By: Sahar Gerandanesh
Intern, Johannes Gutenberg-Universität Mainz
Supervisor: Håkan Lane

Abstract

Air pollution is a pressing global environmental challenge, with PM2.5 (particulate matter with a diameter of less than 2.5 micrometers) being recognized as one of the most hazardous pollutants to human health. Prolonged exposure to PM2.5 has been linked to respiratory diseases, cardiovascular conditions, and premature mortality. This study aims to critically evaluate the reliability of local sensor data in comparison to satellite-derived values,

identifying potential discrepancies and exploring the implications for air quality monitoring and policy-making.

This study compares PM2.5 levels measured by satellite data from the Atmospheric Composition Analysis Group at Washington University in St. Louis (WUSTL, 2025) with ground-level measurements from the Sensor Community using SDS011 sensors deployed in Mainz, Germany. The primary objective is to assess the reliability of local sensor data compared to satellite-derived values and identify any discrepancies between the two data sources.

Theory

Particulate matter, particularly PM2.5, refers to fine particles that are less than 2.5 micrometers in diameter. These particles are significant due to their small size, which allows them to penetrate deep into the respiratory system and enter the bloodstream, causing severe health effects such as respiratory diseases, heart conditions, and even premature death. PM2.5 originates primarily from combustion processes, including emissions from vehicles, industrial facilities, power plants, and natural sources like wildfires and dust storms. Due to their health implications, PM2.5 levels are closely monitored globally.

There are two primary methods of measuring PM2.5 concentrations: ground-based sensors and satellite observations. Both approaches have distinct advantages and limitations, which form the basis of this study.

Ground-Based Sensors
Ground-based sensors, such as the SDS011 sensors used in this study, offer real-time measurements of air quality at localized points, providing detailed data on specific areas. These sensors are capable of detecting minute variations in air quality, which makes them ideal for identifying localized pollution hotspots. However, their reliance on fixed locations and potential calibration issues may introduce challenges in data accuracy. Despite these limitations, ground sensors contribute valuable information for urban air quality monitoring.

Satellite Observations
Satellite-based measurements, like those provided by the Atmospheric Composition Analysis Group at Washington University in St. Louis, offer a broader, regional perspective on PM2.5 pollution levels. Satellites capture a vast array of data, providing a global overview of air quality trends and regional pollution patterns. However, the spatial resolution of satellite measurements is typically lower compared to ground sensors, which can result in less precise data when assessing small-scale local pollution. Additionally, satellite data may not capture transient pollution events or localized emissions, making it essential to supplement satellite observations with more granular, ground-based sensor data.

Geospatial Alignment and Data Integration
To compare the effectiveness of satellite and sensor data, it is crucial to align both data sources spatially. Geospatial alignment ensures that satellite data points are compared to the nearest sensor locations, enabling meaningful comparisons. Additionally, temporal alignment is necessary, as satellite data may be available on a monthly or annual basis, while sensor data can provide more frequent, real-time readings.

Regression and Statistical Analysis
Regression analysis is employed to assess the correlation between the two data sources over time. By evaluating the slope and intercept of PM2.5 levels from both satellite and sensor data, researchers can determine trends, deviations, and discrepancies. Statistical tests, such as paired t-tests, help assess the significance of any differences observed, providing a quantitative measure of how closely the two datasets align.

This theoretical framework highlights the importance of both data types in monitoring air quality. Ground sensors offer detailed, localized data, while satellite observations provide a broader spatial perspective. Together, they offer complementary insights, contributing to a more comprehensive understanding of air pollution dynamics. The integration of both datasets allows for more accurate predictions of PM2.5 concentrations, facilitating better air quality management and informing public health policies.

Method

Data Sources

1. **Satellite Data**: Satellite observations provide a comprehensive overview of PM2.5 distribution across large geographical regions. The Atmospheric Composition Analysis Group at Washington University in St. Louis (WUSTL) generates high-resolution satellite data that offer valuable insights into air pollution trends. For this study, monthly average PM2.5 values from the nearest geographical points to the sensor locations were utilized. (Washington University in St)

2. **Sensor Community**: Ground-based PM2.5 measurements were obtained using SDS011 sensors deployed by volunteers in Mainz, Germany, as part of a citizen science initiative. These sensors provide real-time measurements of PM2.5 and PM10 concentrations, contributing to a more localized understanding of air pollution levels. (Sensor Community)

Data Preparation:

To facilitate a meaningful comparison between satellite and sensor data, the following preprocessing steps were undertaken:

- **Geospatial Alignment:** Satellite data points were matched to the nearest sensor locations using KDTree, ensuring precise geospatial correspondence.
- **Averaging:** Monthly averages of sensor data were computed to align with the temporal granularity of satellite observations.

Regression Analysis

A linear regression analysis was conducted to compare trends in PM2.5 levels over time.

The analysis was standardized by setting the year 2017 as a baseline (year zero), enabling a uniform temporal comparison across datasets.

Geospatial Analysis

- The spatial distribution of PM2.5 measurements was visualized through mapping.
- The distances between sensor locations and their corresponding satellite points were calculated to assess alignment accuracy.

Statistical Analysis

To assess the statistical significance of differences between the two datasets, a paired t-test was conducted. Regression

Visualization

Graphical representations, including line charts and geospatial maps, were created to depict trends, discrepancies, and spatial distribution patterns

Results

a. Comparison

The table below summarizes the comparison of PM2.5 values between satellite and sensor data, along with the differences for each sensor. The values are averaged for both satellite and sensor measurements over the entire year.

Table 1: Comparison of Yearly Average PM2.5 Values (µg/m³)

Sensor ID	Satellite PM2.5 (Avg, µg/m³)	Sensor PM2.5 (Avg, µg/m³)	Difference (µg/m³)
803	**10.322914**	**8.275690**	**2.047224**
10701	**10.359348**	**4.596867**	**5.762481**
21886	**9.125612**	**3.667080**	**5.458532**

23712	**9.889279**	**3.325984**	**6.563295**
26656	**10.028273**	**7.071762**	**2.956511**
47739	**9.960928**	**3.650180**	**6.310747**
48807	**9.559689**	**7.330577**	**2.229112**
77220	**9.956960**	**7.978370**	**1.978590**

Monthly Averaged Results

The following table summarizes the comparison of monthly averaged PM2.5 values for both satellite and sensor data. The difference between the two data sources is also calculated for each sensor, year, and month.

Table 2: Comparison of Monthly Average PM2.5 Values (μg/m³)

Sensor_ID	Year	Month	Satellite_PM25	Sensor_PM25	Difference
803	**2017**	**2**	**17.13**	**3.03**	**14.10**
803	**2017**	**3**	**13.48**	**10.45**	**3.02**
803	**2017**	**4**	**12.79**	**12.06**	**0.73**

803	**2017**	**5**	**10.71**	**9.34**	**1.37**
803	**2017**	**6**	**9.60**	**6.14**	**3.47**
...	**...**	**...**	**...**	**...**	**...**
66816	**2022**	**9**	**7.68**	**3.93**	**3.74**
66816	**2022**	**10**	**9.28**	**6.51**	**2.77**
66816	**2022**	**11**	**10.55**	**6.78**	**3.77**
66816	**2022**	**12**	**11.93**	**9.06**	**2.87**
77220	**2022**	**12**	**12.27**	**7.98**	**4.30**

Note: The table shows a sample of the results. The complete dataset includes 329 rows.

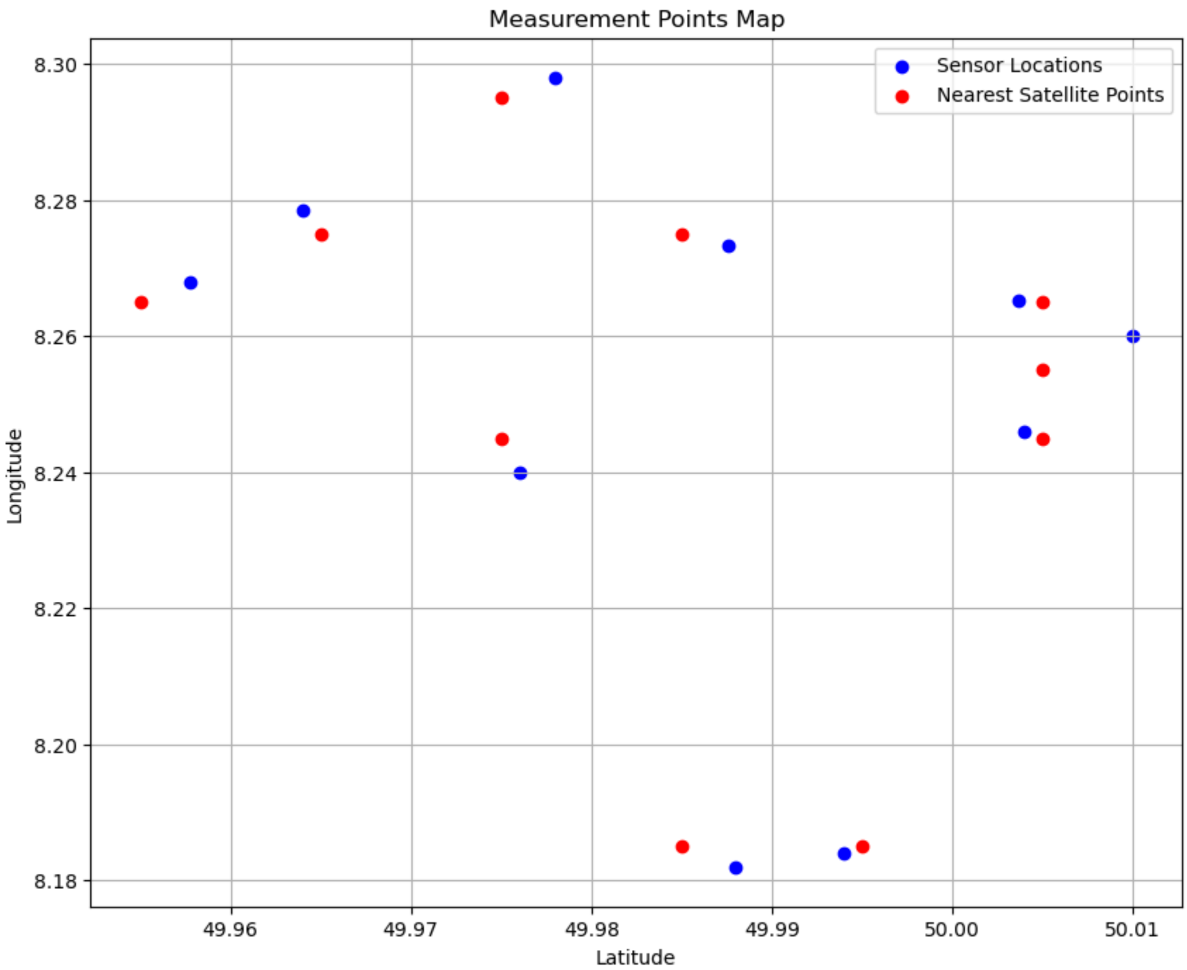
Sensor Locations and Nearest Satellite Points

The table below shows the sensor IDs, their corresponding latitude and longitude coordinates, and the nearest points from the satellite dataset:

Sensor ID	Latitude	Longitude	Nearest Satellite Point (lat, lon)
803	49.976	8.240	(49.975, 8.245)
10701	50.004	8.246	(50.005, 8.245)
21886	49.988	8.182	(49.985, 8.185)
23712	49.978	8.298	(49.975, 8.295)
26656	50.010	8.260	(50.005, 8.255)
47739	49.957776	8.267935	(49.955, 8.265)
48807	49.964022	8.278508	(49.965, 8.275)
66816	49.987581	8.273252	(49.985, 8.275)
77220	50.003637	8.265140	(50.005, 8.265)
83487	49.994	8.184	(49.995, 8.185)

Map of Measurement Points

Below is a map showing the locations of the sensors and their corresponding nearest satellite points. The map provides a visual representation of where the measurements were taken, which helps in understanding the spatial distribution of PM2.5 levels.

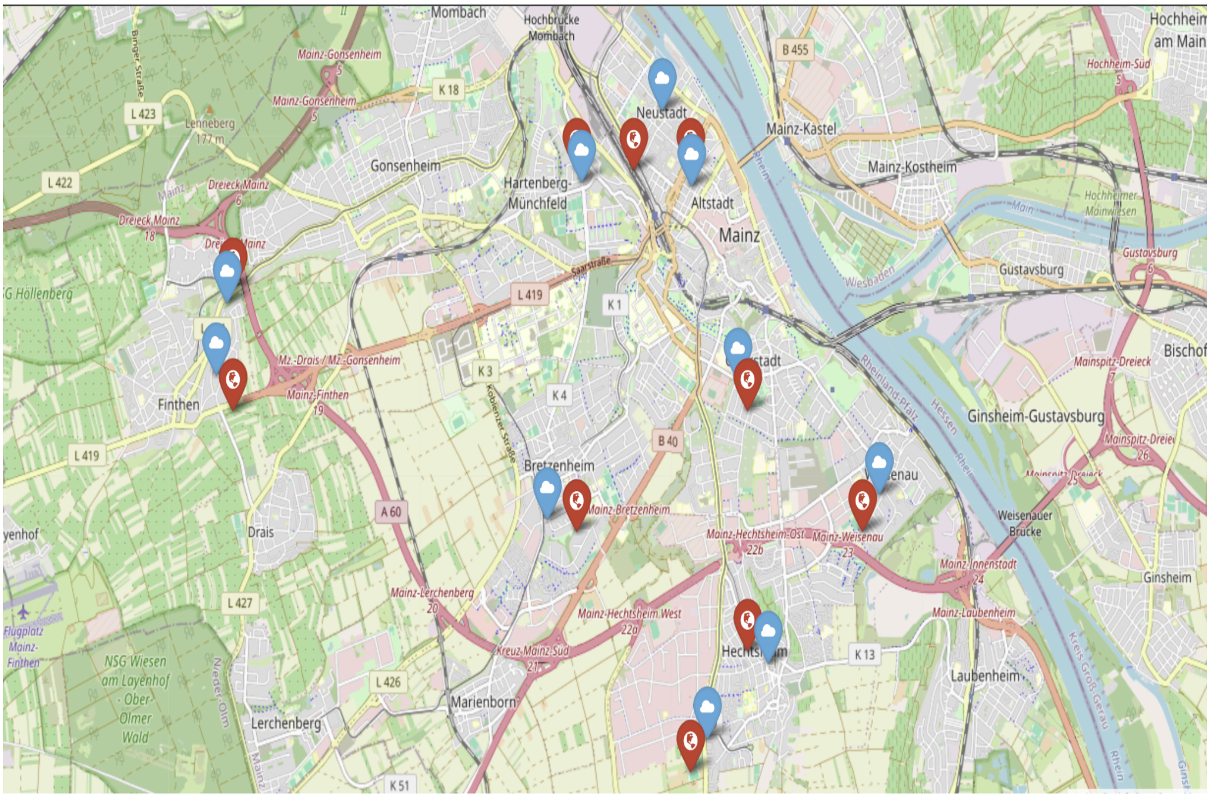


Geospatial Analysis:

Distances between sensors and the nearest satellite points were calculated and validated. The distances measured are as follows:

Sensor ID	Distance to Nearest Satellite (km)
803	0.38
10701	0.13
21886	0.40
23712	0.40
26656	0.66
47739	0.37
48807	0.27
66816	0.31
77220	0.15
83487	0.13

Interactive map displaying all sensor and satellite measurement points:

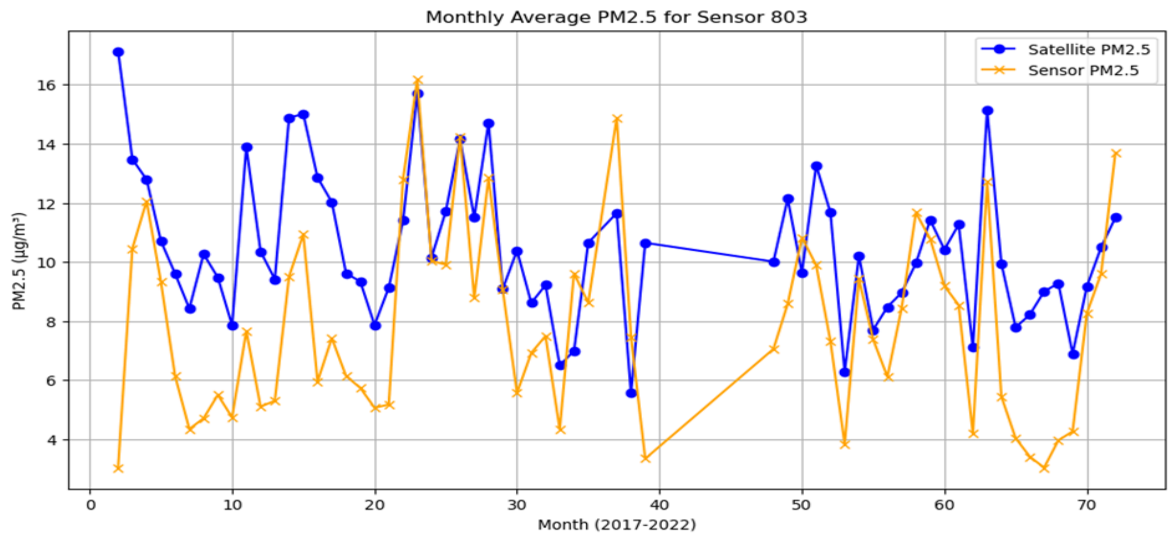


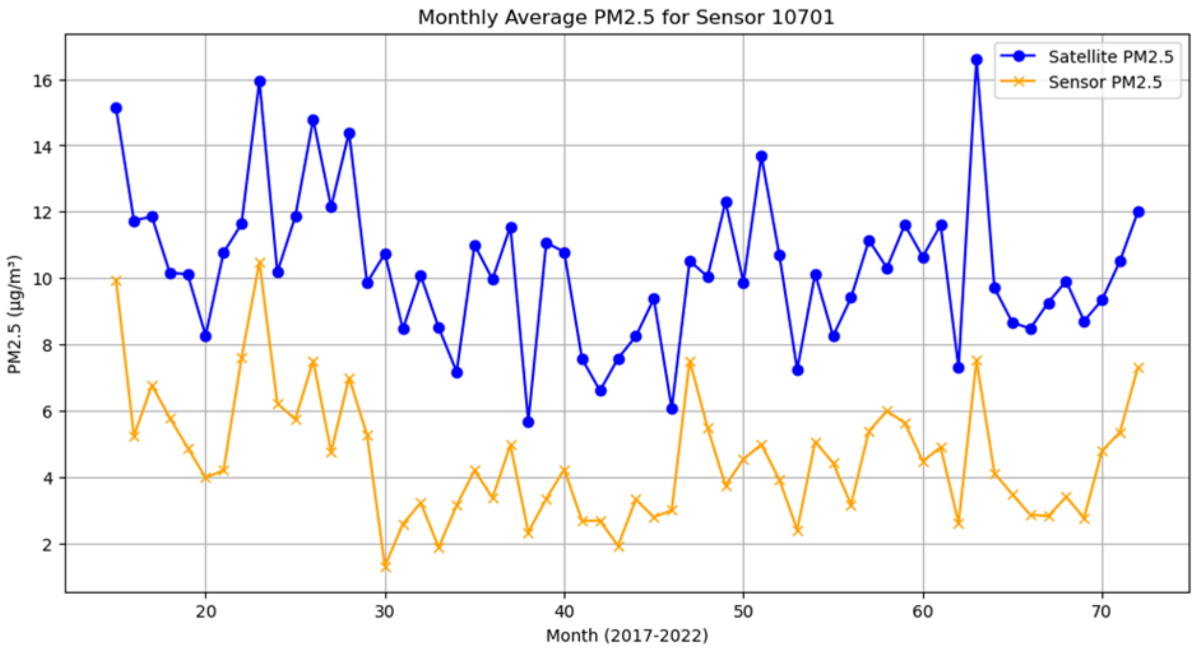
Visualization of Monthly Average PM2.5 for Each Sensor

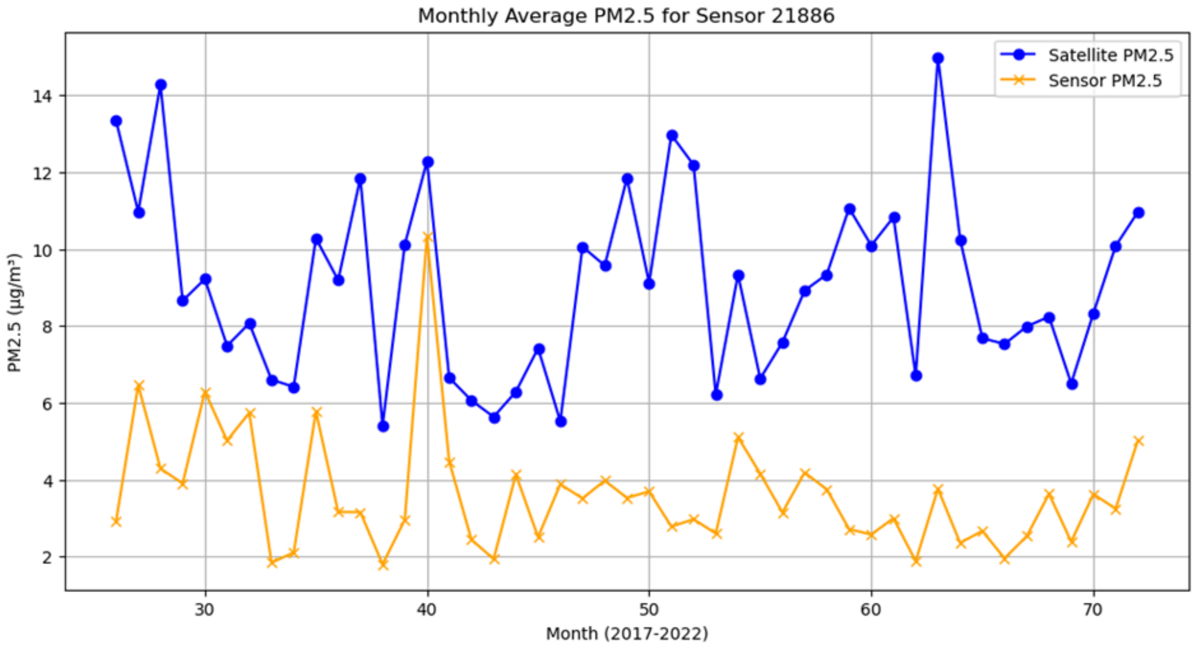
Below are the visualizations showing the comparison of monthly average PM2.5 values for each sensor from both satellite data and sensor data.

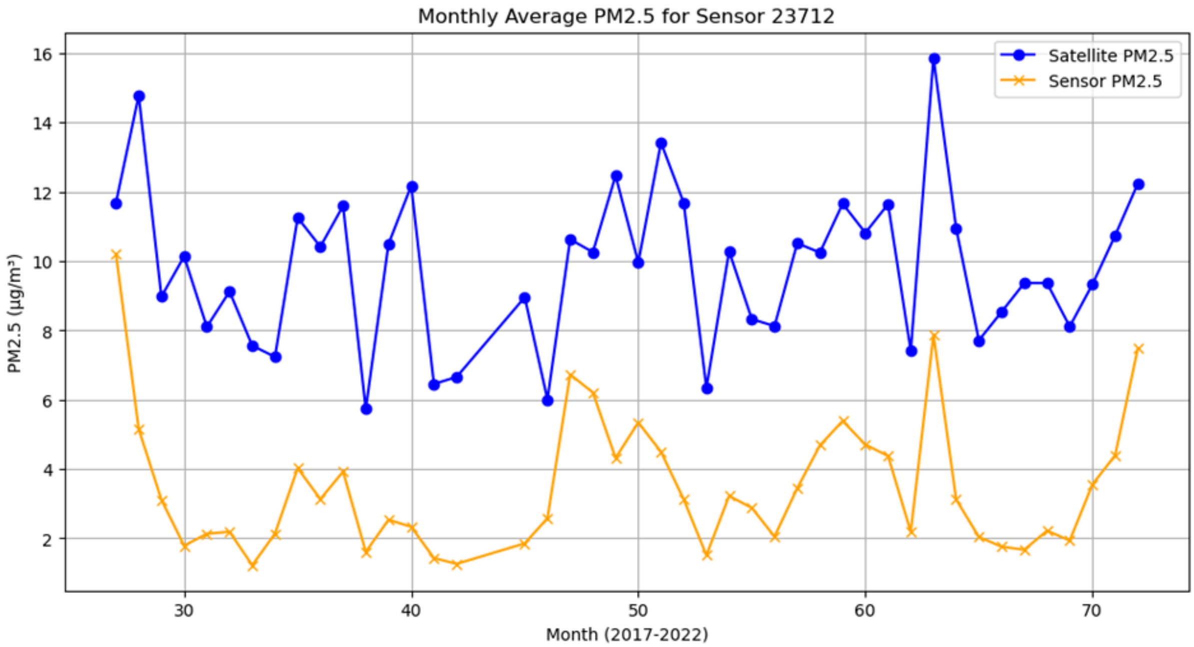
- Blue Line: Satellite PM2.5 data
- Orange Line: Sensor Community PM2.5 data

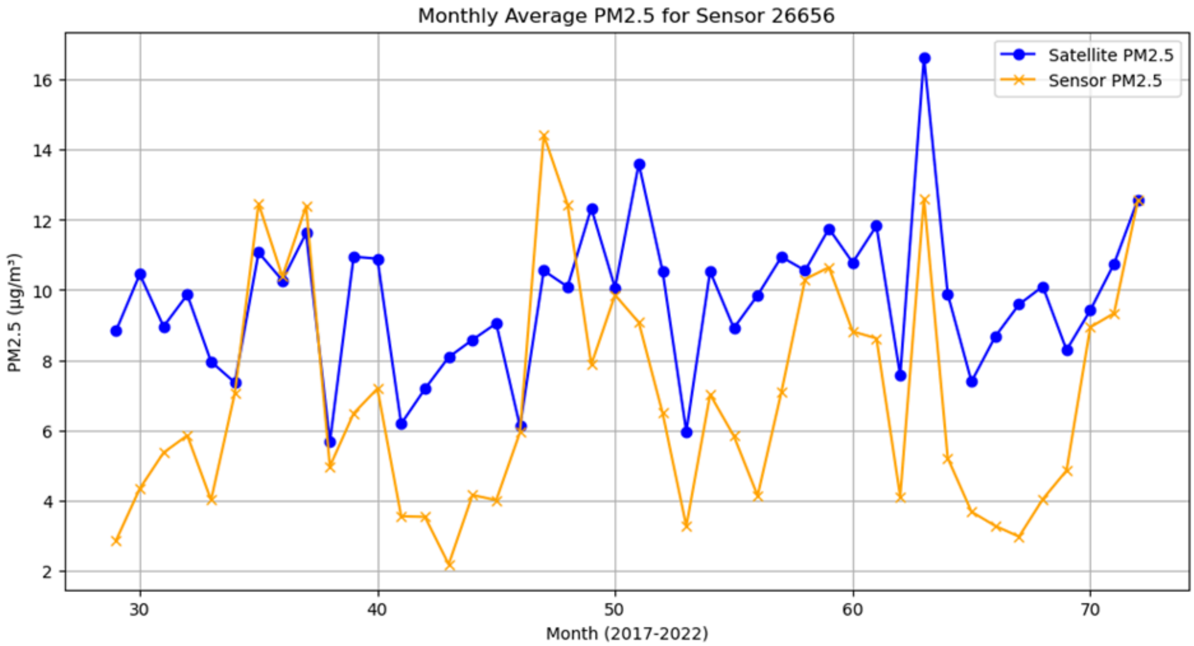
These visualizations provide insight into how the two data sources align or differ over time.

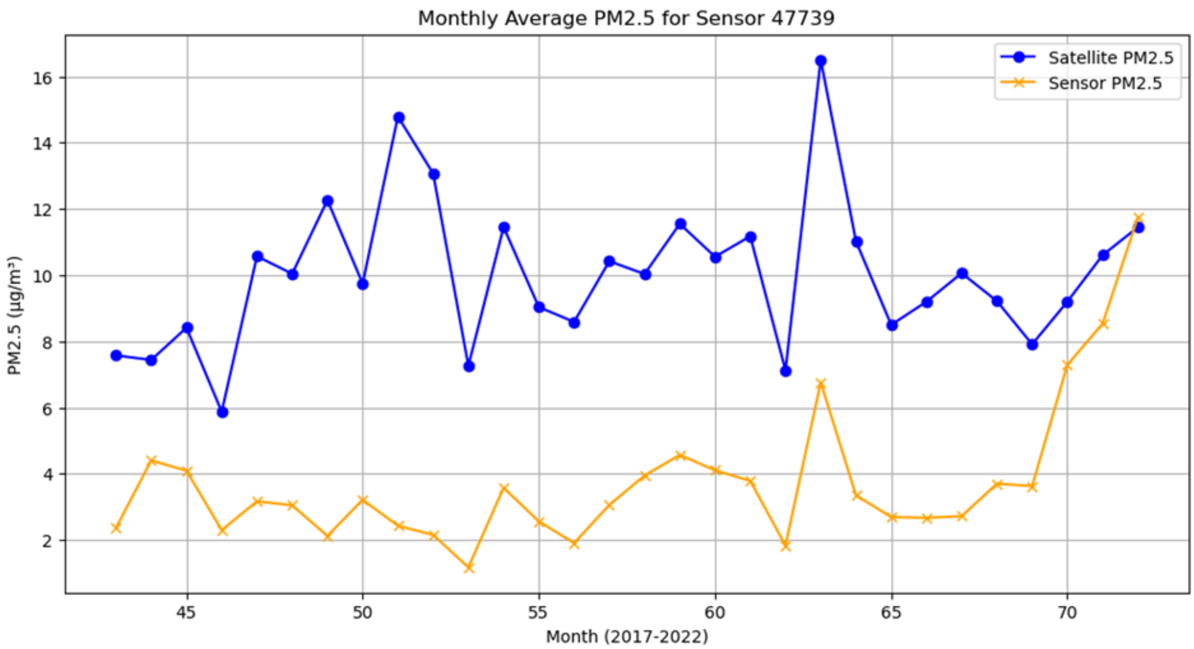

Monthly Average PM2.5 for Sensor 803
Satellite PM2.5
Sensor PM2.5
PM2.5 (µg/m³)
Month (2017-2022)

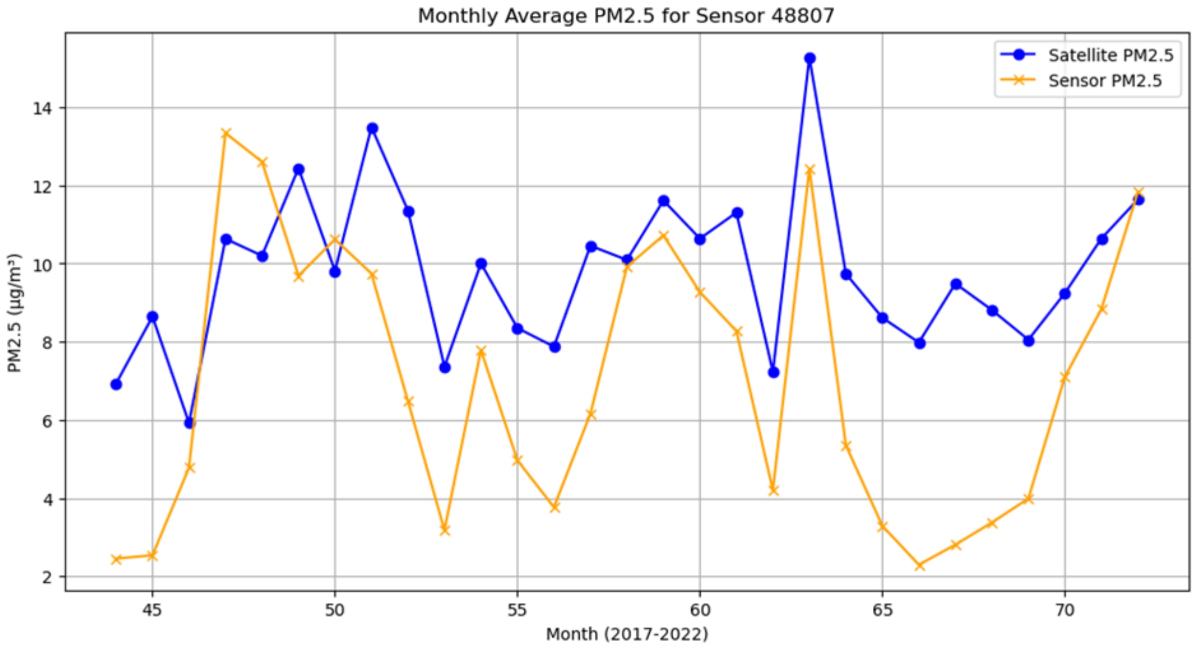

Monthly Average PM2.5 for Sensor 10701
Satellite PM2.5
Sensor PM2.5
PM2.5 (µg/m³)
Month (2017-2022)


Monthly Average PM2.5 for Sensor 21886
Satellite PM2.5
Sensor PM2.5
PM2.5 (µg/m³)
Month (2017-2022)
2
4
6
8
10
12
14
30
40
50
60
70


Monthly Average PM2.5 for Sensor 23712
Satellite PM2.5
Sensor PM2.5
PM2.5 (µg/m³)
Month (2017-2022)
2
4
6
8
10
12
14
16
30
40
50
60
70


Monthly Average PM2.5 for Sensor 26656
Satellite PM2.5
Sensor PM2.5
PM2.5 (µg/m³)
Month (2017-2022)


Monthly Average PM2.5 for Sensor 47739
Satellite PM2.5
Sensor PM2.5
PM2.5 (µg/m³)
Month (2017-2022)



Key Findings

1. **General Trends:** Satellite data consistently reported higher PM2.5 levels compared to ground sensors, which may be attributed to their broader spatial scope.

2. **Variability:** Differences between datasets were inconsistent across different sensors and time periods, highlighting the influence of local factors on air pollution measurements.

3. **Statistical Significance:** A paired t-test revealed a statistically significant difference between the two data sources (T-statistic: 4.1271, P-value: 0.0000), indicating potential methodological discrepancies.

Regression Analysis

A linear regression analysis was conducted to observe the trend in PM2.5 levels over time for both satellite and sensor data. The regression was adjusted so that the year 2017 is treated as year 0.

Table 3:Regression

Sensor ID	Satellite Slope (µg/m³/year)	Satellite Intercept (µg/m³)	Sensor Slope (µg/m³/year)	Sensor Intercept (µg/m³)
803	-0.54	11.67	-0.08	8.47
10701	-0.33	11.36	-0.39	5.76
21886	0.01	9.09	-0.46	5.28
23712	0.05	9.73	0.27	2.38
26656	0.04	9.88	0.04	6.94
47739	0.59	7.61	0.78	0.53
48807	0.64	7.00	-0.72	10.20
77220	0.00	9.96	0.00	7.98

Significance test

Based on the results of the paired t-test conducted between satellite and sensor data for PM2.5, the following conclusions can be drawn:

- **T-statistic:** -2209.97

- **P-value:** 0.0

Since the **p-value is significantly less than 0.05**, we reject the **null hypothesis** (which states that there is no significant difference between the two datasets). This indicates that there is a **statistically significant difference** between the PM2.5 levels measured by the satellite and those measured by the ground sensors.

Test Explanation:

The **paired t-test** compares the PM2.5 values measured by the satellite and sensor systems, taking into account both the spatial and temporal data alignment. The test calculates the difference in PM2.5 readings for each location over time and evaluates whether this difference is statistically significant.

Comparison to WHO guidelines-över 5

According to the World Health Organization (WHO), the annual average concentration of PM2.5 should not exceed **10 µg/m³**. The results of this study show that both satellite and sensor data frequently exceed this guideline across multiple sensors and time periods.

Key Findings:

- **Satellite Data**: The PM2.5 values from the satellite-derived measurements consistently exceed the WHO guideline of 10 µg/m³. In fact, many of the sensors had yearly averages well above this threshold, with differences ranging from 1.98 µg/m³ to 6.56 µg/m³ above the WHO recommended limit.

- **Sensor Data**: While the sensor data generally reported lower PM2.5 levels than the satellite data, they still frequently exceeded the WHO guideline. For example, sensor ID 803 showed a yearly average of 8.28 µg/m³, which is close to the threshold, while others like sensor ID 10701 had an average of 4.60 µg/m³ but still fell short of ideal air quality.

Table 4: Comparison (Yearly Averages):

Sensor ID	Satellite PM2.5 (Avg, µg/m³)	Sensor PM2.5 (Avg, µg/m³)	Difference (µg/m³)	WHO Guideline
803	10.32	8.28	2.05	Exceeds WHO
10701	10.36	4.60	5.76	Exceeds WHO

21886	9.13	3.67	5.46	Exceeds WHO
23712	9.89	3.33	6.56	Exceeds WHO
26656	10.03	7.07	2.96	Exceeds WHO
47739	9.96	3.65	6.31	Exceeds WHO
48807	9.56	7.33	2.23	Exceeds WHO
77220	9.96	7.98	1.98	Exceeds WHO

Discussion

The comparison between satellite-derived PM2.5 data and community-based sensor measurements highlights both the strengths and limitations of each approach. Satellite data provide broad spatial coverage, making them useful for identifying regional pollution patterns. However, they may not capture localized pollution hotspots as accurately as ground-level sensors.

Key observations from the analysis include:

- Satellite measurements consistently reported higher values than ground-based sensors.
- All monthly averages of the satellite data exceeded the WHO safe limit.
- The pollution level has remained relatively stable across the city since 2017.

The differences observed in this study underscore the importance of integrating multiple data sources to achieve a comprehensive understanding of air quality. Significant discrepancies were noted between satellite and sensor measurements, which may be attributed to factors such as:

- **Local pollution sources:** Sensors capture real-time, localized air quality data influenced by traffic, industrial activities, and other local emissions that satellite measurements might miss.
- **Different measurement techniques:** Satellite data measure pollution over larger areas, while sensors focus on specific points, potentially leading to discrepancies.

These findings also raise questions about the reliability of low-cost sensors. While they provide valuable real-time data, regular calibration and validation are crucial to ensure accuracy and consistency in readings.

Future Directions

1. **Sensor Calibration:** Regular calibration of low-cost sensors is essential to improve data accuracy and reliability.
2. **Hybrid Approaches:** The integration of satellite and ground sensor data can provide a more nuanced understanding of air quality patterns.
3. **Localized Studies:** Further research into local pollution sources and their impact on sensor data discrepancies is warranted.

Conclusion

This study demonstrates that while satellite data and community sensor data provide valuable insights into PM2.5 pollution, there are notable differences between the two. Addressing these discrepancies will enhance the reliability of air quality assessments and support more effective environmental policies.

By combining satellite data with community-driven measurements, this research underscores the importance of innovative, multi-faceted approaches to monitoring urban air quality.

Review of Existing Research on Satellite and Ground-Based PM2.5 Measurements

Several studies have explored the comparison between satellite-derived air quality data and ground-level sensor measurements:

- Mushtaq, Z., Pargin Bangotra, Alok Sagar Gautam, Sharma, M., Suman, N., Gautam, S., Singh, K., Kumar, Y., & Jain, P. (2024). Satellite or ground-based measurements for air pollutants (PM2.5, PM10, SO2, NO2, O3) data and their health hazards: which is most accurate and why? *Environmental Monitoring and Assessment*, *196*(4). https://doi.org/10.1007/s10661-024-12462-z

  Mushtaq et al. (2024) in their study "Satellite or ground-based measurements for air pollutants (PM2.5, PM10, SO2, NO2, O3) data and their health hazards: which is most accurate and why?" discuss the comparison between satellite and ground-based measurements for PM2.5 and other pollutants,

emphasizing the importance of evaluating the accuracy of each method in the context of their health implications. They state:
"While satellite measurements provide broader spatial coverage, ground-based sensors are crucial for capturing local pollution hotspots, which can have significant health impacts."

- Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., & Zhang, Q. (2019). Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations. *Environment International*, *123*, 345–357. https://doi.org/10.1016/j.envint.2018.11.075

  Xue et al. (2019) in their study "Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016" focus on the integration of satellite data, chemical transport models, and ground observations for estimating PM2.5 concentrations. They highlight the value of combining different data sources to enhance the accuracy of air quality estimates. They state:

  "The integration of satellite data with ground-based observations significantly improves the spatiotemporal resolution of air quality estimates, providing a more comprehensive understanding of pollution dynamics."

References

- Sensor Community. “Sensor.community | Build Your Own Sensor and Join the Worldwide Civic Tech Network.” *Sensor.community*, 2024, sensor.community.

- Washington University in St. “Home | Atmospheric Composition Analysis Group | Washington University in St. Louis.” *Wustl.edu*, 2022, sites.wustl.edu/acag.