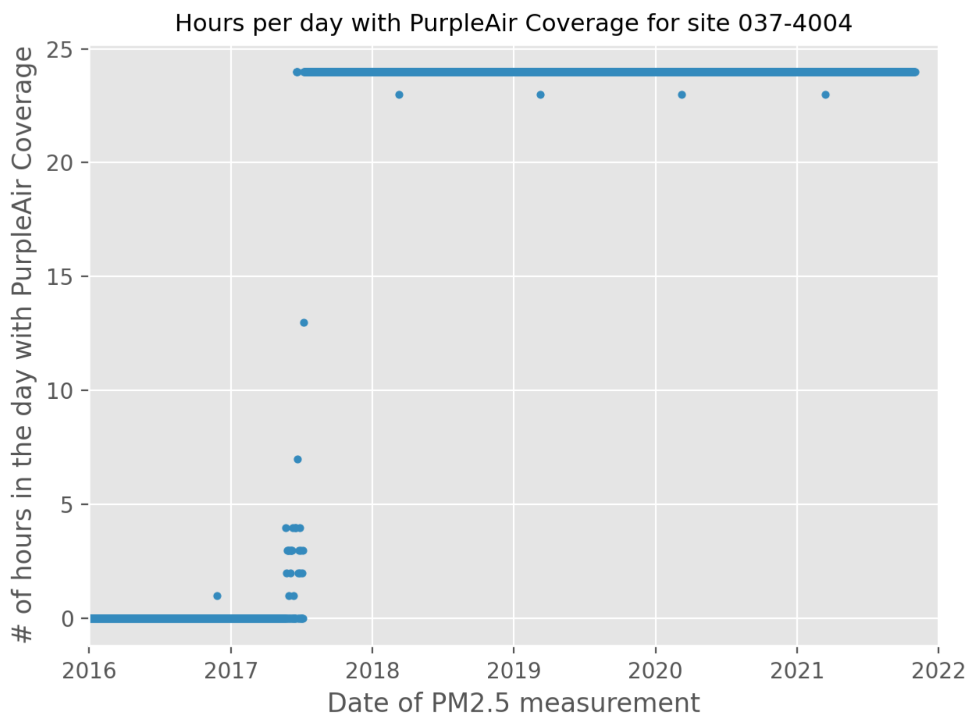

Figure 4: Scatter plot indicating the number of hours in each day that this NAAQS-primary monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles.

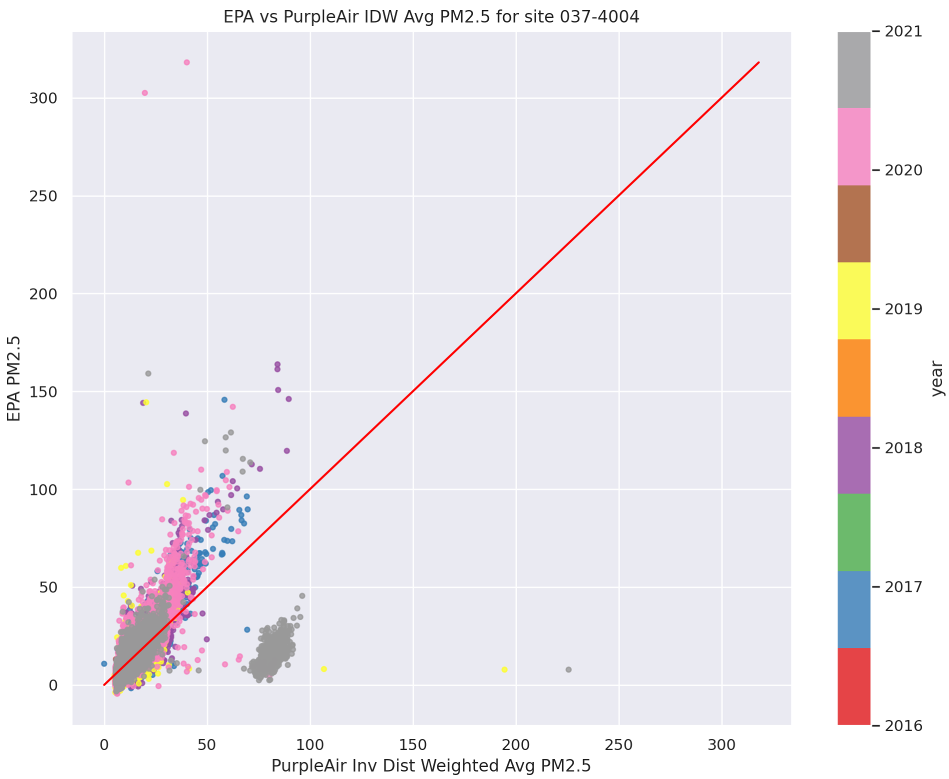

Figure 5: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. For this site, we can see the PurpleAir average is skewed to the right for readings from 2021. This is likely from a PurpleAir sensor coming online that was placed near a source of localized pollution that is not being picked up by the NAAQS-primary monitor.

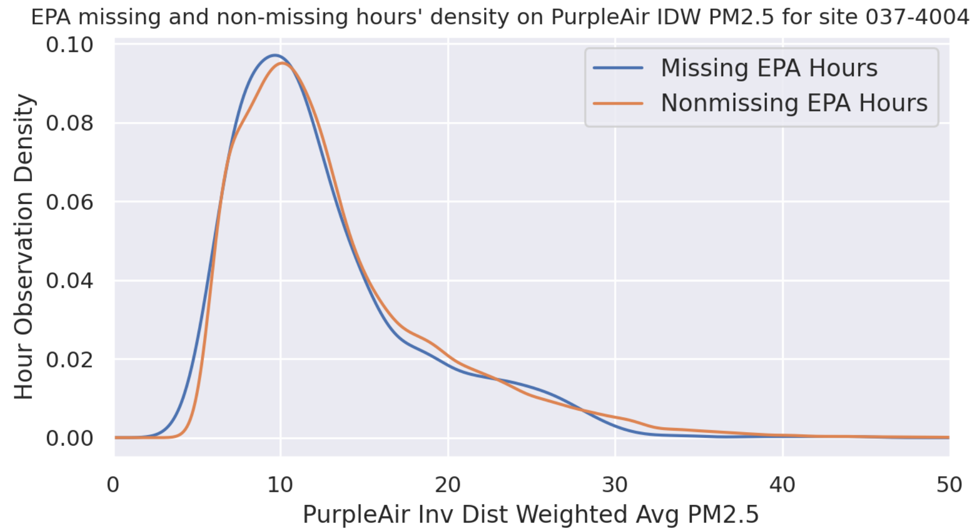

Figure 6: Comparison of PM2.5 concentration densities for two sets of hours: reported (orange) and missing (blue) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location.

4.3 Estimating Regulation-grade Readings with PurpleAir Sensors

I use inverse-distance weighting (IDW) with a power of 1 on the denominator. In their discussion of IDW in ambient pollution estimation, de Mesnard[9] derives that a power between 1 and 3 is appropriate for diffuse particle distributions. I use a power of 1 here because I find evidence that some PurpleAir sensors that are very close to the NAAQS monitor are not very good predictors of the monitor's PM2.5 levels. These PurpleAir sensors seem to still have reliable estimates[10], and anecdotally seem to have very high PM2.5 readings when they disagree with the NAAQS monitor. This suggests they are measuring localized pollution that is out of the range of the NAAQS monitor (these sensors could be located next to a highway, for example).

[9] Fix this citation: de Mesnard "Pollution Models and Inverse Distance Weighting."

[10] Each PurpleAir sensor has two internal sensors that measure PM2.5. Reliability of the PurpleAir sensors is determined by the agreement of the two sensors' hourly averages.

I plan to fix this issue with a future implementation of a better prediction model[11]. In this iteration, I have implemented the sub-optimal IDW average to avoid excluding entire PurpleAir sensors and removing potentially useful sensor data.

5 Empirical Framework

- Describe hourly IDW average

6 Results and Discussion

- None of the 15 sites I tested had noticeable differences in the PM2.5 distributions for missing and reported hours (e.g., Fig. 6).
- I would like to calculate two sets of design values: one set for all the reported PM2.5 hourly measurements from the NAAQS monitor, and another set for the reported hourly measurements combined with estimated hourly measurements during the hours that were missing in the data. I would like to report the difference in those sets of design values.
- I have not yet calculated the design values, but based on the high level of visual agreement in the density plots of the reported and missing PM2.5, I do not think they will be significantly different after including the estimated PM2.5 values for the missing hours.

Future Work

- I have only tested 15 of 388 possible sites. Now that I have written the bulk of the code the manage the data, most of the future work lies in acquiring the rest of the data for the NAAQS monitors and PurpleAir monitors.

[11] See Appendix section 8.1

- I am in the process of negotiating a data usage agreement with PurpleAir staff to get the entire historical dataset of all US sensors. This is the main bottleneck.
- I have written code to download any arbitrary NAAQS sensor data between two dates. However, some of the NAAQS monitors report on a less frequent basis than hourly, so the analysis code will need to be generalized for other reporting frequencies.
- I have written code to download wind velocity data, but parsing the files to get wind velocity at a particular location and time requires more work. This will be used in my predictive model for missing PM2.5 data at the NAAQS monitor.

7 Summary and Concluding Remarks

References

91st US Congress (1970, December). Clean Air Amendments of 1970.

Currie, J., J. Voorheis, and R. Walker (2020, January). What Caused Racial Disparities in Particulate Exposure to Fall? New Evidence from the Clean Air Act and Satellite-Based Measures of Air Quality. Working Paper 26659, National Bureau of Economic Research. Series: Working Paper Series.

EPA, E. P. A. (2017). Appendix N to Part 50, Title 40 – Interpretation of the National Ambient Air Quality Standards for PM2.5.

Fowlie, M., E. Rubin, and R. Walker (2019, May). Bringing Satellite-Based Air Quality Estimates Down to Earth. *AEA Papers and Proceedings 109*, 283–288.

Grainger, C. and A. Schreiber (2019, May). Discrimination in Ambient Air Pollution Monitoring? *AEA Papers and Proceedings 109*, 277–282.

Grainger, C., A. Schreiber, and W. Chang (2019). Do Regulators Strategically Avoid Pollution Hotspots when Siting Monitors? Evidence from Remote Sensing of Air Pollution. Working Paper.

Mu, Y., E. A. Rubin, and E. Zou (2021, April). What's Missing in Environmental (Self-)Monitoring: Evidence from Strategic Shutdowns of Pollution Monitors. Working Paper 28735, National Bureau of Economic Research. Series: Working Paper Series.

Sullivan, D. M. and A. Krupnick (2018). Using satellite data to fill the gaps in the US air pollution monitoring network. *Resources for the Future Working Paper*, 18–21.

Zou, E. Y. (2021, July). Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality. *American Economic Review 111*(7), 2101–2126.

8 Appendix

8.1 Improving Prediction of PM2.5 at the NAAQS Monitor Location

- Add short description of the implementation
- Add description of the NOAA NCEP NARR wind data – u-v components and extraction needed from layer.

$$EPA_{i,t} = \gamma_{i,0} + \sum_{j \in J_i} \sum_{k=1}^{7} \gamma_{i,j,k} PA_{j,t} \cdot Winddir_{i,t,k} + u_{i,t}$$

- Each EPA monitor i has it's own set of weights for the PA sensors around it.
- Analysis is done at the quarter level; suppressing quarter subscript.
- t is a unique hour within a given quarter.
- EPA monitor i at time t reads PM2.5 pollution $EPA_{i,t}$.
- For each EPA monitor i, there are J_i Purple Air monitors within a 5-mile radius.
- Purple Air monitor $j \in J_i$ at time t reads PM2.5 pollution $PA_{j,t}$.
- $Winddir_{i,t,k}$ is a wind direction indicator; 1 if the prevailing wind near station i at time t is in the k^{th} bucket (of 8 buckets).
- I will also estimate a version with wind speed interacted in the sum. This could allow for sensors further away to have more predictive power when the winds are strong.
- This regression could be run as a LASSO first to determine which of the interactions for each PurpleAir sensor have the most predictive power.

8.2 Plots for Other California Hourly NAAQS Monitors

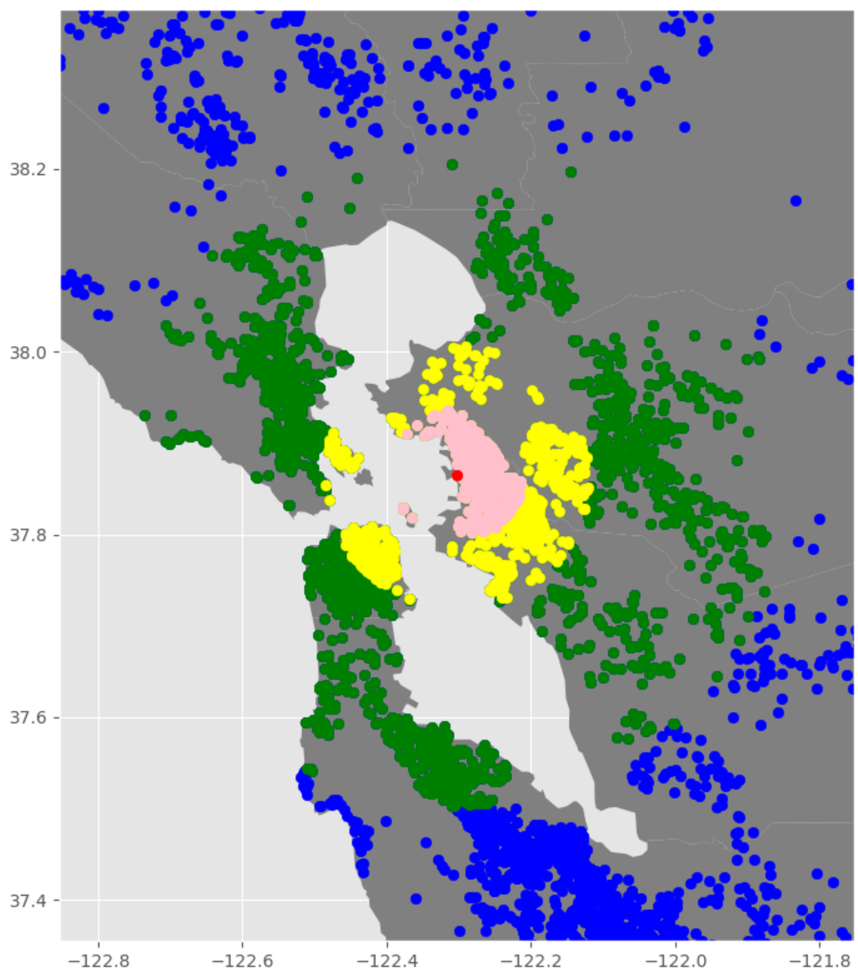

Figure 7: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0013 in county 001 (FIPS code).

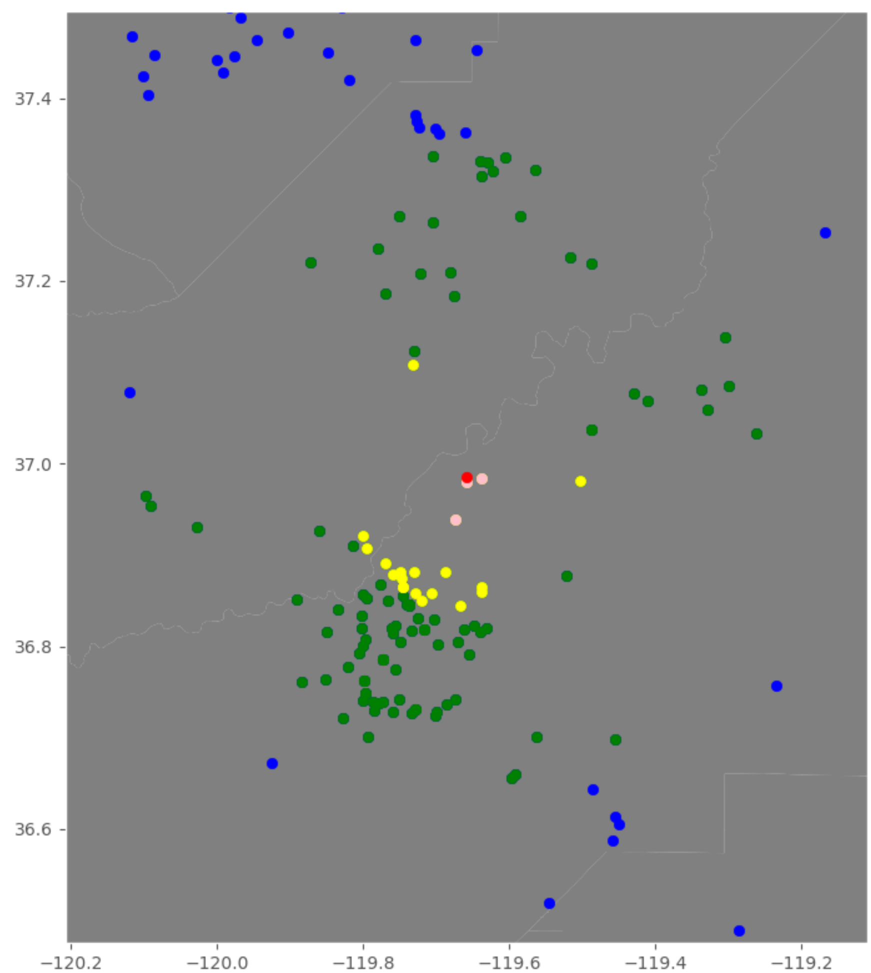

Figure 8: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0500 in county 019 (FIPS code).

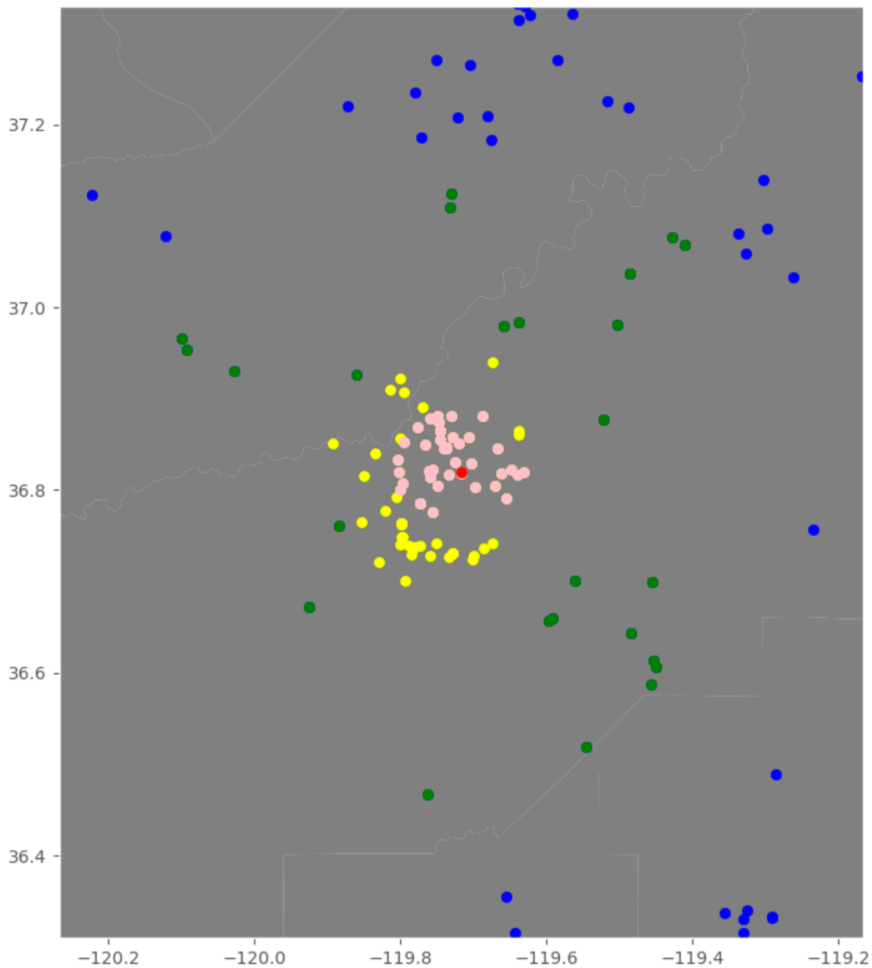

Figure 9: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 5001 in county 019 (FIPS code).

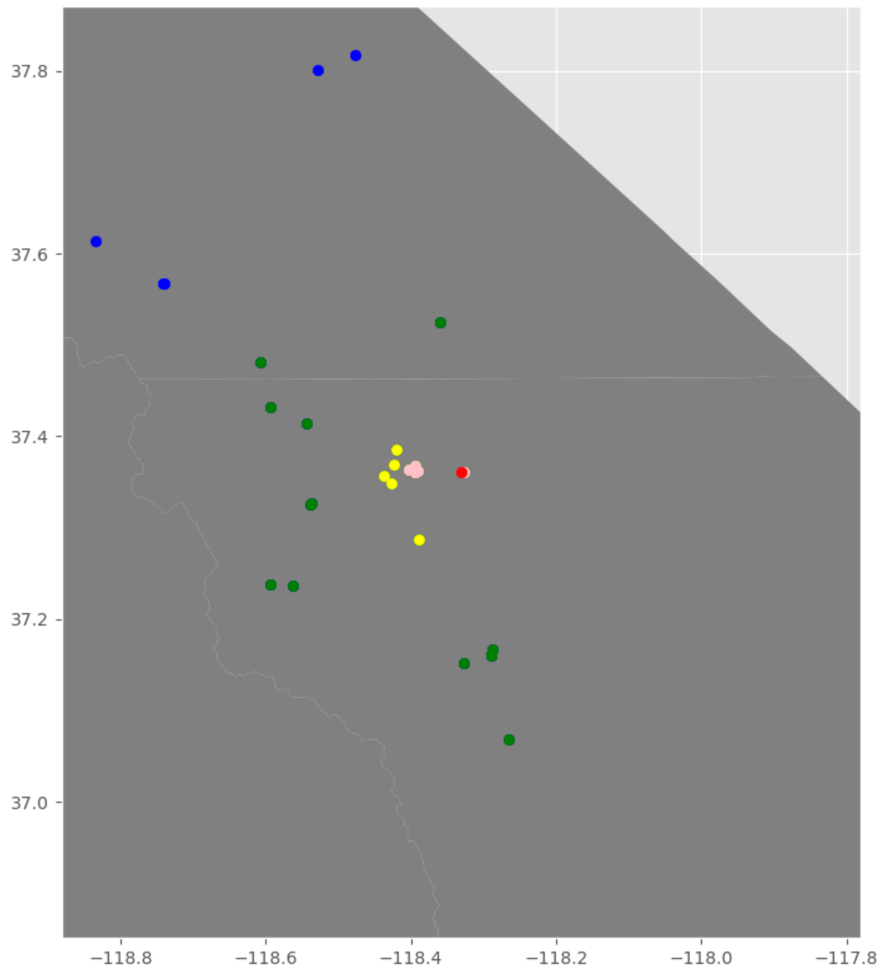

Figure 10: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0002 in county 027 (FIPS code).

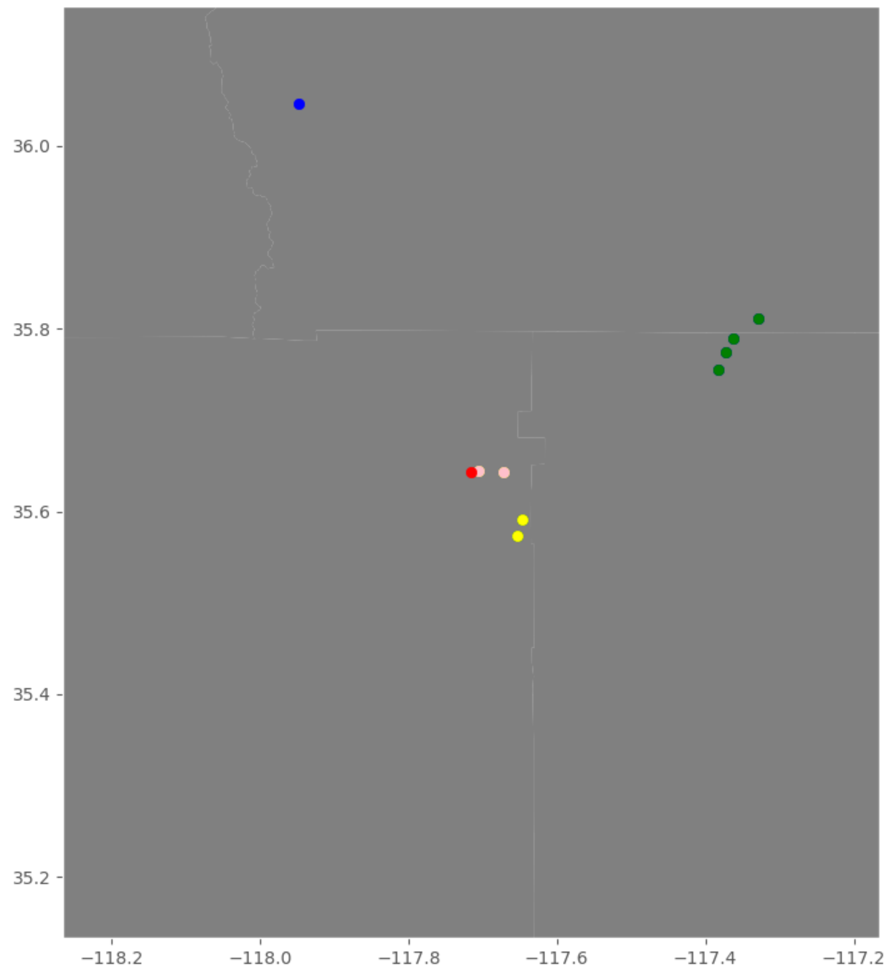

Figure 11: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0018 in county 029 (FIPS code).

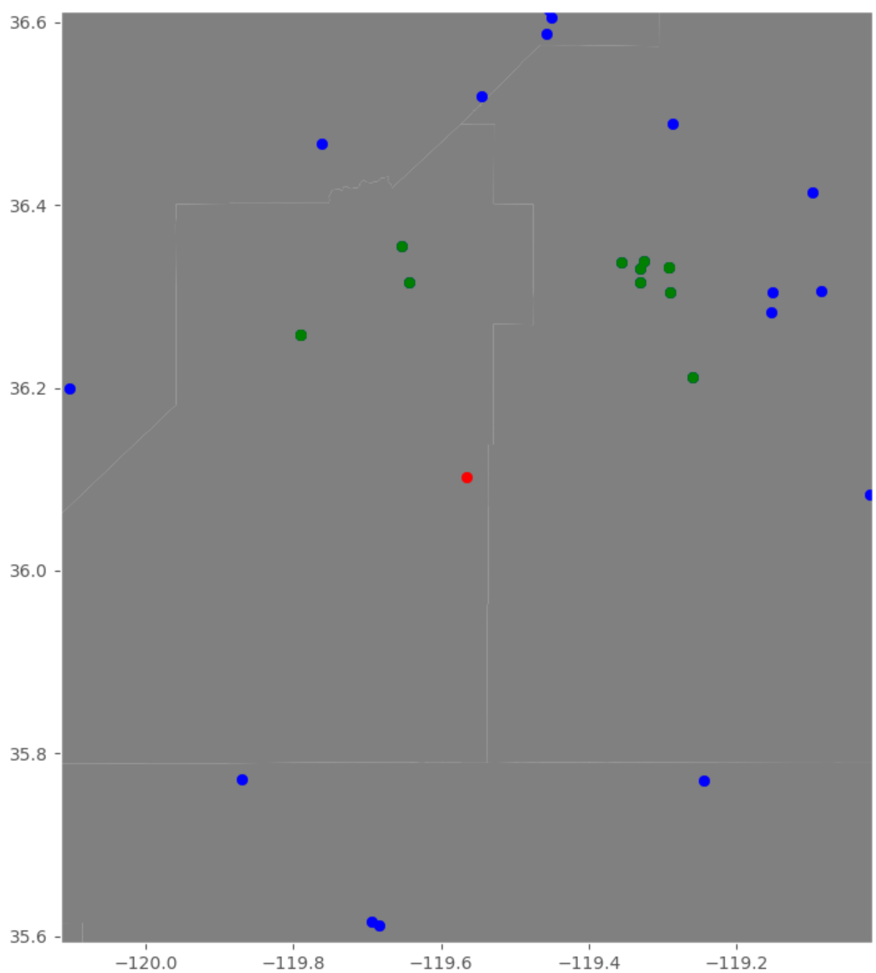

Figure 12: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0004 in county 031 (FIPS code).

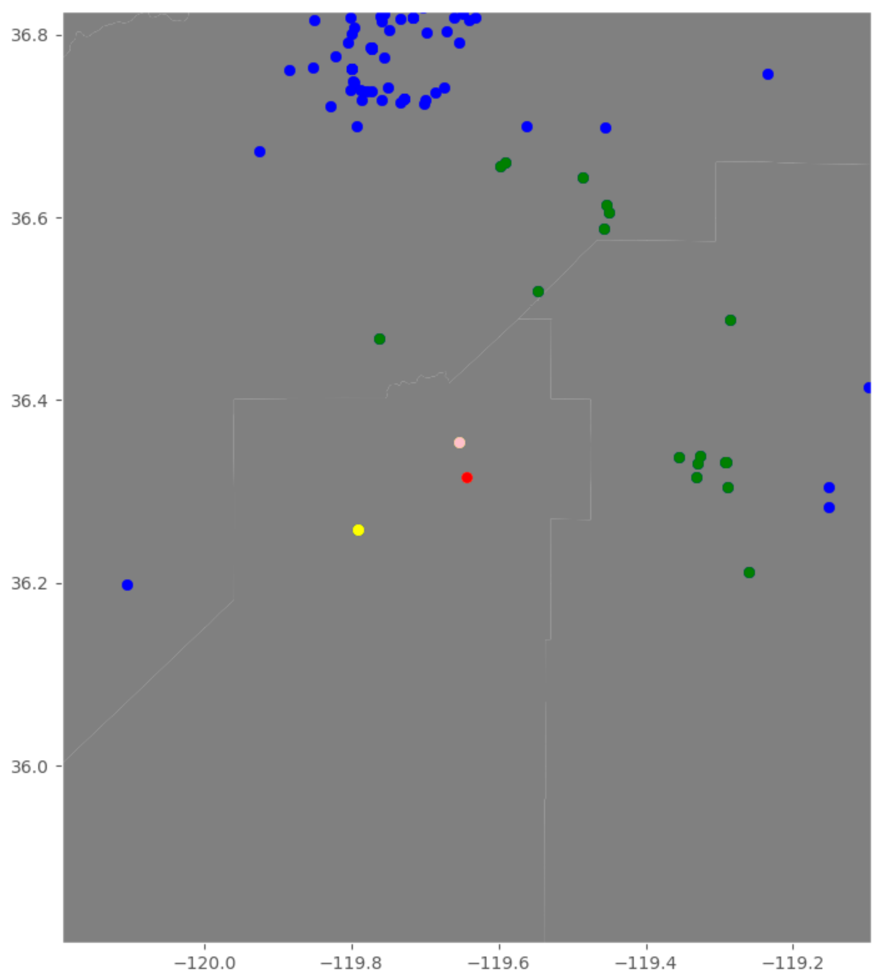

Figure 13: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 1004 in county 031 (FIPS code).

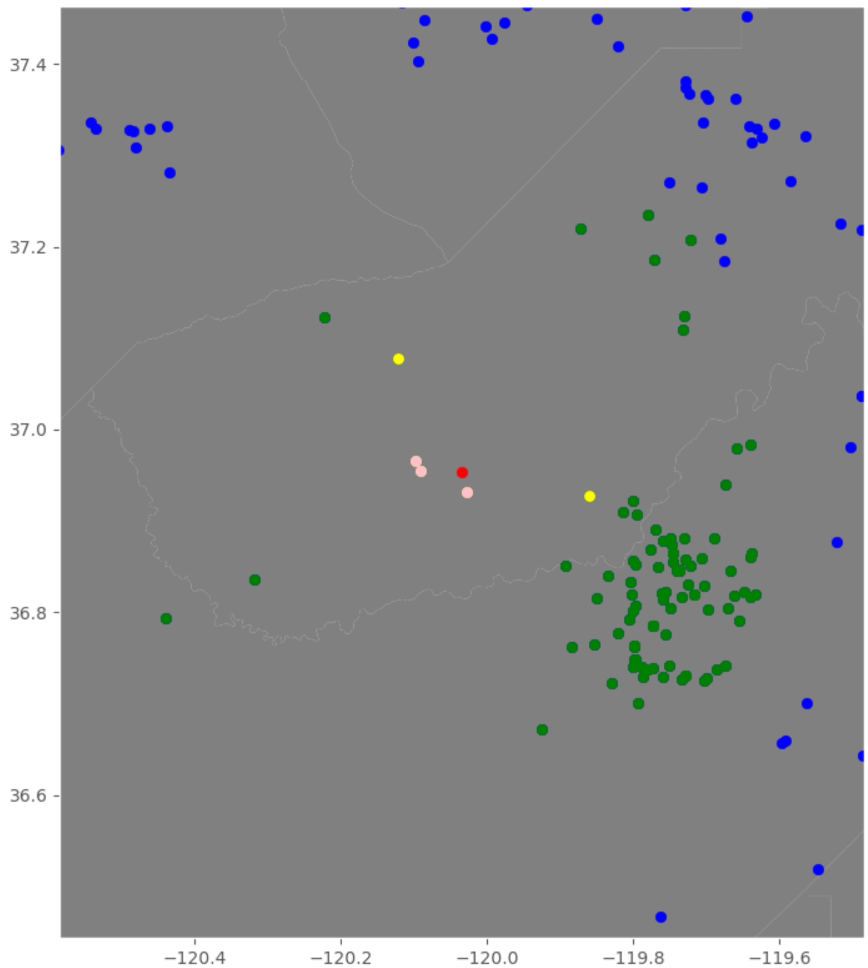

Figure 14: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 2010 in county 039 (FIPS code).

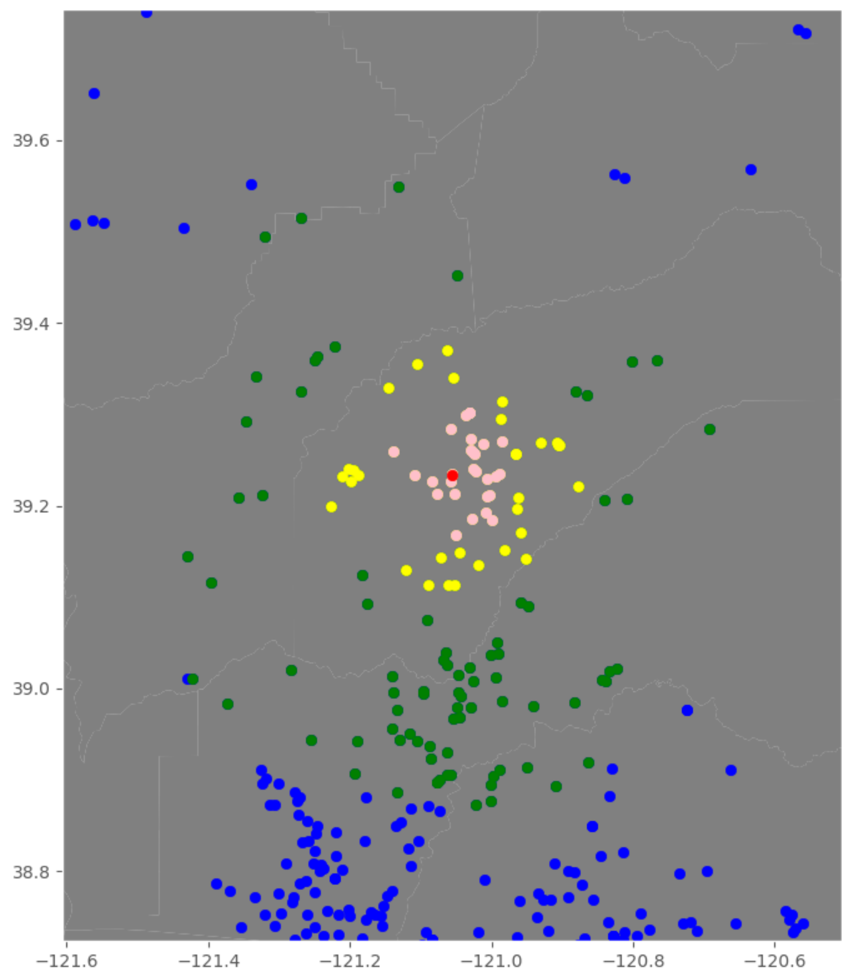

Figure 15: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0005 in county 057 (FIPS code).

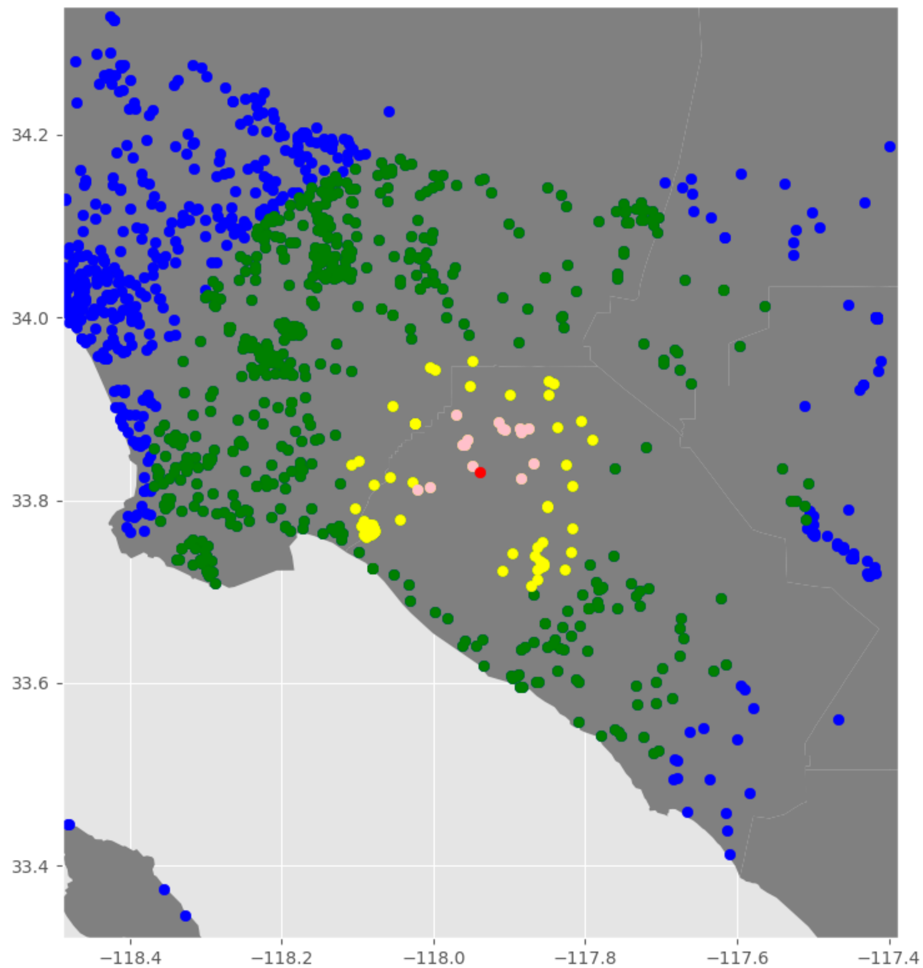

Figure 16: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0007 in county 059 (FIPS code).

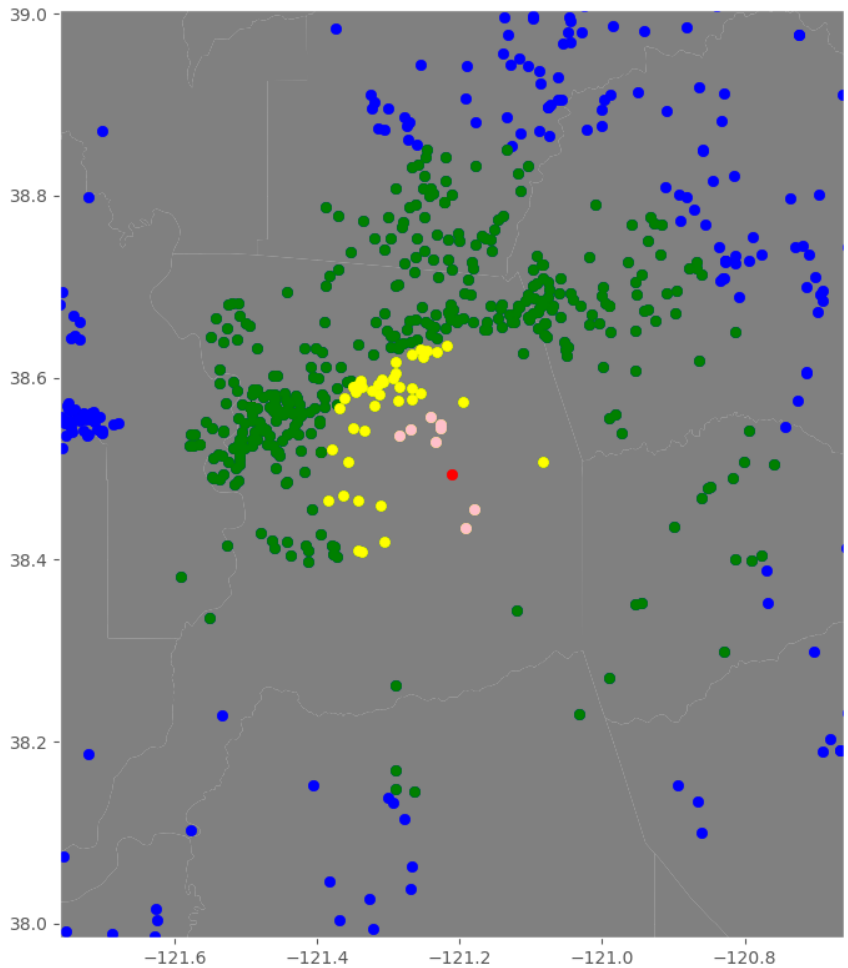

Figure 17: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 5003 in county 067 (FIPS code).

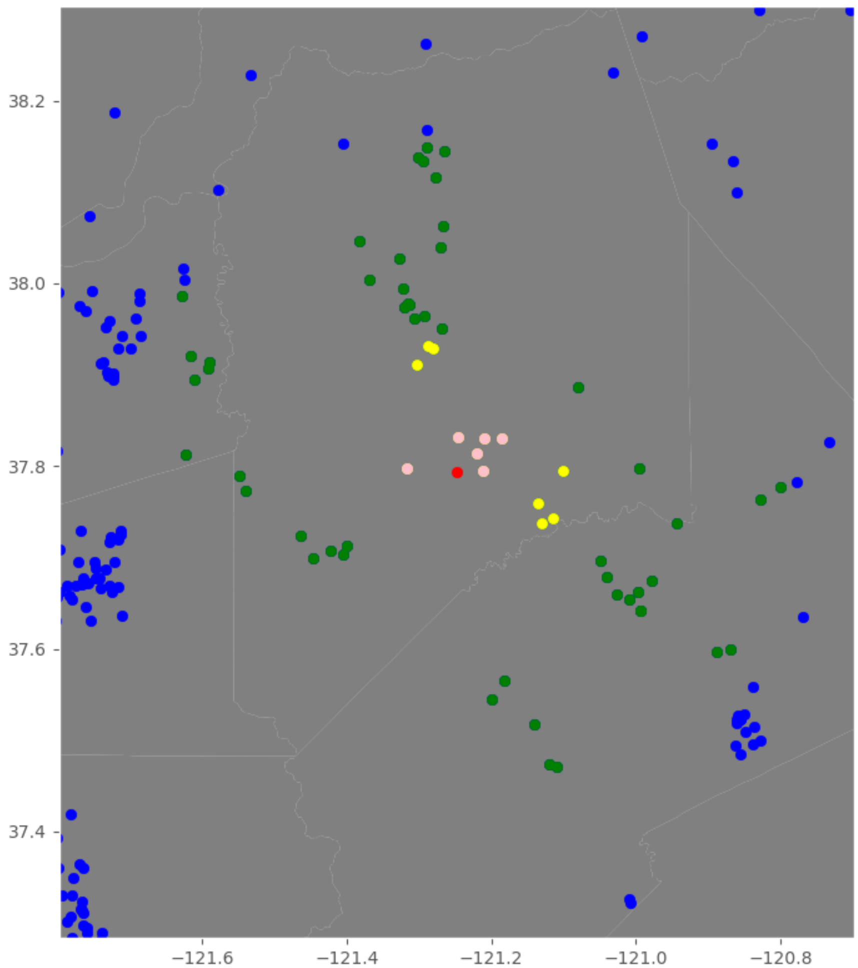

Figure 18: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 2010 in county 077 (FIPS code).

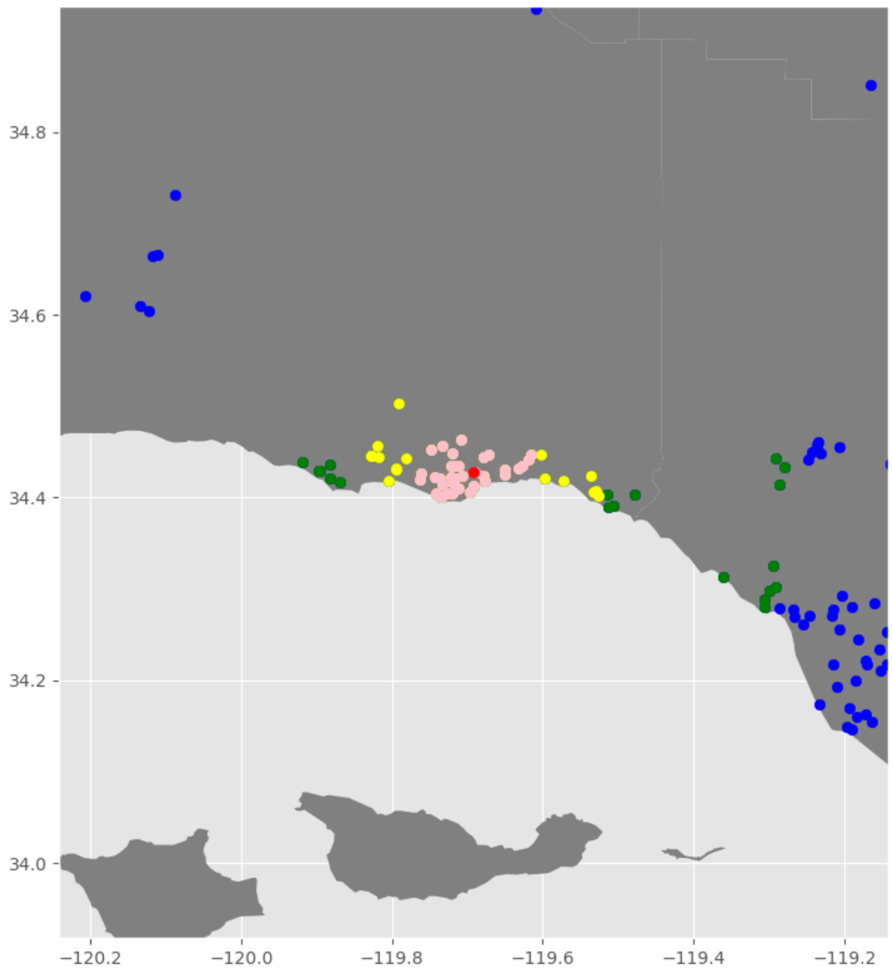

Figure 19: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0011 in county 083 (FIPS code).

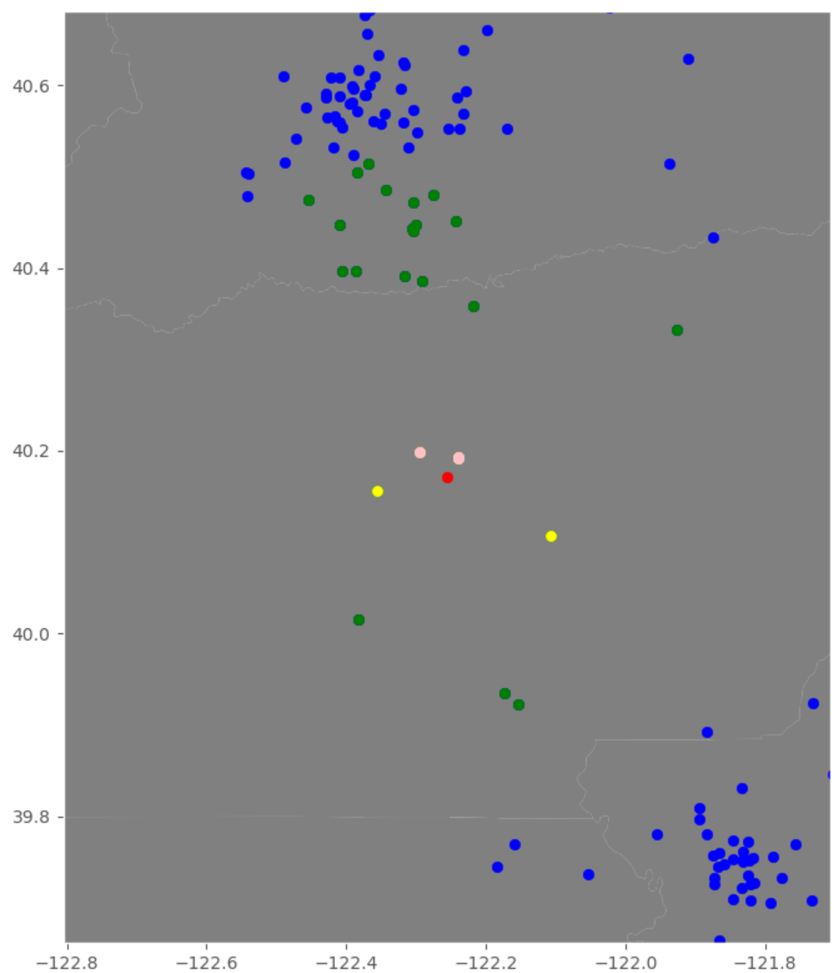

Figure 20: Map of EPA NAAQS-primary monitoring station (red) surrounded by PurpleAir monitors within 5-mile (pink), 10-mile (yellow), and 25-mile (green) radii.This preliminary analysis uses the PurpleAir sensors within 5 miles (pink markers). This monitor is at site 0007 in county 103 (FIPS code).

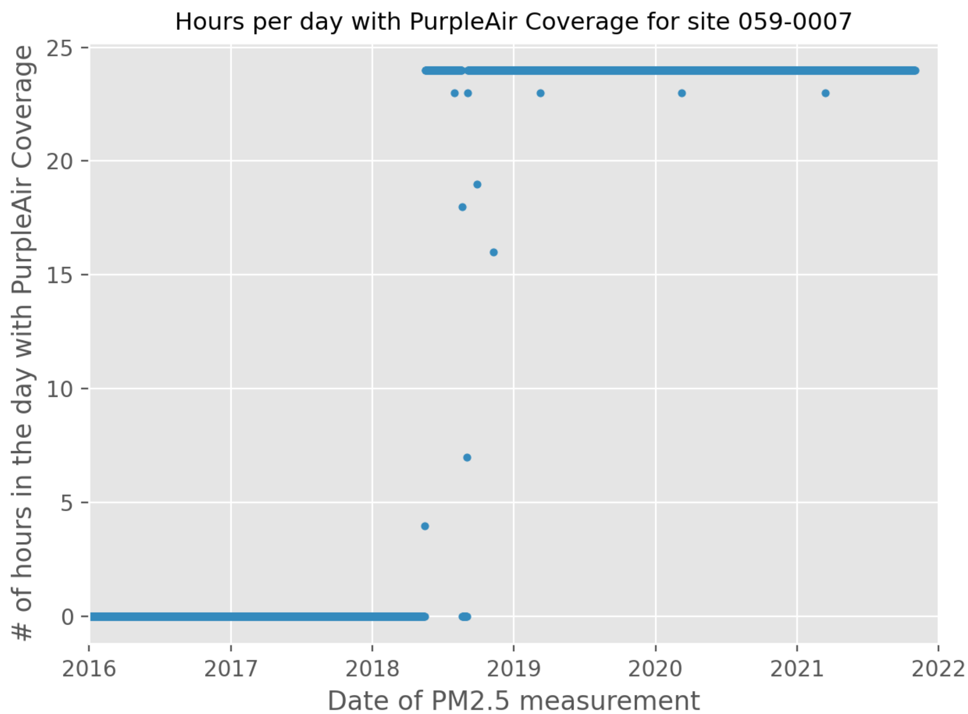

Figure 21: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0007 in county 059 (FIPS code).

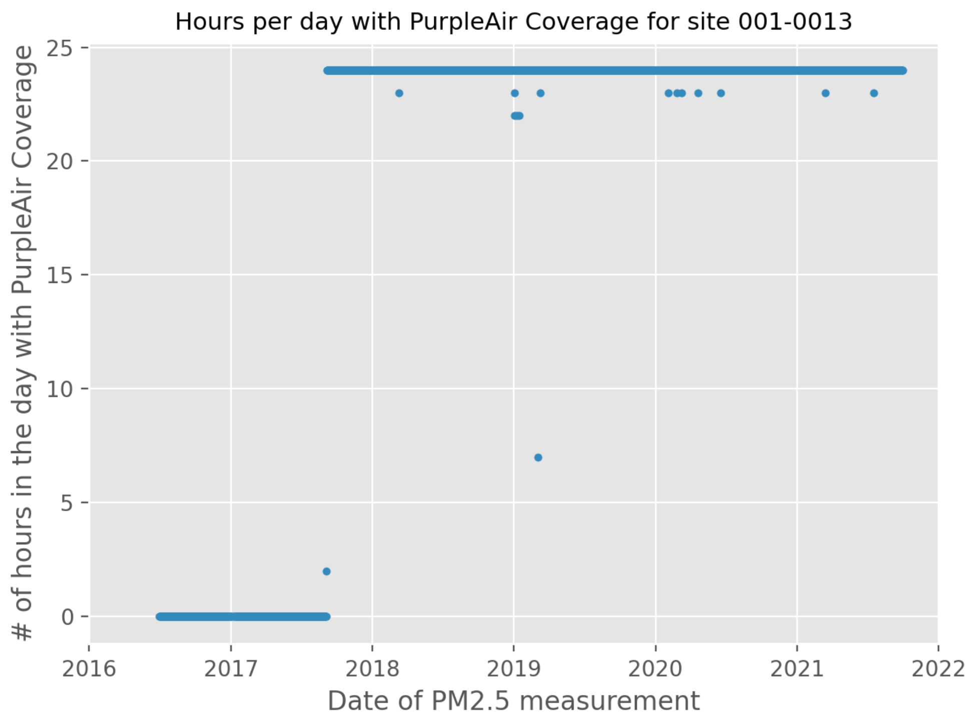

Figure 22: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0013 in county 001 (FIPS code).

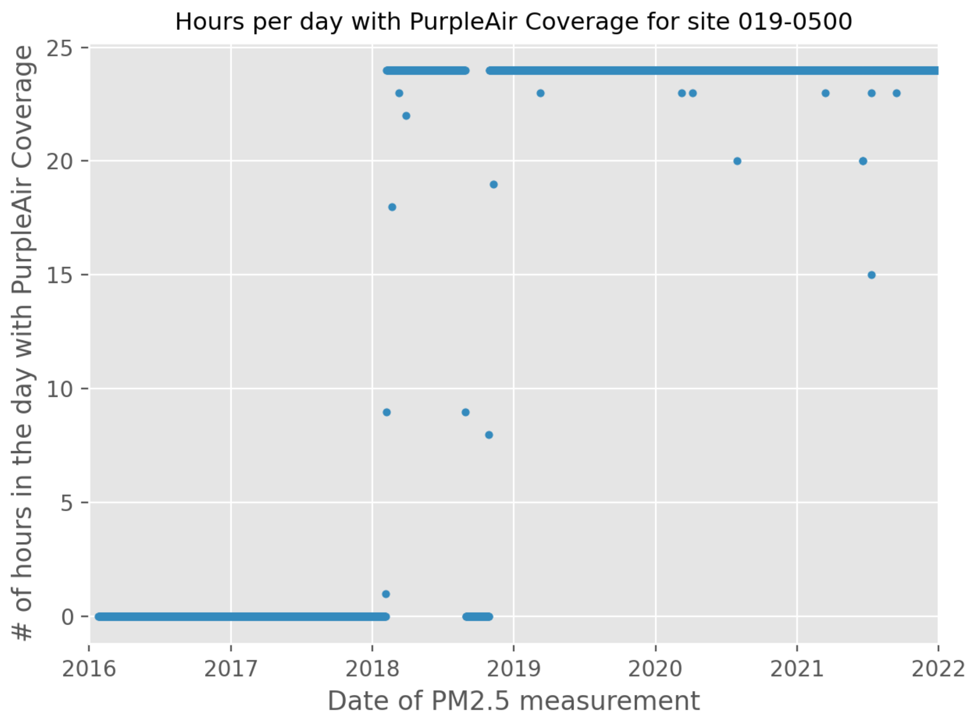

Figure 23: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0500 in county 019 (FIPS code).

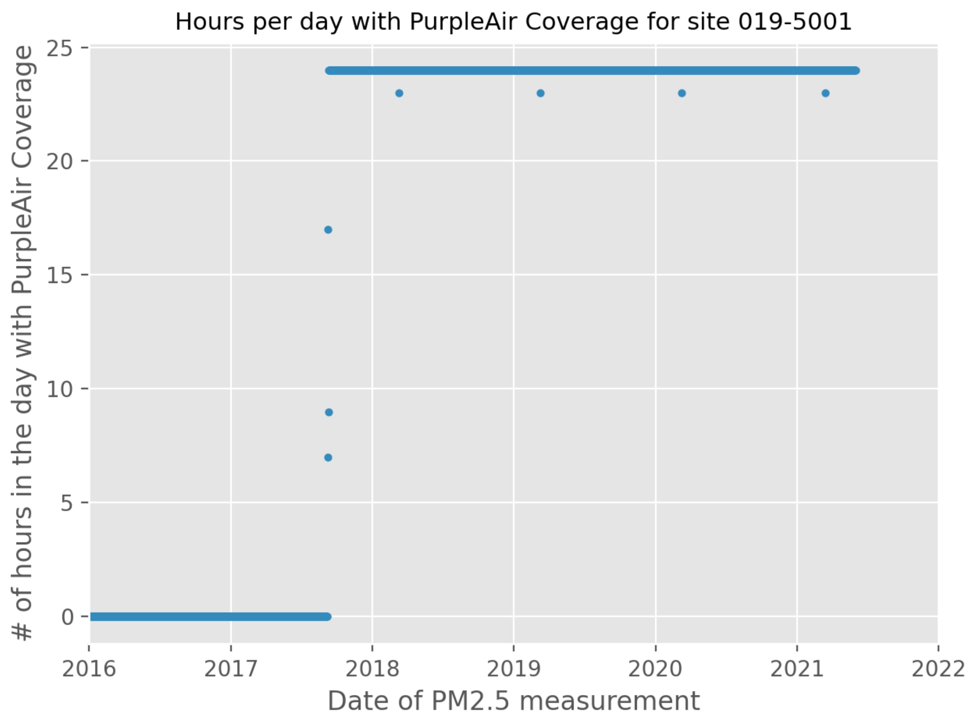

Figure 24: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 5001 in county 019 (FIPS code).

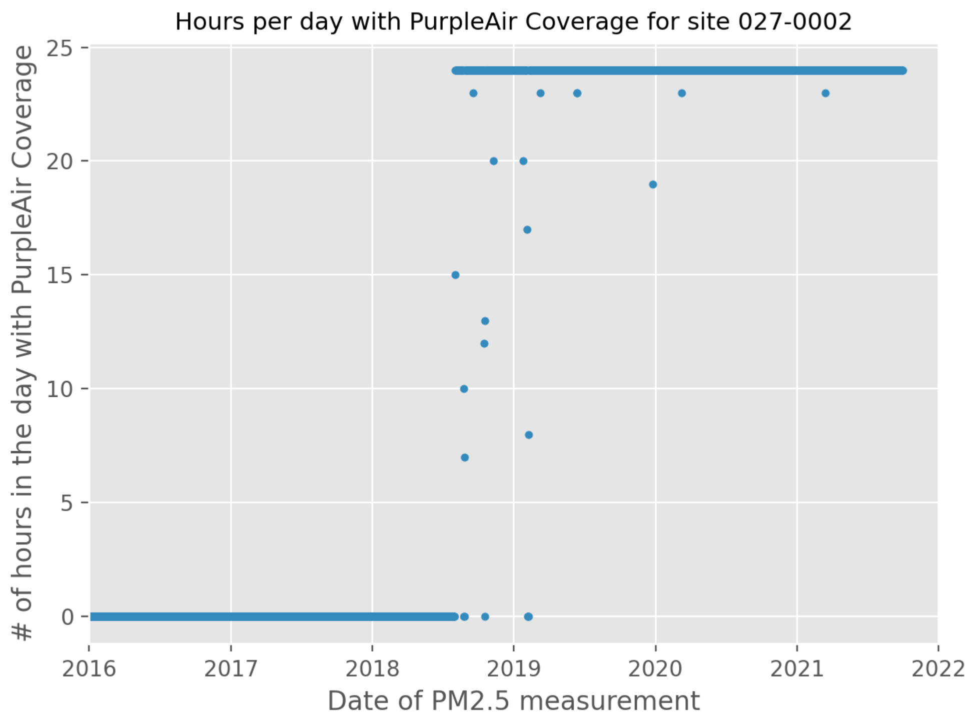

Figure 25: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0002 in county 027 (FIPS code).

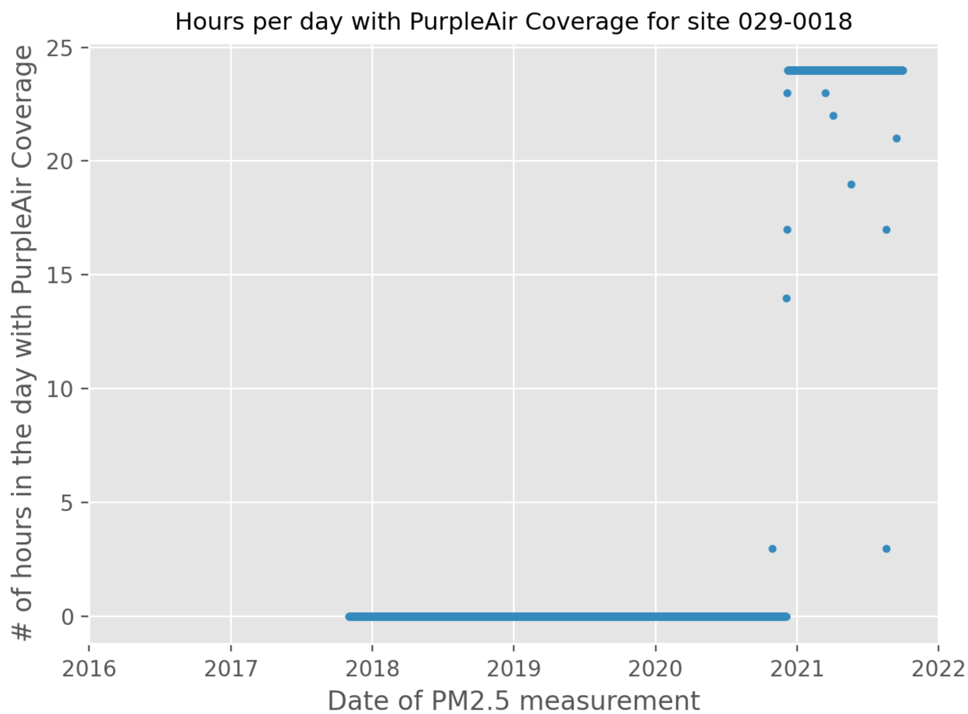

Figure 26: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0018 in county 029 (FIPS code).

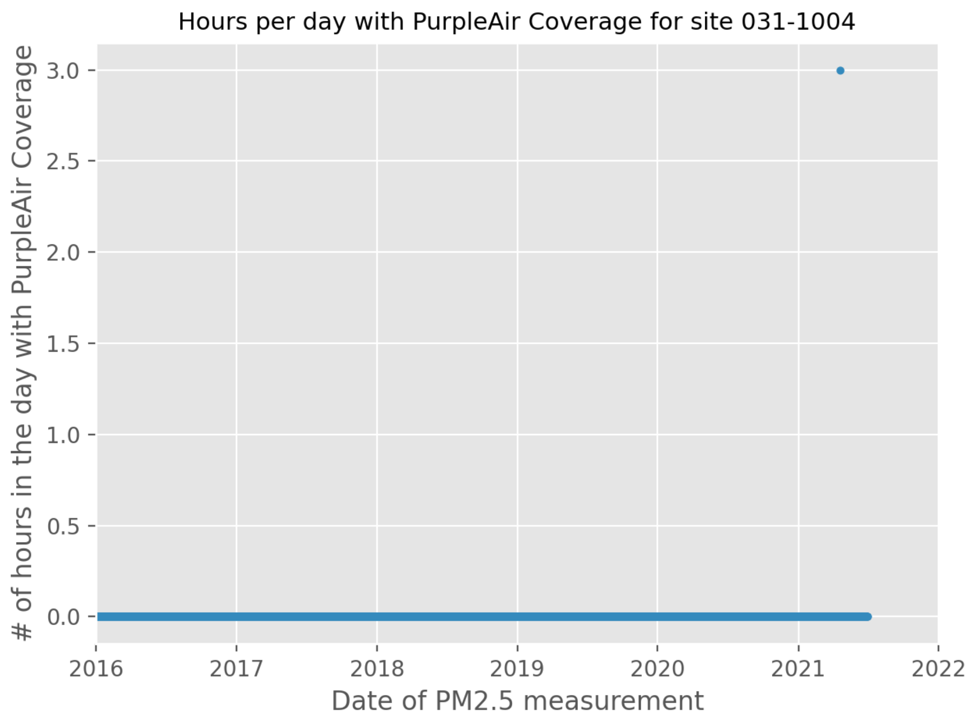

Figure 27: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 1004 in county 031 (FIPS code).

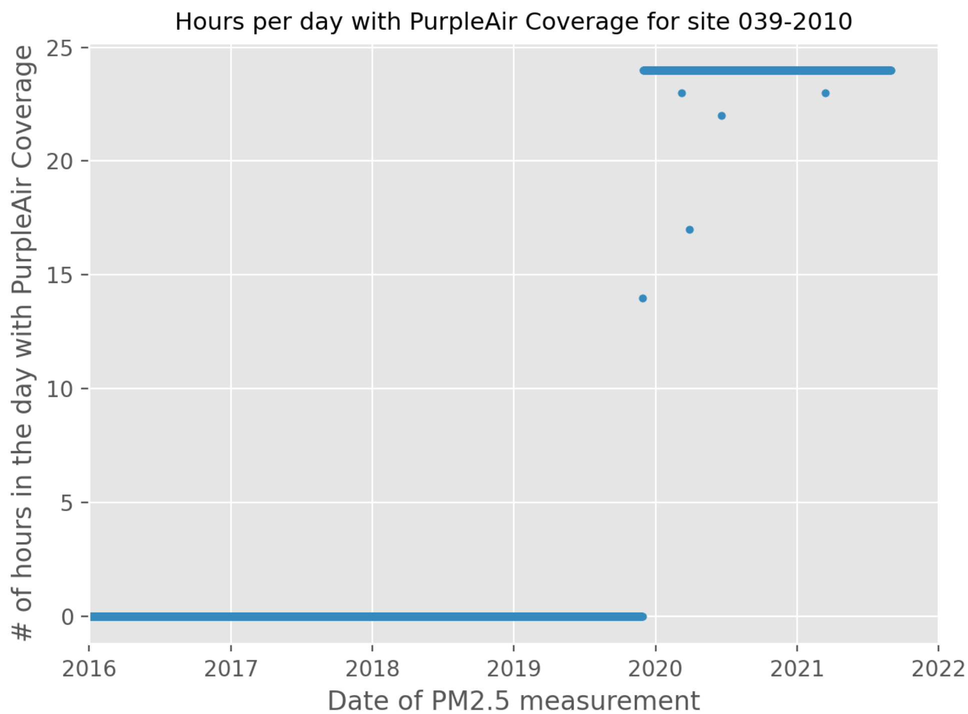

Figure 28: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 2010 in county 039 (FIPS code).

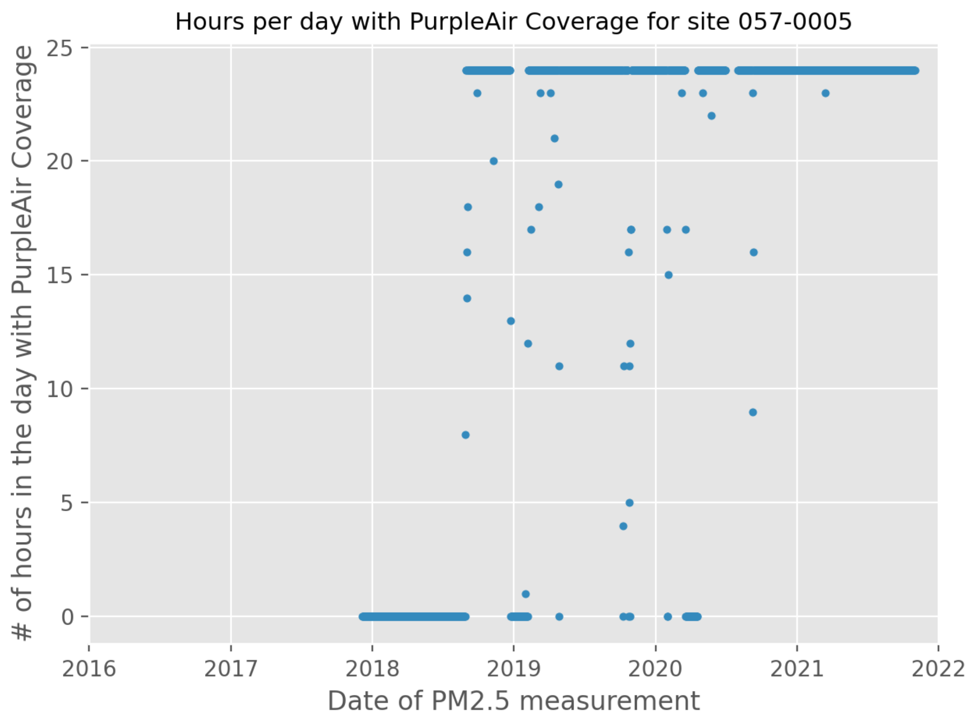

Figure 29: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0005 in county 057 (FIPS code).

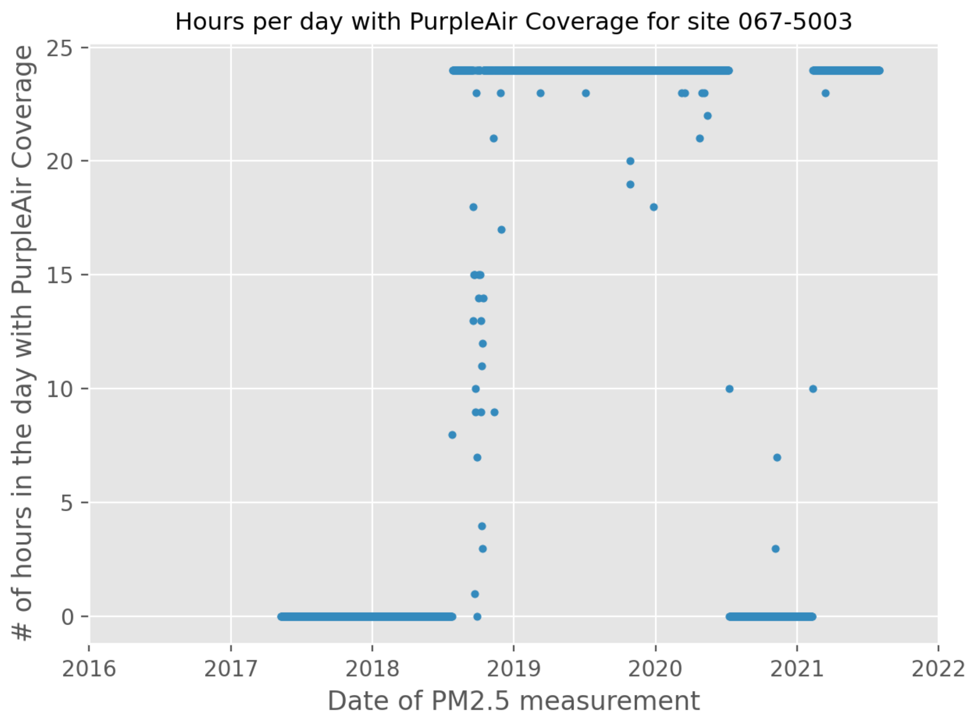

Figure 30: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 5003 in county 067 (FIPS code).

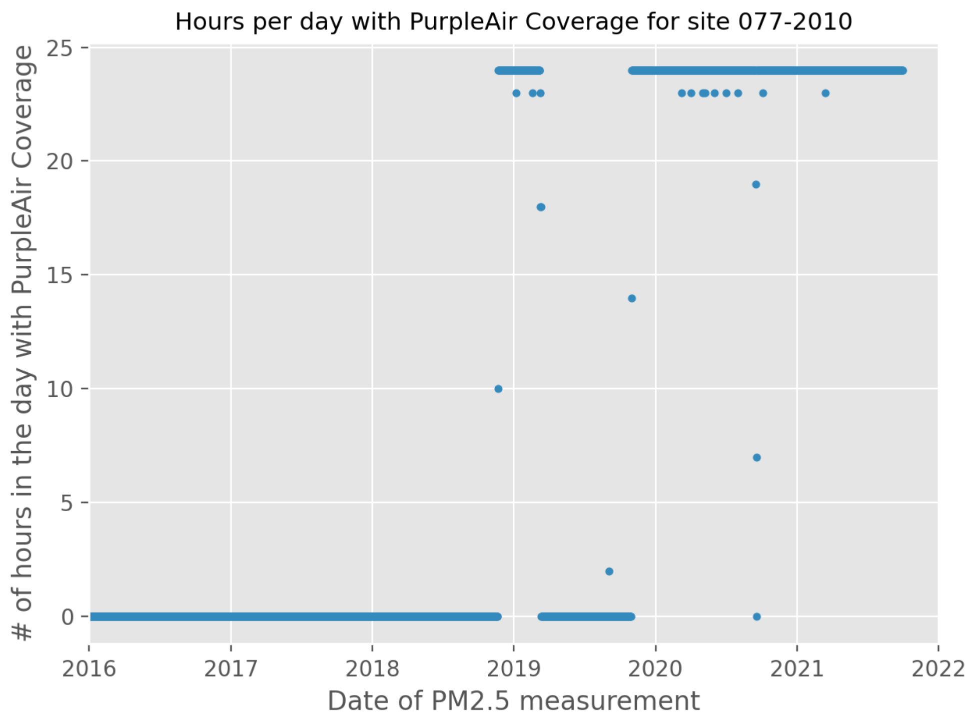

Figure 31: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 2010 in county 077 (FIPS code).

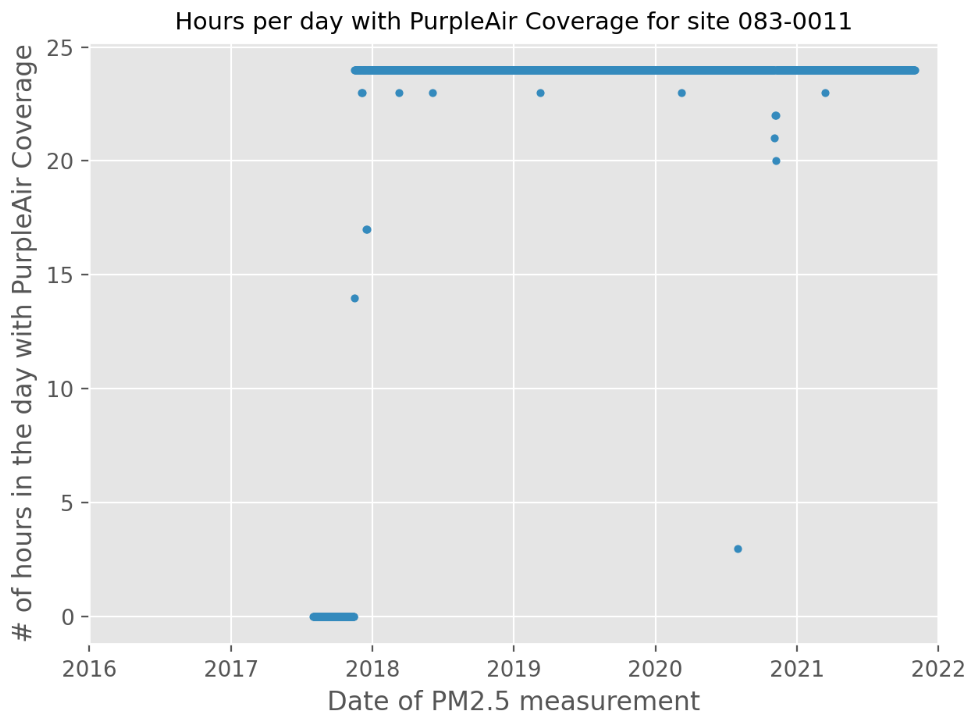

Figure 32: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0011 in county 083 (FIPS code).

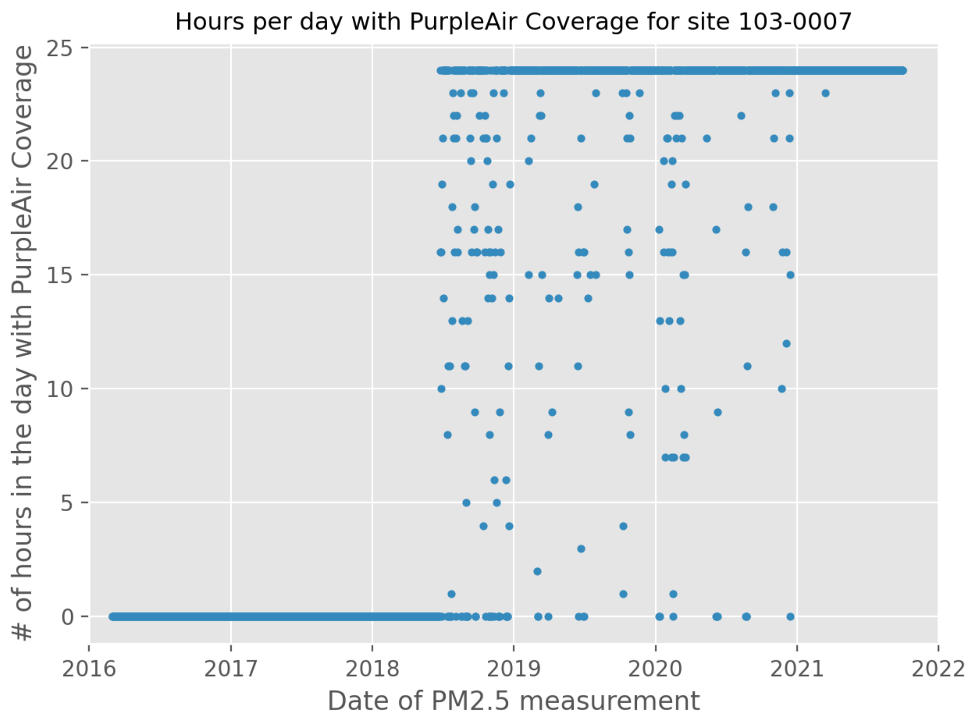

Figure 33: Scatter plot indicating the number of hours in each day that this NAAQS monitor has PurpleAir coverage. An hour has PurpleAir coverage if there are any PurpleAir sensor readings within the 5-mile radius of the monitor site for that hour. The weighted average is calculated for that hour using all the available PurpleAir readings within 5 miles. This monitor is at site 0007 in county 103 (FIPS code).

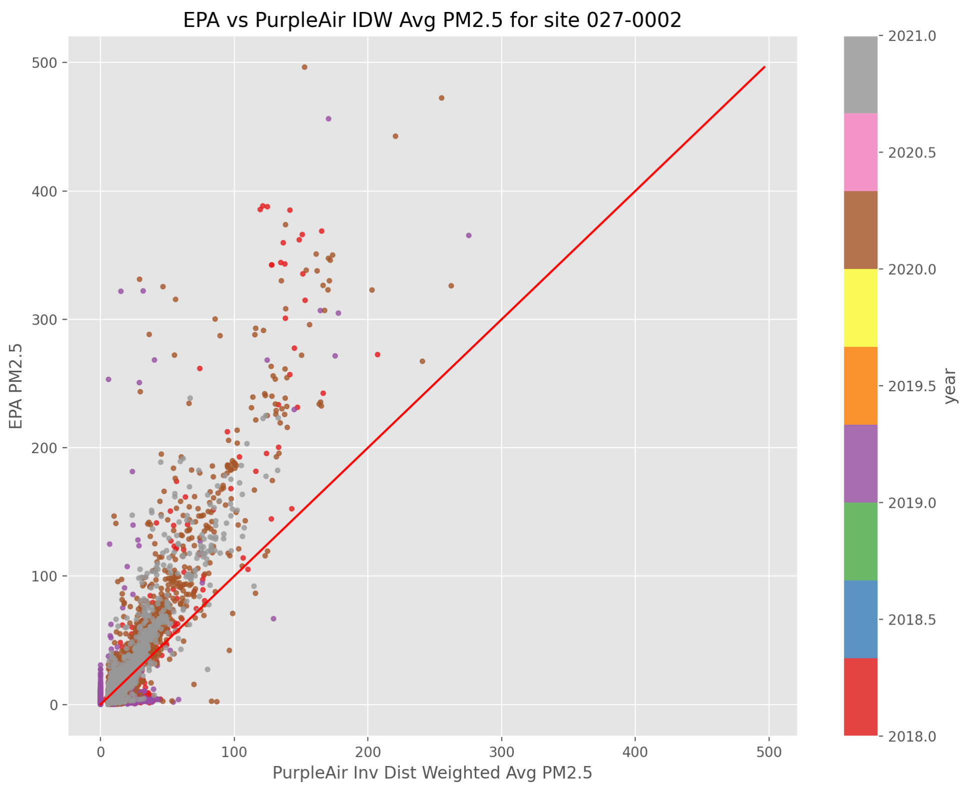

Figure 34: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0002 in county 027 (FIPS code).

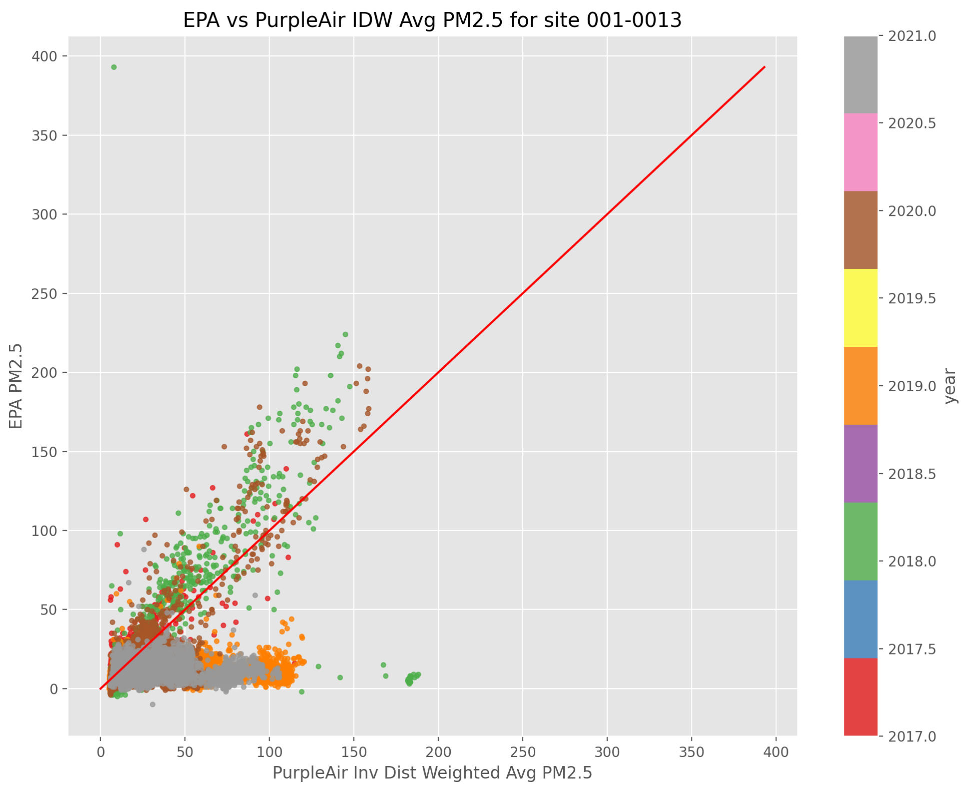

Figure 35: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0013 in county 001 (FIPS code).

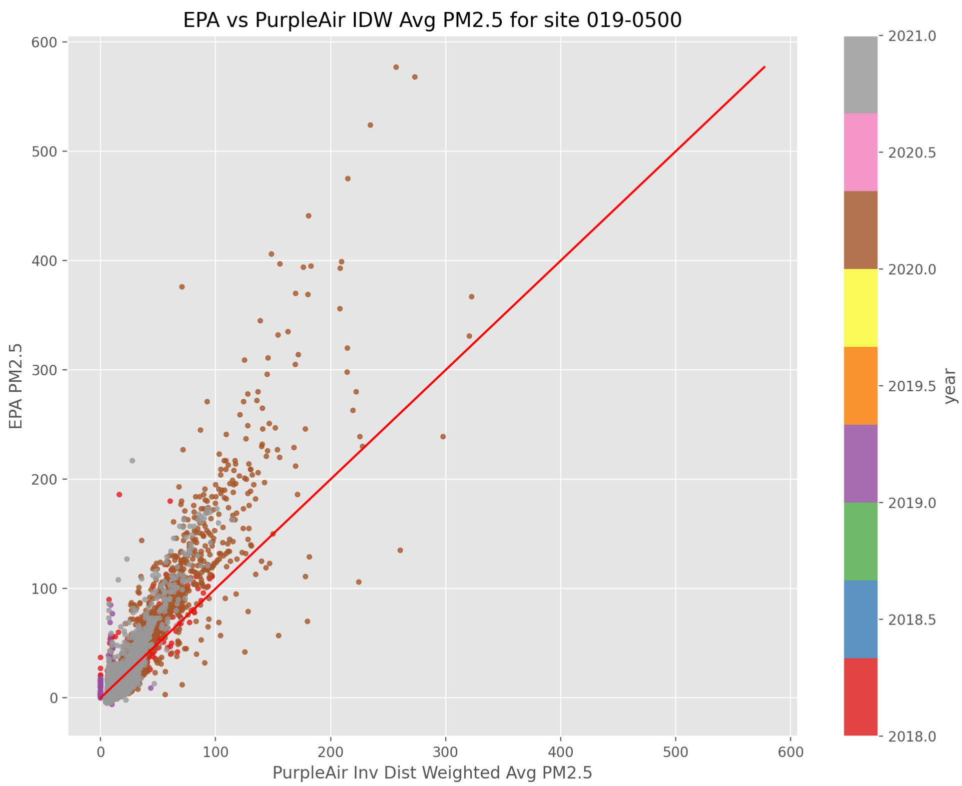

Figure 36: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0500 in county 019 (FIPS code).

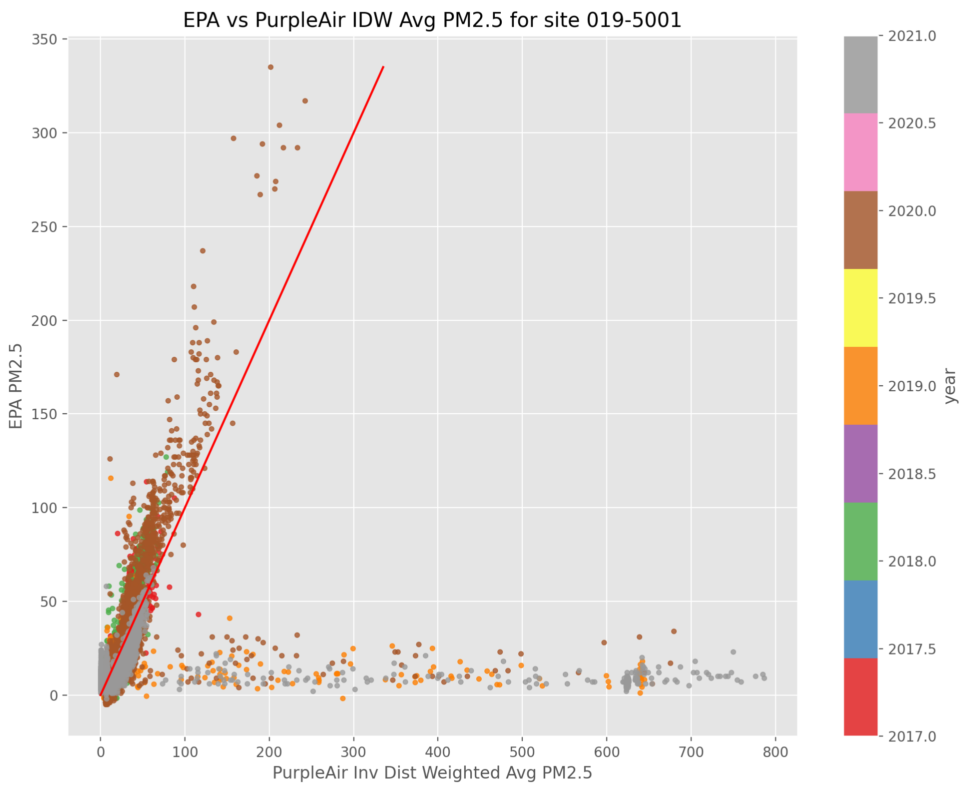

Figure 37: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 5001 in county 019 (FIPS code).

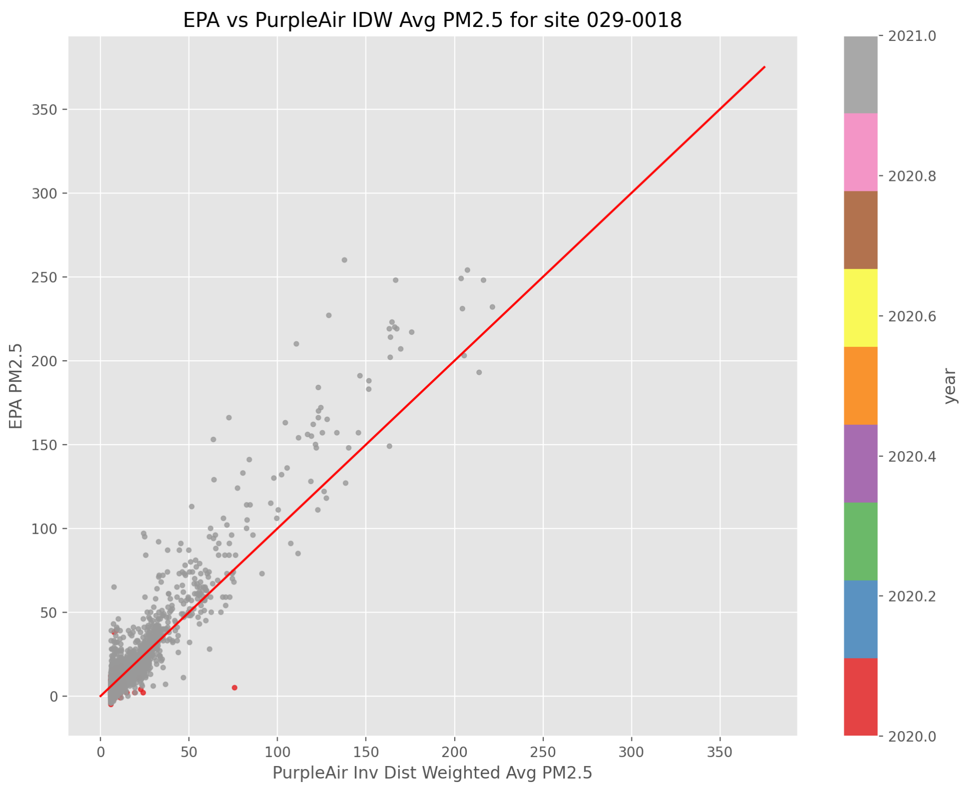

Figure 38: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0018 in county 029 (FIPS code).

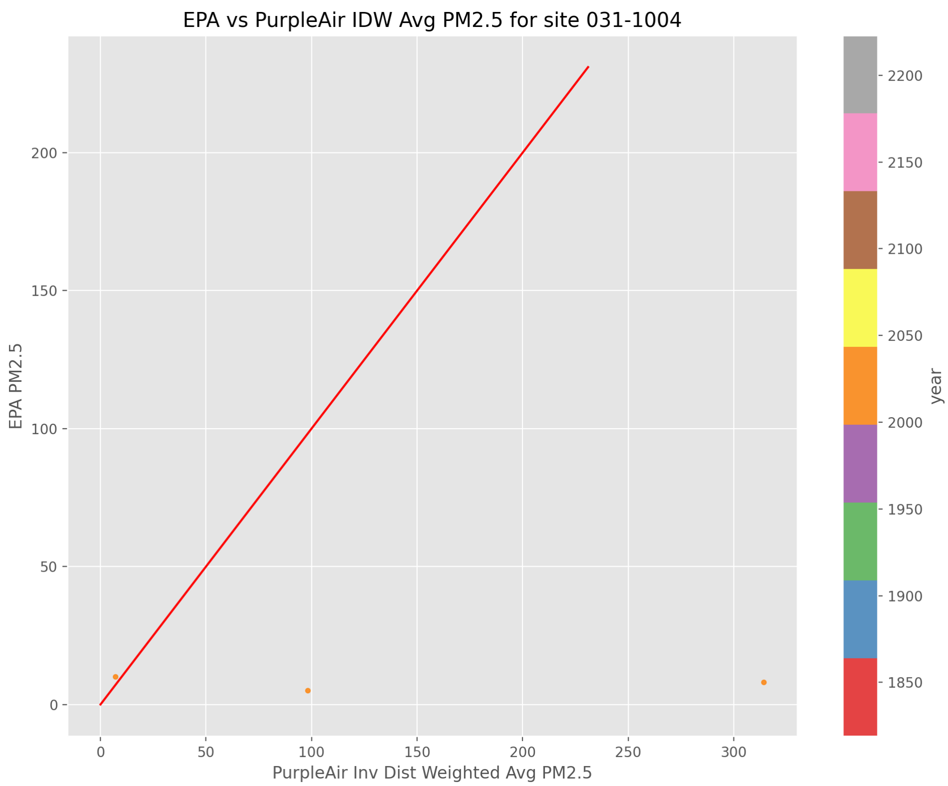

Figure 39: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 1004 in county 031 (FIPS code).

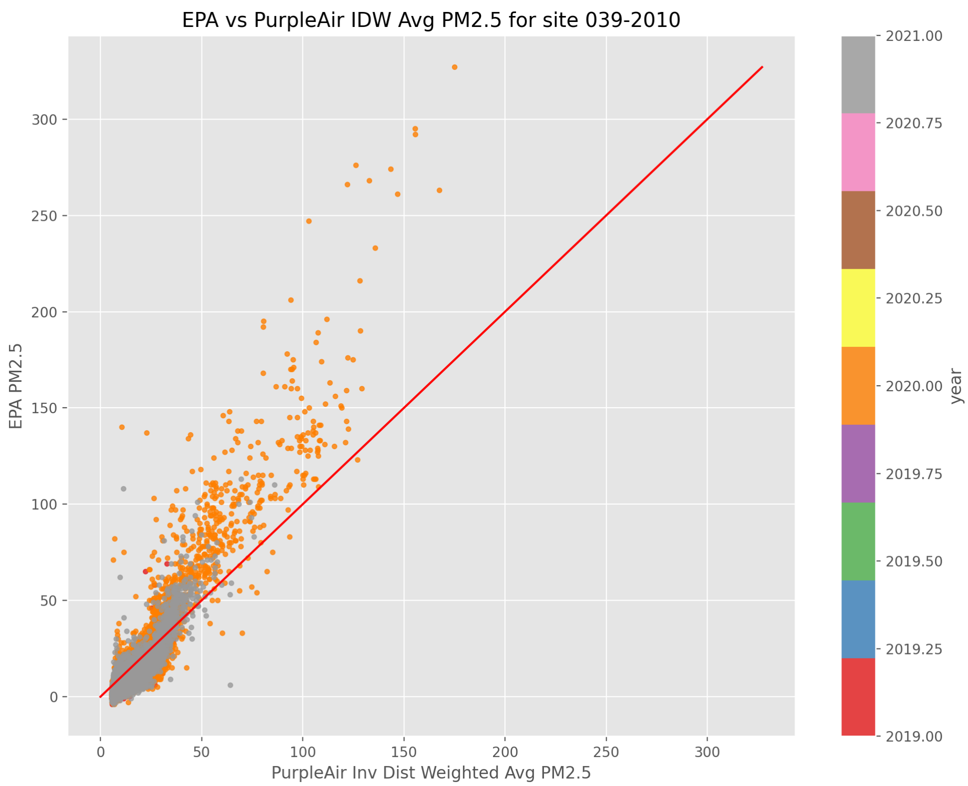

Figure 40: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 2010 in county 039 (FIPS code).

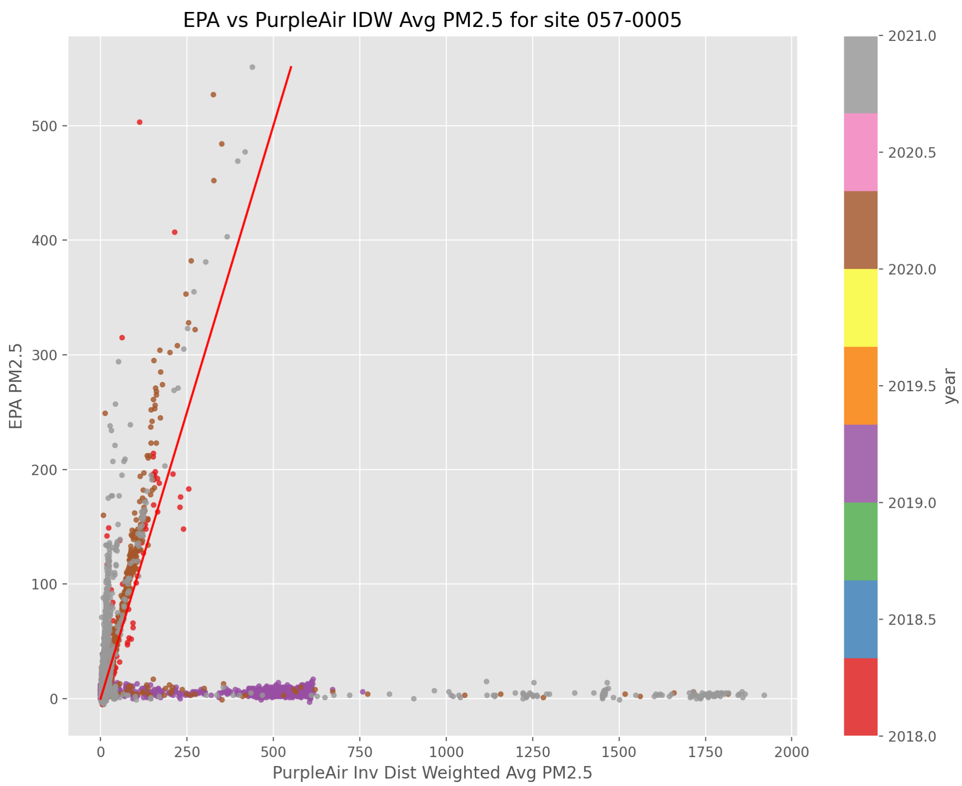

Figure 41: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0005 in county 057 (FIPS code).

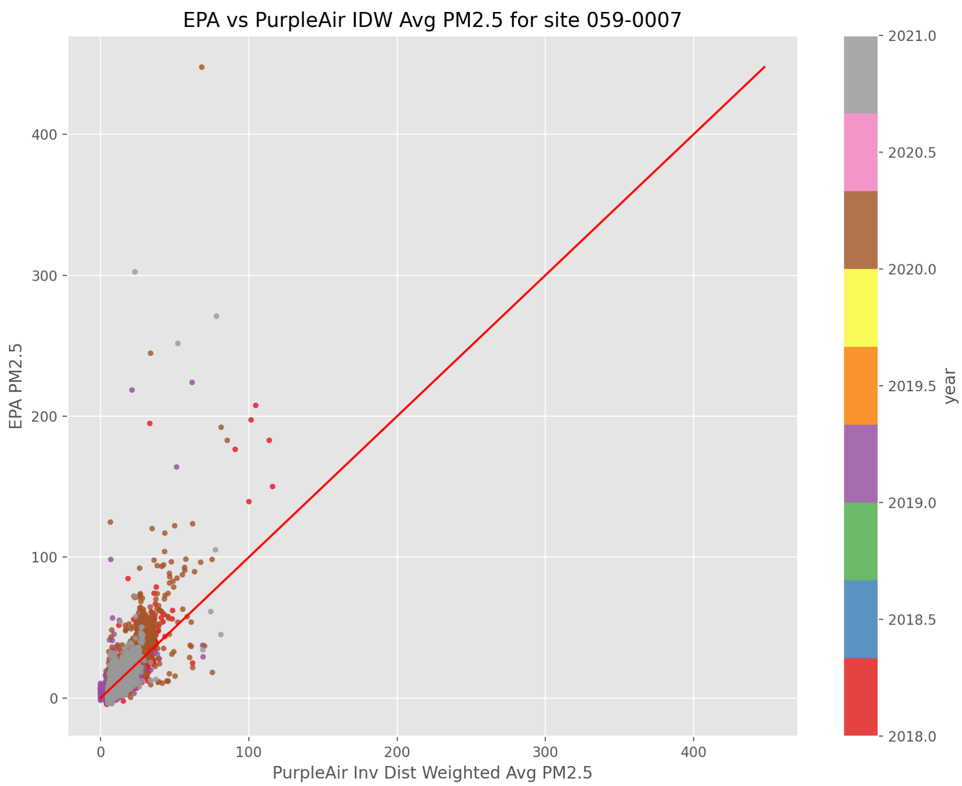

Figure 42: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0007 in county 059 (FIPS code).

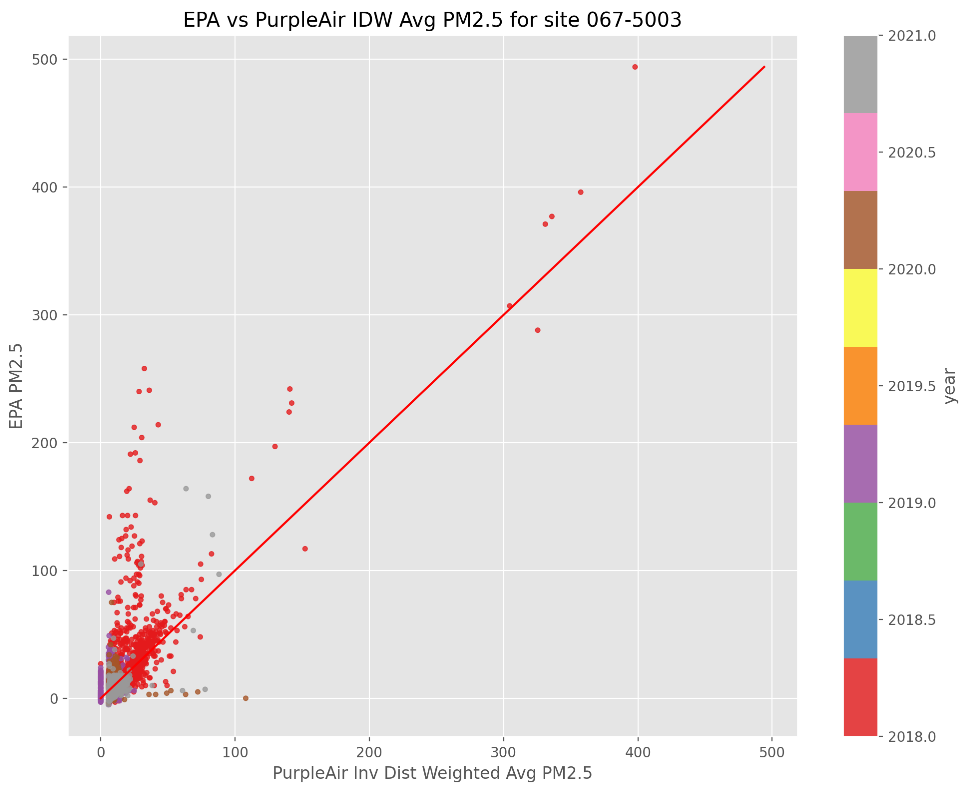

Figure 43: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 5003 in county 067 (FIPS code).

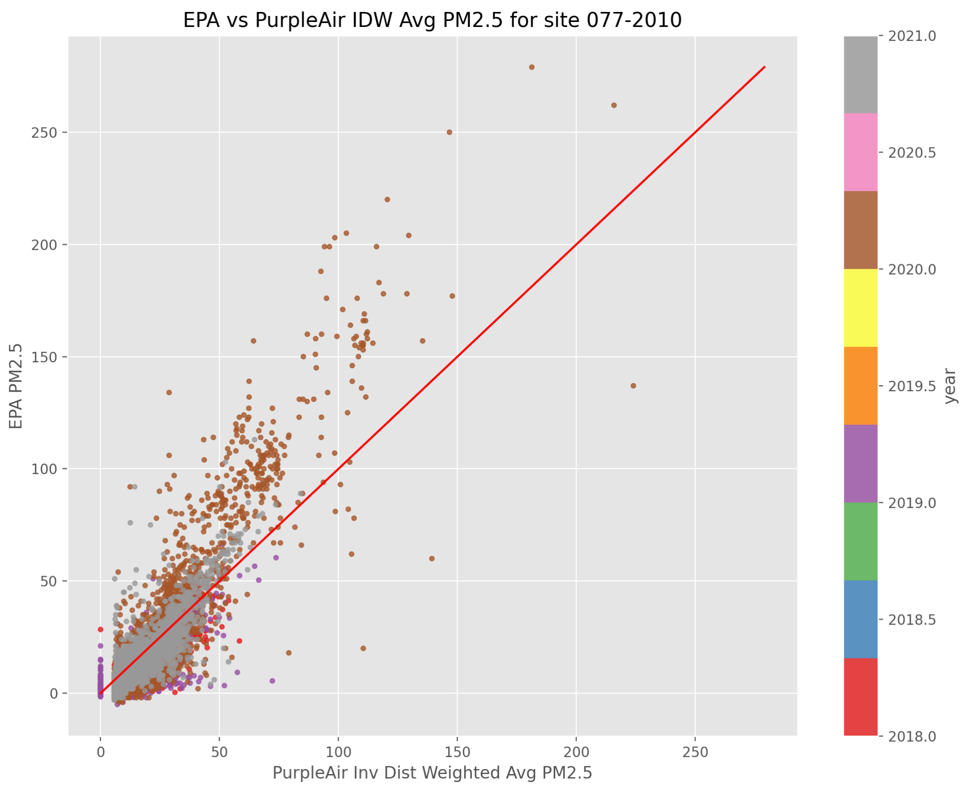

Figure 44: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 2010 in county 077 (FIPS code).

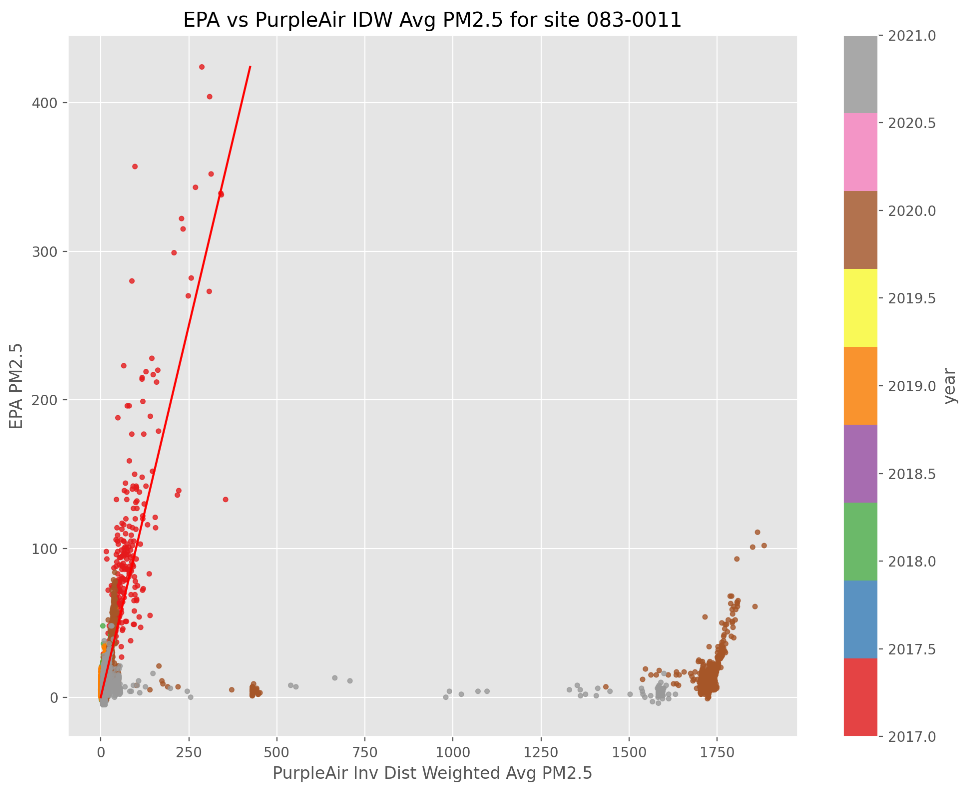

Figure 45: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0011 in county 083 (FIPS code).

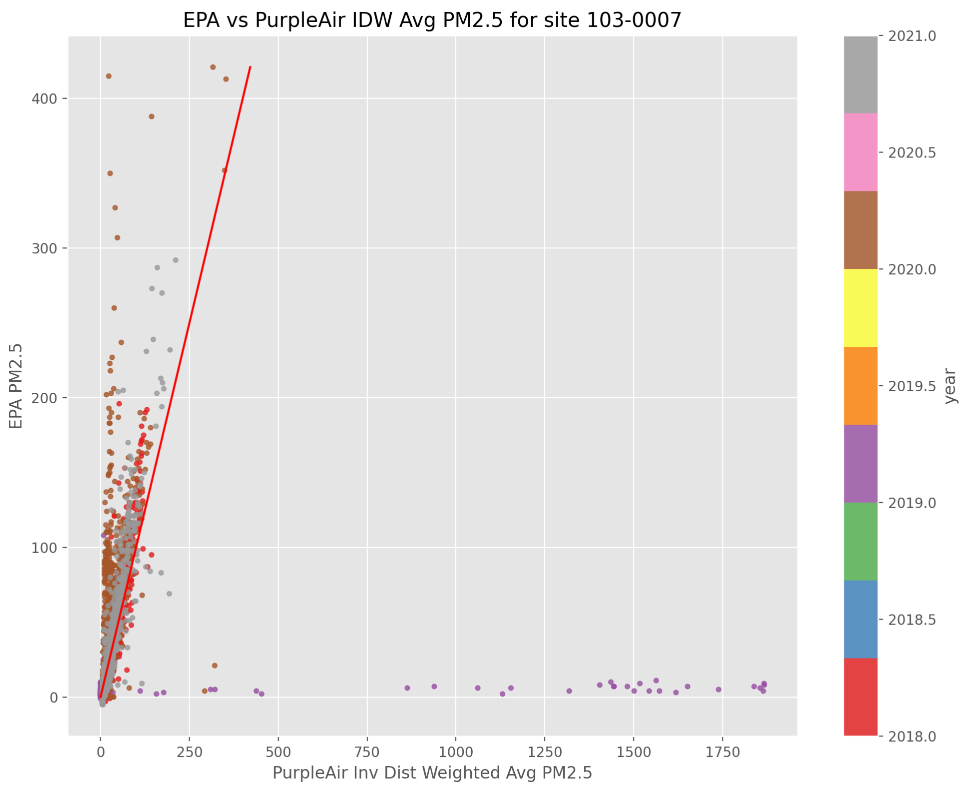

Figure 46: Scatter plot comparing reported hourly PM2.5 measurements: the x-axis represents the IDW-weighted average of PurpleAir measurements, the y-axis represents reported NAAQS-primary monitor measurements. The red line is a 45° line, representing perfect correlation between the PurpleAir average and the NAAQS-primary monitor. This monitor is at site 0007 in county 103 (FIPS code).

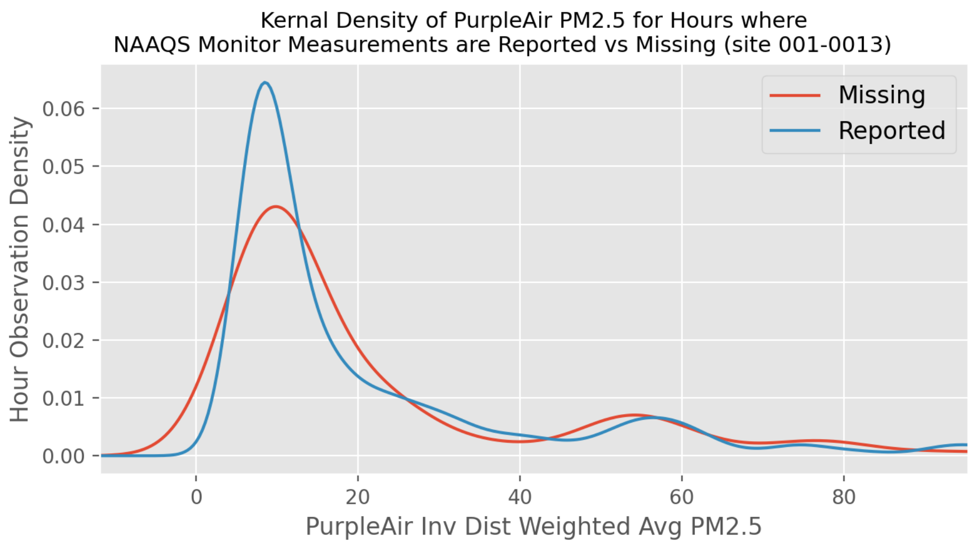

Figure 47: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0013 in county 001 (FIPS code).

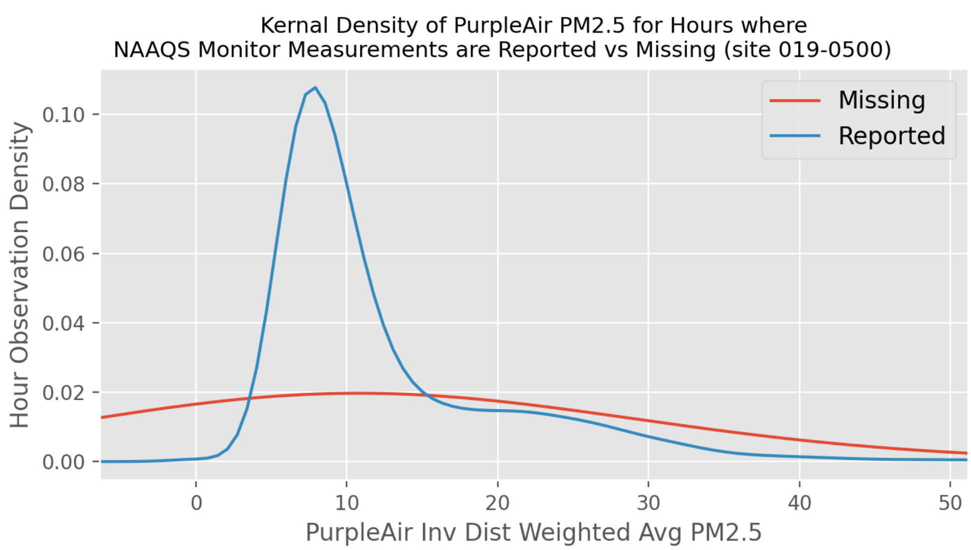

Figure 48: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0500 in county 019 (FIPS code).

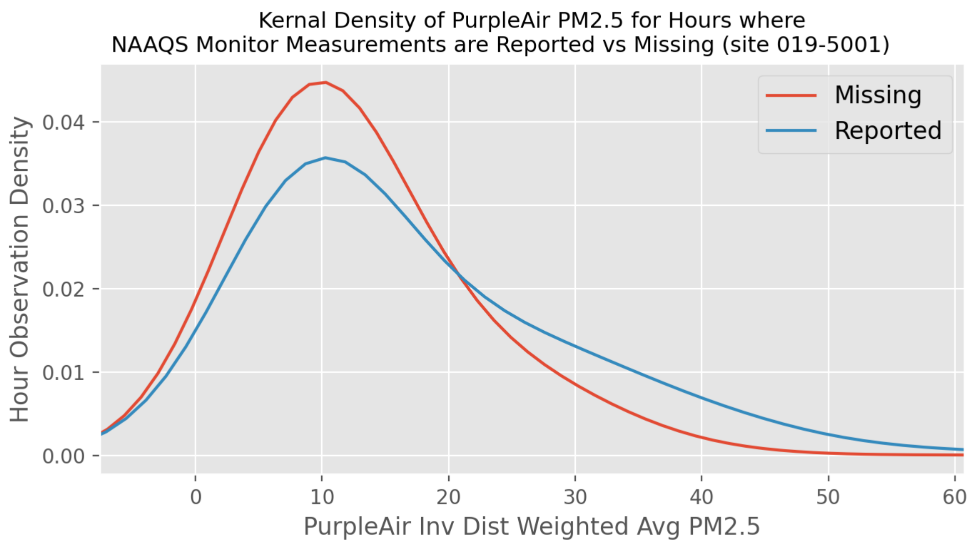

Figure 49: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 5001 in county 019 (FIPS code).

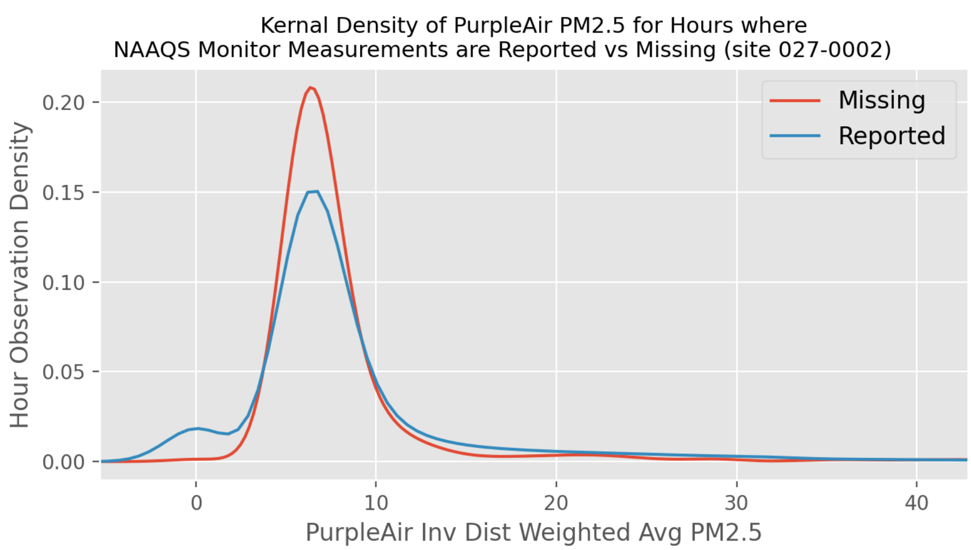

Figure 50: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0002 in county 027 (FIPS code).

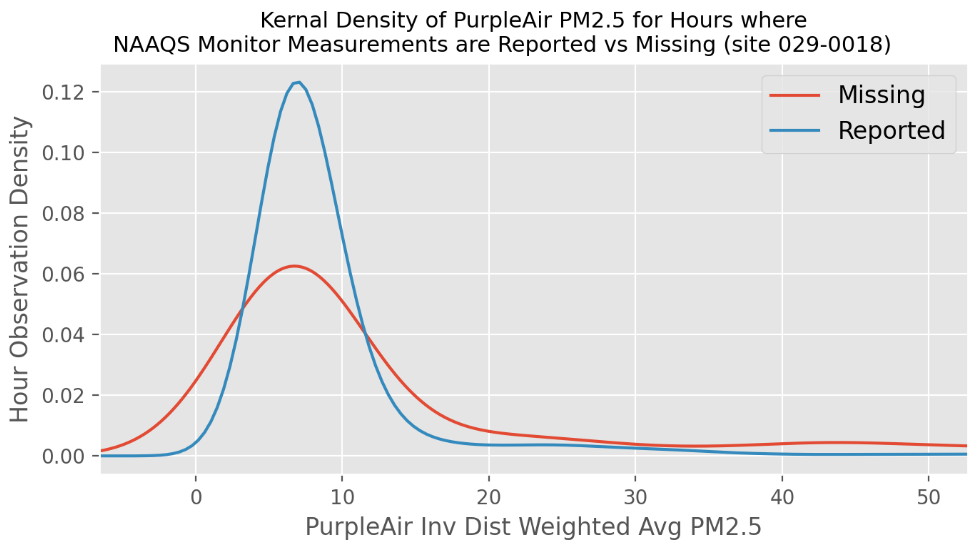

Figure 51: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0018 in county 029 (FIPS code).

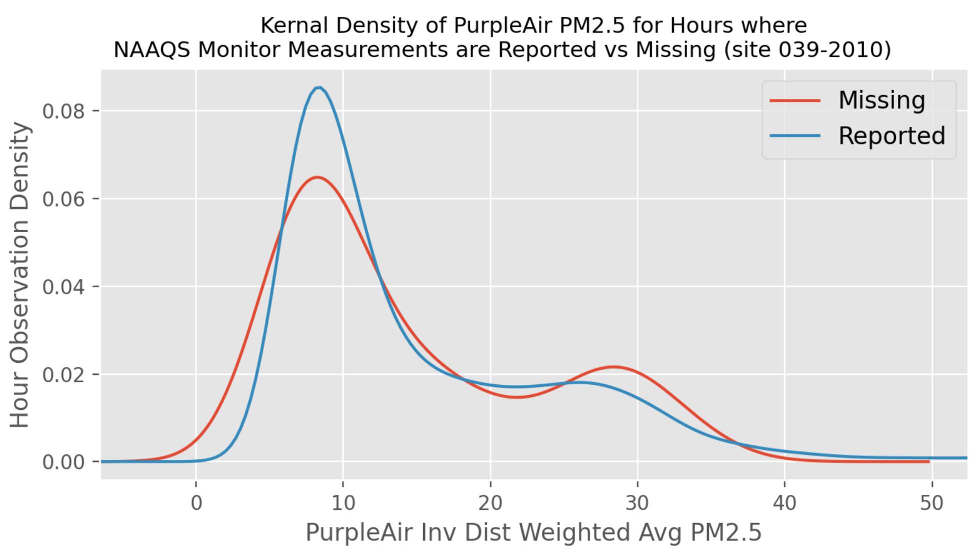

Figure 52: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 2010 in county 039 (FIPS code).

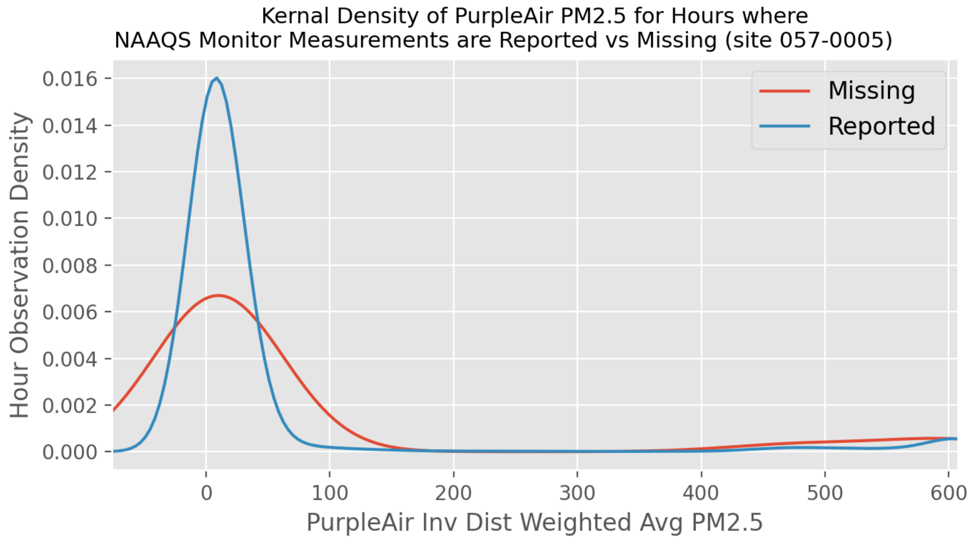

Figure 53: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0005 in county 057 (FIPS code).

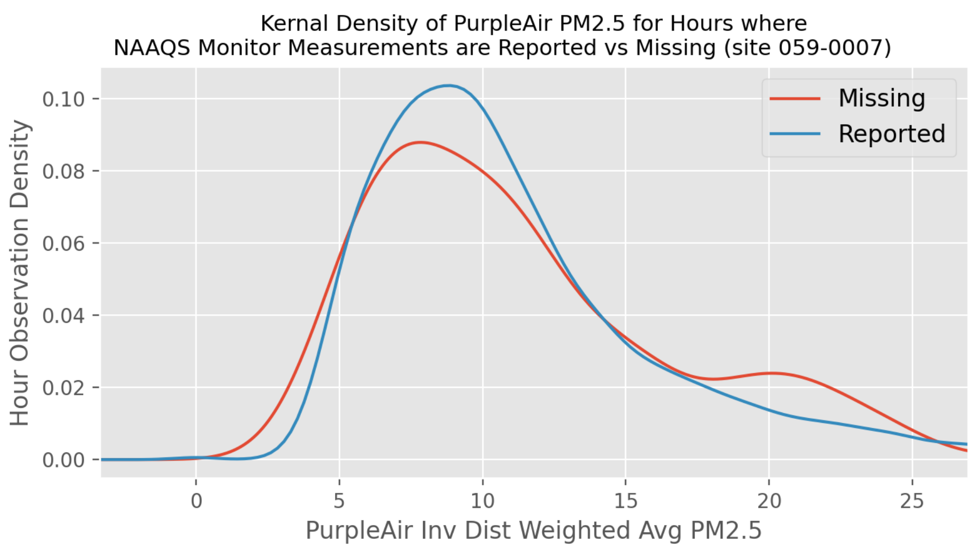

Figure 54: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0007 in county 059 (FIPS code).

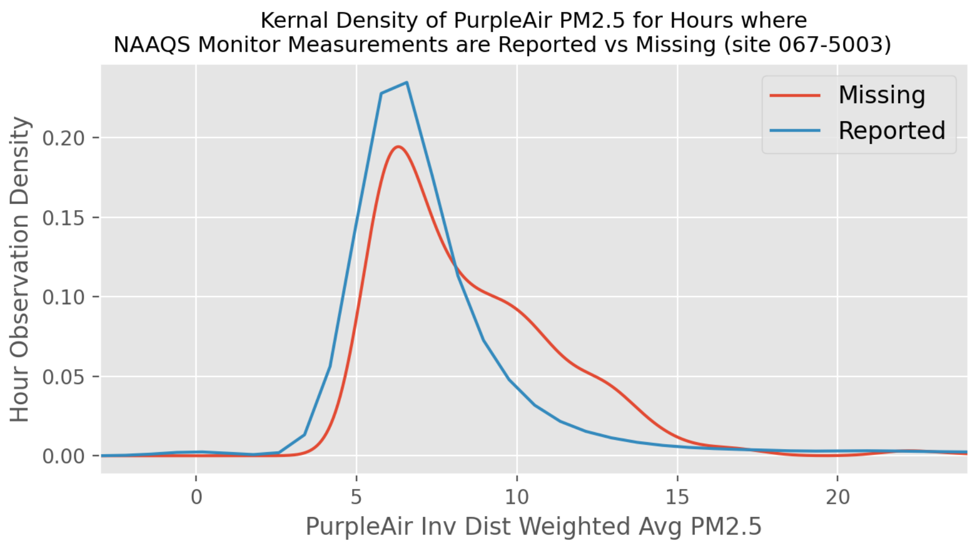

Figure 55: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 5003 in county 067 (FIPS code).

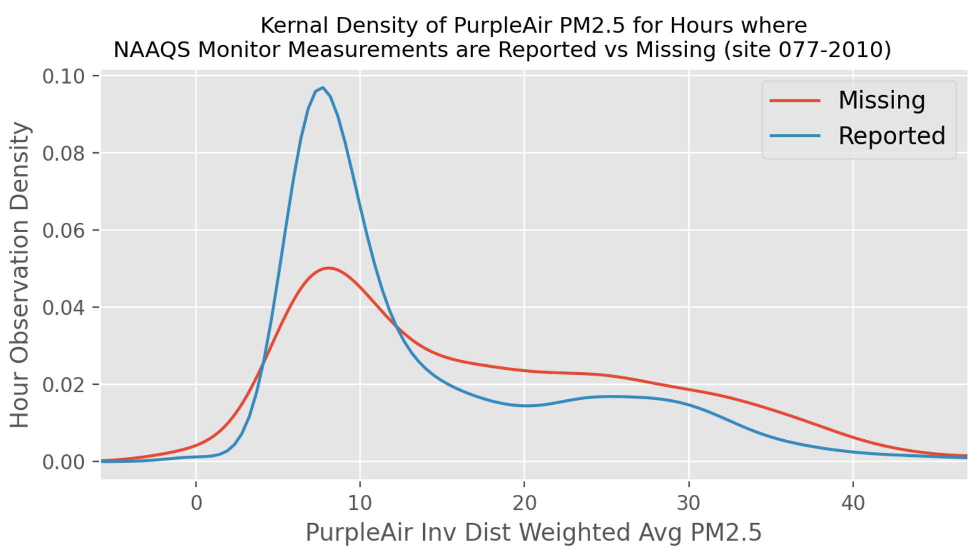

Figure 56: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 2010 in county 077 (FIPS code).

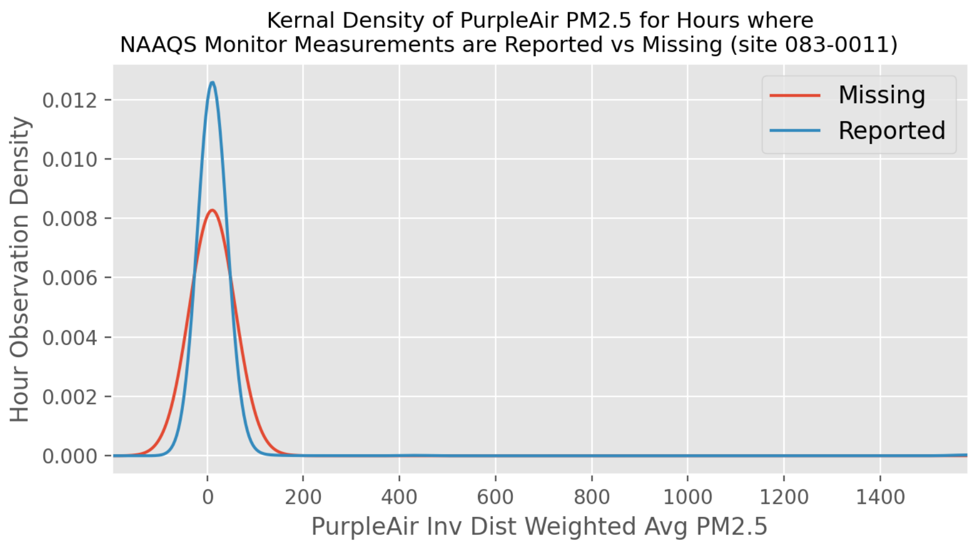

Figure 57: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0011 in county 083 (FIPS code).

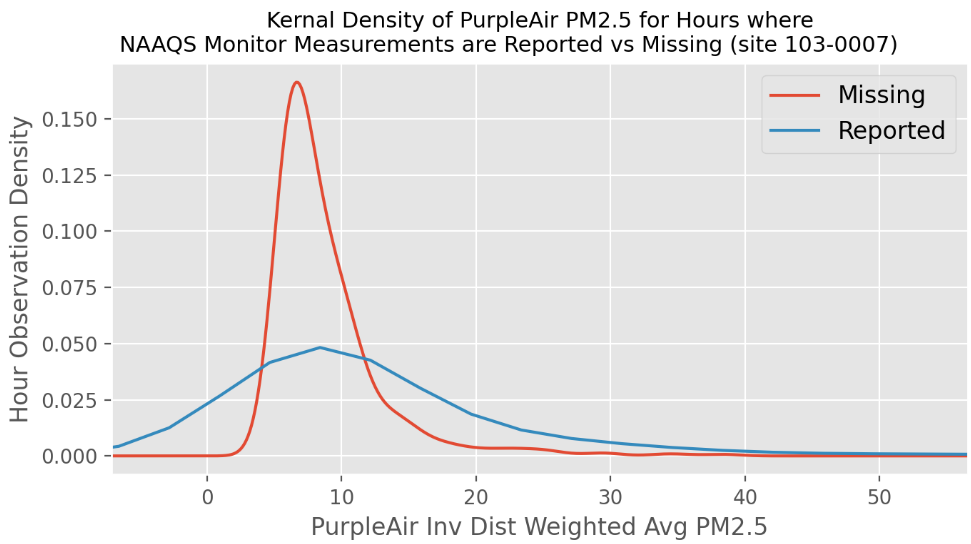

Figure 58: Comparison of PM2.5 concentration densities for two sets of hours: reported (blue) and missing (red) hourly observations of the NAAQS monitor. Both densities use the hourly PurpleAir PM2.5 concentration estimates for this site, calculated using the IDW average of PurpleAir sensors within 5 miles of the NAAQS monitor location. This monitor is at site 0007 in county 103 (FIPS code).