FROM FIXED POINTS TO FLUID MEASUREMENTS

STRATEGIZING MOBILE AIR QUALITY MONITORING WITH STATIONARY DATA IN CHICAGO

Pouyan Hatami, PhD

2023

Introduction

Air pollution refers to the presence of pollutants in the air which are detrimental to human health and the environment. Most air pollution in the United States is due to automobile traffic, but other sources of air pollution include power plants, agricultural land, cities, wildfires, and volcanoes. Poor air quality poses serious economic, social, environmental, and health consequences. In 2019, 99% of the world's population was living in places where air quality guidelines levels were not met (WHO, 2021). In the United States, about 111 million Americans - 35% of the population - live in counties with unhealthy air (Peltier, 2019). In 2016, 4.2 million deaths occurred worldwide due to ambient air pollution (WHO, 2021). Air pollution is also now empirically linked to negative psychological and societal outcomes such as increasing depression, impairing cognitive abilities, hurting worker productivity, and exacerbating criminal behavior (Church, 2020).

The EPA is required by the U.S. Clean Air Act (CAA) to establish National Ambient Air Quality Standards (NAAQS) for six common air pollutants that are known as "criteria pollutants". They are termed criteria because their thresholds are based on human health (primary standards) or environmental criteria (secondary standards) including visibility impairment, climate effects and material effects. The pollutants are particulate matter (PM), ground-level ozone (O3), carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2) and lead. Particle pollution and ground-level ozone are the most prevalent health threats.

According to the EPA's air quality standards, any $PM_{2.5}$ level below 12 μg/m³ and 15 μg/m³ is considered good air quality for primary and secondary standards, respectively (EPA, 2022). Specifically, $PM_{2.5}$ can pose extreme health threats because these tiny particles matriculate deep into human lungs. Exposure to $PM_{2.5}$ has been linked to an increased risk of cardiovascular disease (Jiang et al., 2016). Furthermore, previous research has confirmed that negative health outcomes due to $PM_{2.5}$ exposure can result in premature death (Cohen et al., 2017).

Past studies have identified variables like traffic density as influences on the distribution of $PM_{2.5}$. In fact, exhaust from diesel-powered vehicles has been determined as the primary cause of roadway pollutant concentrations (Fruin et al., 2008). Several studies have confirmed this effect of traffic on air pollution levels (Cummings et al., 2021; deSouza et al., 2020; Apte et al., 2017). Other studies have also shown that land cover (Shakya et al., 2019), building characteristics (Li et al., 2019), and airport proximity (Hudda et al., 2014) are also strong predictors of pollutant concentrations in urban areas.

While air pollution affects everyone, people who live in urban areas like Chicago suffer the most exposure to air pollution as compared to those living in more rural areas. Furthermore, the increased burden of air pollution on large cities is not distributed equally. Data from Microsoft shows that certain Chicago communities such as Little Village, Austin, Englewood, Auburn Gresham, Irving Park, and Avondale have the highest levels of pollutant particulate matter due to heavier traffic and closer proximity to industrial areas (Gupta et al., 2022).

To create effective mitigation plans that accurately reflect the unequal distribution of pollution, air quality monitoring is crucial. Air quality monitoring measurements can reveal hotspots, or areas with the most pollution, by measuring pollutant concentrations over set periods of time in different locations. Commonly used forms of data collection for air quality monitoring are stationary and mobile systems. Stationary air quality monitoring systems provide data from fixed locations. Stationary systems are used most often because of their high accuracy and ability to monitor long-term air quality. On the other hand, mobile monitoring which typically refers to collecting air quality data in real-time while moving whether by vehicle, foot, or other method of transportation, is a particularly reliable, low-cost option that can acquire data at a high spatial resolution (i.e., hyperlocal). They can characterize smaller-scale variability, and measure a widespread area. The major air quality measurement programs in the U.S. include those run by the EPA, National Parks Service, NASA, NOAA, and the National Weather Service. In Chicago, prominent air quality monitoring includes data collection by Microsoft, the Chicago Department of Public Health (CDPH), and the EPA. However, of the U.S.-based studies that solely use mobile air quality monitoring (Cummings et al., 2021; Shakya et al., 2019; Cabral et al., 2021; deSouza et al., 2020; Anjomshoaa et al., 2018; Hankey and Marshall, 2015; Apte et al., 2017; Fruin et al., 2008; Hudda et al., 2014; Chen et al., 2022), almost all are vehicle-based which is less accessible and requires more resources, and they have short data collection periods of less than a year.

The temporal variability and long-term patterns of air pollution are perfectly captured by stationary sensors. However, due to the complexity of urban air pollution and the significant spatial variability it exhibits within communities, a vast number of stationary air quality sensors dispersed across the community would be necessary to map the fine resolution spatial variability of air pollution. As a result of this shortcoming, mobile monitoring can be used to measures pollution levels on accessible road networks in metropolitan communities, to complement stationary monitoring methods (Chen et al., 2022). Aclima (2020) demonstrated that from one block to another, air pollution can fluctuate by six to eight times. According to their findings, mobile air quality monitoring offers hyperlocal spatial representativeness that exhibits significant fluctuations in pollutant concentrations at distances between 100 and 300 meters.

When compared to stationary sensors, the mobile sampling system suffers from the limited time spent at each location visited during the data gathering procedure. Due to regular or sporadic emission plumes from adjacent sources, mobile measurement data may contain strong concentration spikes within a few seconds. Understanding the number of repeat visits that may be needed to attain a meaningful approximation of the spatial air quality pattern in a particular area is a key question in relation to mobile monitoring. Few studies have attempted to address this subject, and the challenge is that the number of repeats needed can vary depending on the location characteristics and the target air pollution variable. Chen et al. (2022) argues the local variability of air pollution can be a dependent on the type of the pollution, spatiotemporal dimensions (location, and time including hours, weekday, and season) of the measurement, the deviation from the central tendency measures, and the local climatology.

The present study seeks to enhance our understanding of hyperlocal air quality in Chicago by employing two data sets specific to the city. The first source is data from stationary sensors in Chicago, under the Project Eclipse Network, which is owned and operated by Microsoft. Since July 2021, this network has been continuously monitoring air quality in Chicago for three pollutants: $PM_{2.5}$, O_3, and NO_2. The network consists of over 100 stationary sensors. The second source is ELPC mobile monitoring data collected using AirBeam devices. These devices are low-cost, open-source, handheld particulate matter monitors that ELPC has been using to measure short-term $PM_{2.5}$ exposure in Chicago since 2017. The data collected from Microsoft stations aids in capturing the temporal variability of air quality in Chicago across various time scales. This information can subsequently be used to efficiently implement a comprehensive mobile monitoring system with AirBeam devices and provide recommendations for future mobile monitoring endeavors.

Data

ELPC Mobile Air Quality Monitoring Project

The AirBeam is a low-cost, open-source, handheld air monitor that ELPC has been utilizing to measure short-term $PM_{2.5}$ exposure since 2017. Despite being inexpensive, the measures are in good agreement with USEPA federal regulatory monitors (FRM). The AirBeam measures the amounts of $PM_{2.5}$ by taking air samples at 1 HZ (1 per second intervals). One may view the particulate matter levels in real time as they move throughout the city thanks to the concentrations being transmitted immediately to a mobile device via Bluetooth.

In 2017, ELPC initiated a community science air quality monitoring project and partnered with neighborhood residents, community organizations, and students, to conduct air quality monitoring, collecting and mapping small particulate levels. Every spring and summer, ELPC collaborated with neighborhood-based organizations to train locals in tracking air pollution, recognizing their exposure to $PM_{2.5}$, and making better health-related decisions. While citywide monitoring was carried out, ELPC concentrated on gathering information on the west, south, and southeast portions of the city. This was mostly based on earlier studies on the city of Chicago, where industrial districts, high traffic corridors, and asthma rates suggested that some locations should be given top priority because they are more likely to be exposed to $PM_{2.5}$.

While relying primarily on walking-based mobile monitoring, over 14 million data points were collected between 2017 and 2021 by ELPC and its community partners, and 10.47 (an equivalent of 2905.5 hours or 121 full days of data collection) million were within the city limits of Chicago. The temporal breakdown of the collected data is displayed in figures 1 to 3. The figures show that the majority of the data was collected in July, and that the hourly range with the highest rate of monitoring was from 9 A.M. to 4 P.M. Figure 4 displays the number of measurements made in the top 26 neighborhoods where more than 100,000 data points were collected, accounting for 76% of the total data set. Lincoln Park had the most measurements made with over three million data points.

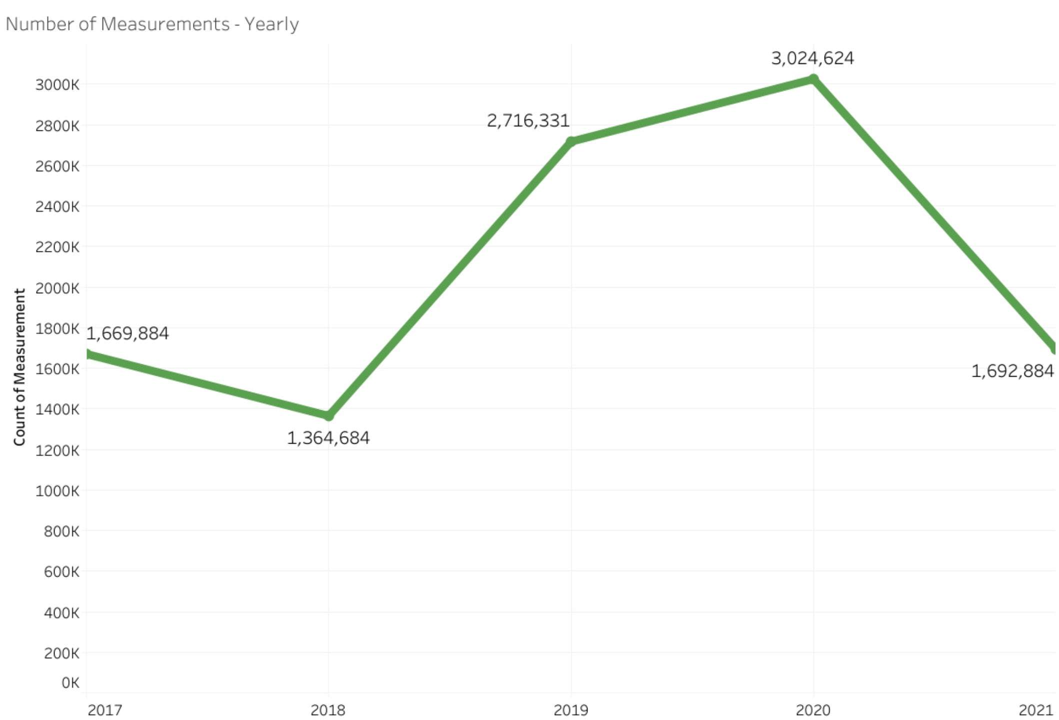

Figure 1: Number of measurements collected by ELPC team yearly, 2017-2021. Shows the most data was collected in 2019 and 2020.

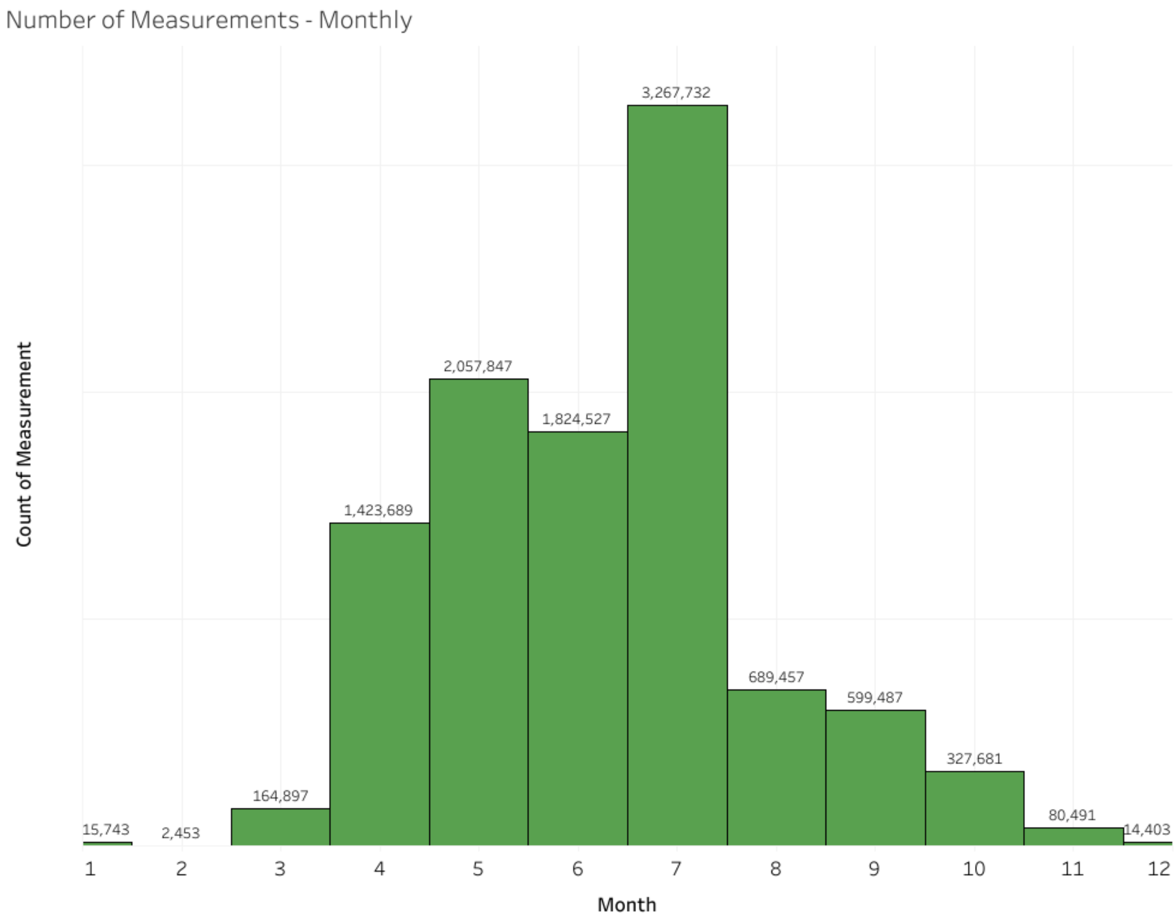

Figure 2: Measurements collected by ELPC team, monthly, 2017-2021.
Shows the most data was collected during the warmer months, especially July.

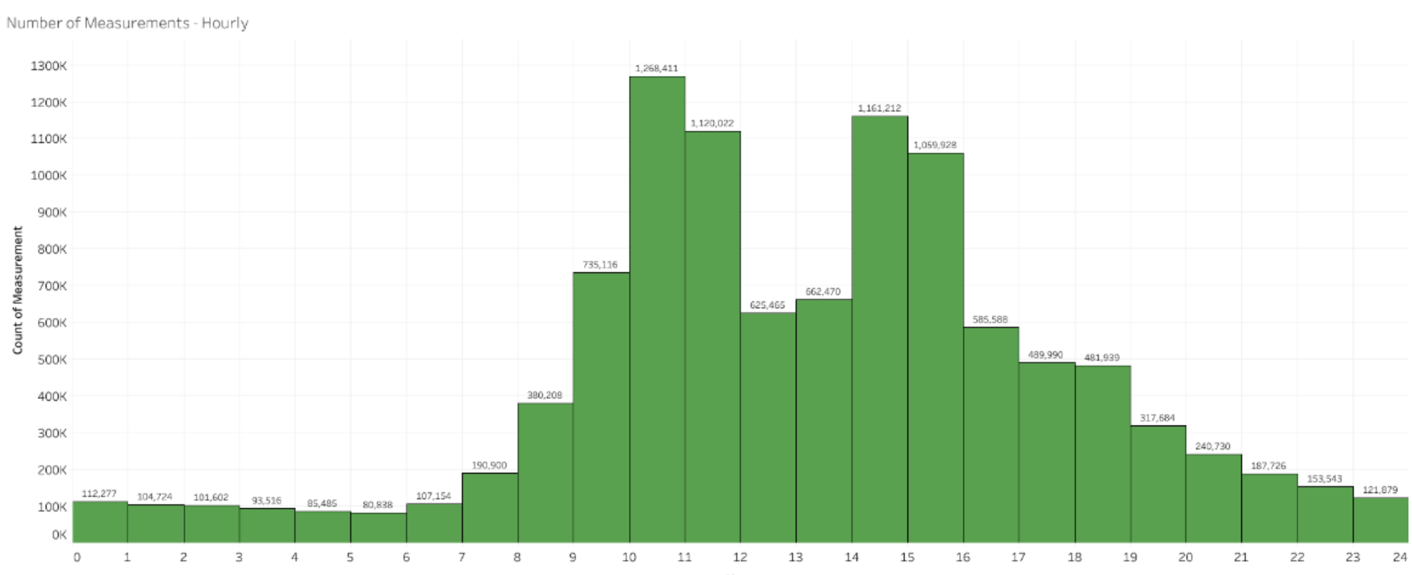

Figure 3: Measurements collected by ELPC team, by time of day, 2017-2021. Shows the most data was collected in the morning and afternoons, especially 10am-12pm and 2pm-4pm.

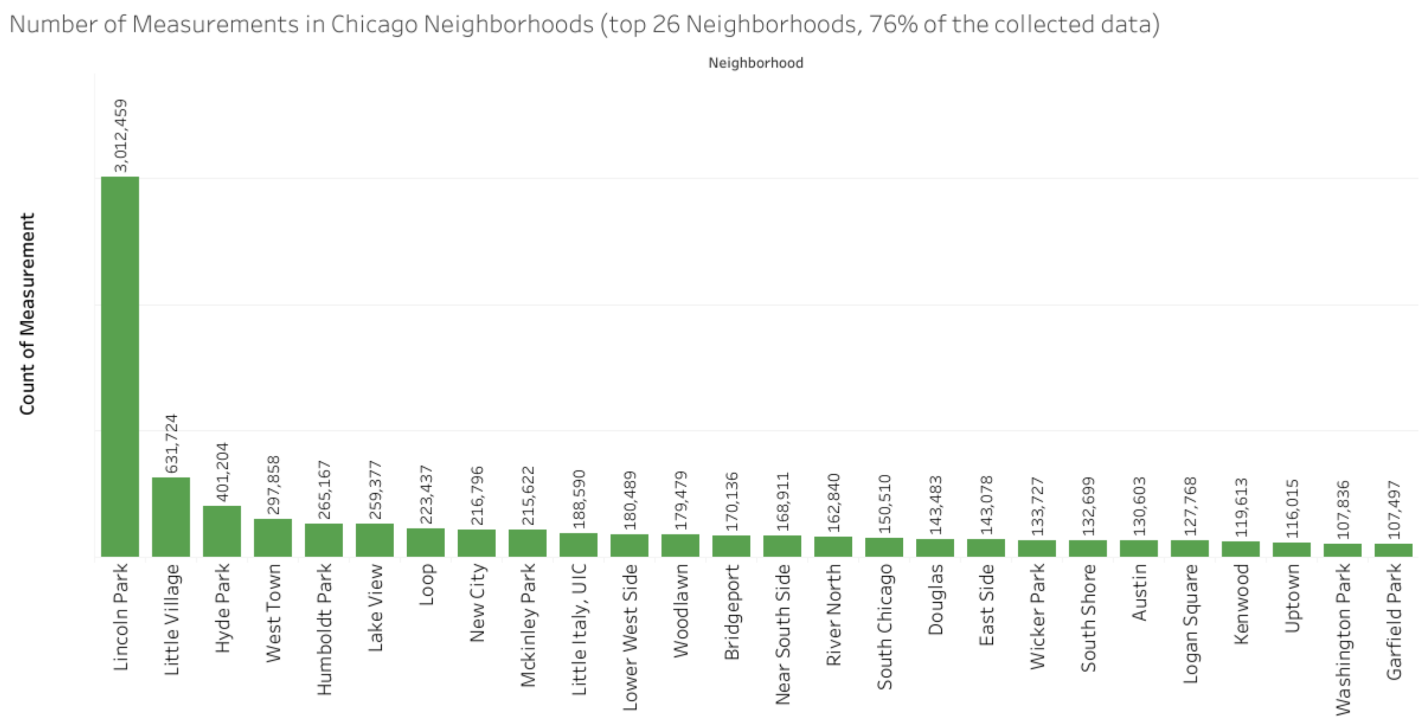

Figure 4: Measurements collected by ELPC team, by the top 26 neighborhood, 2017-2021, accounting for 76% of all readings. Shows the most data was collected in Lincoln Park, followed by Little Village, Hyde Park, West Town, Humboldt Park, and Lakeview.

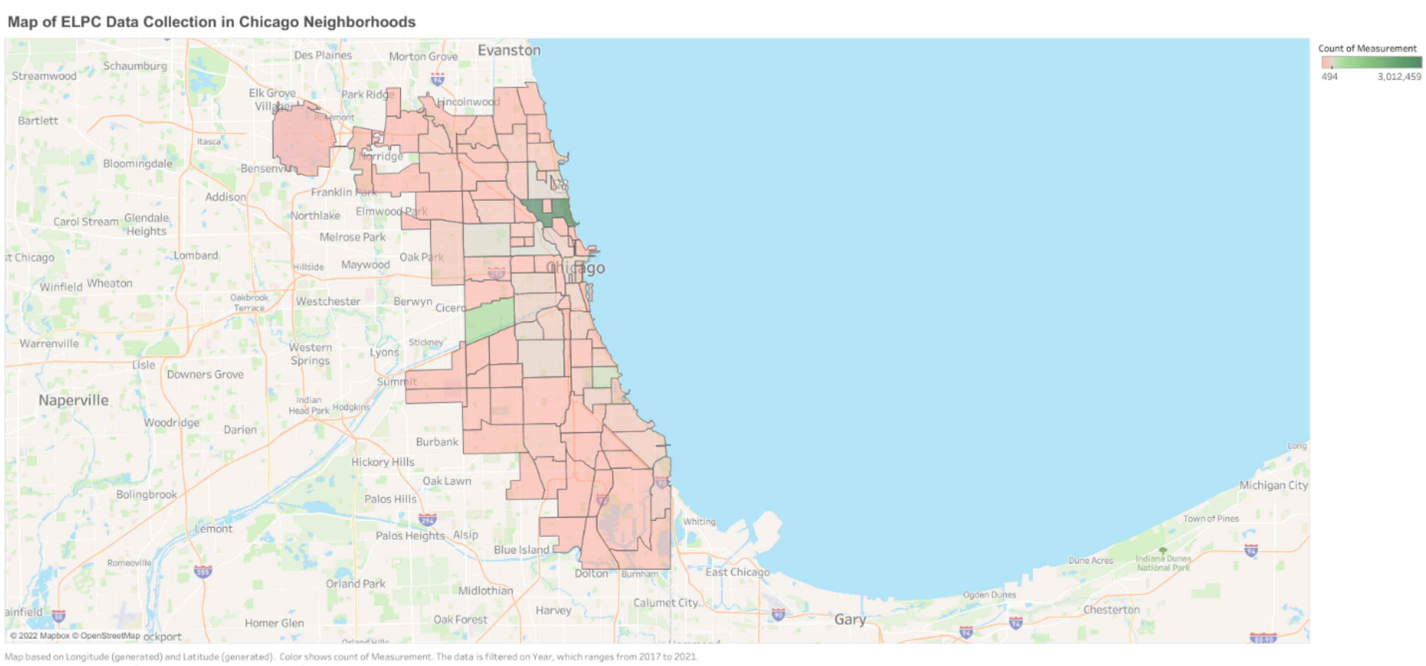

Figure 5: ELPC data map shows the most documented neighborhoods in green, and least documented neighborhoods in red. Click to see an interactive map: [Map of ELPC Data Collection in Chicago Neighborhoods.]

Figures 6 and 7, for instance, show the coverage and density (i.e., greater sampling) of ELPC's mobile air quality monitoring in the Little Village neighborhood, respectively.

Little village Actual

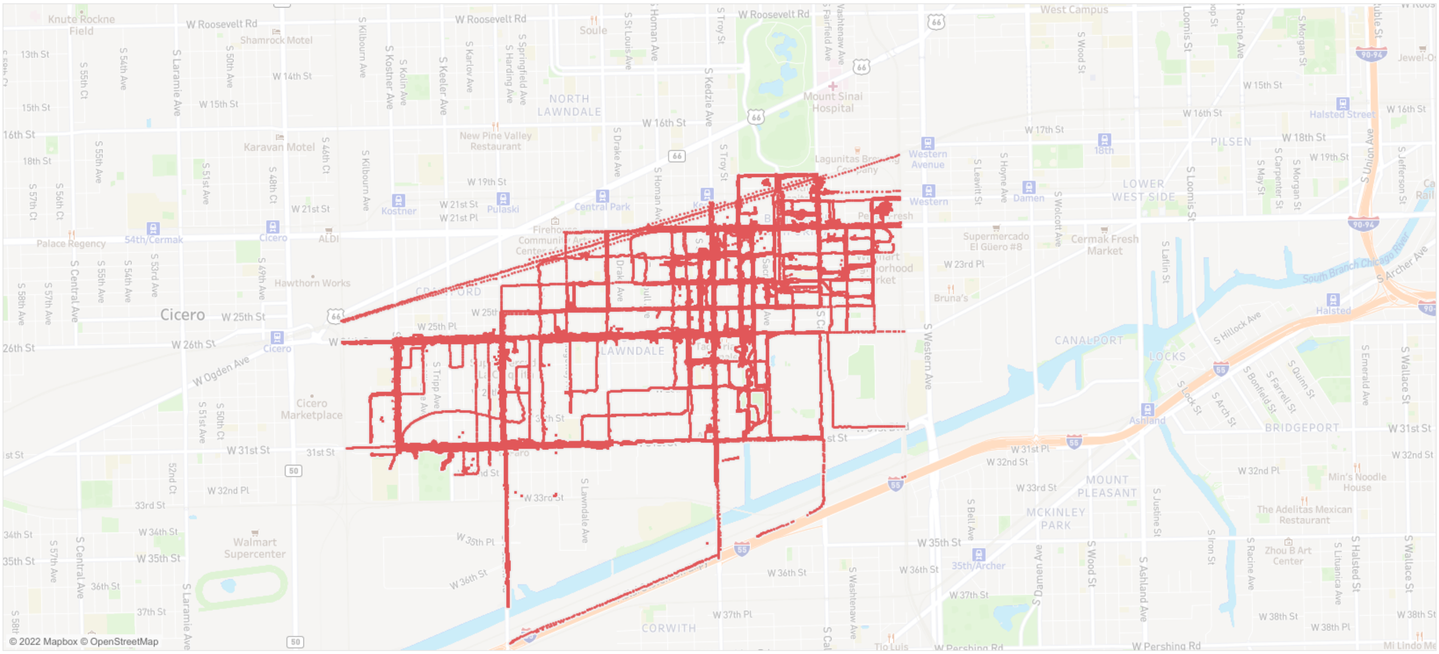

Figure 6: Map of Little Village neighborhood shows the geographic spread of Airbeam data collected by the ELPC team. Each red dot represents a measurement collected, and together they form a grid pattern tracing the neighborhood's streets and sidewalks.

Little village Density

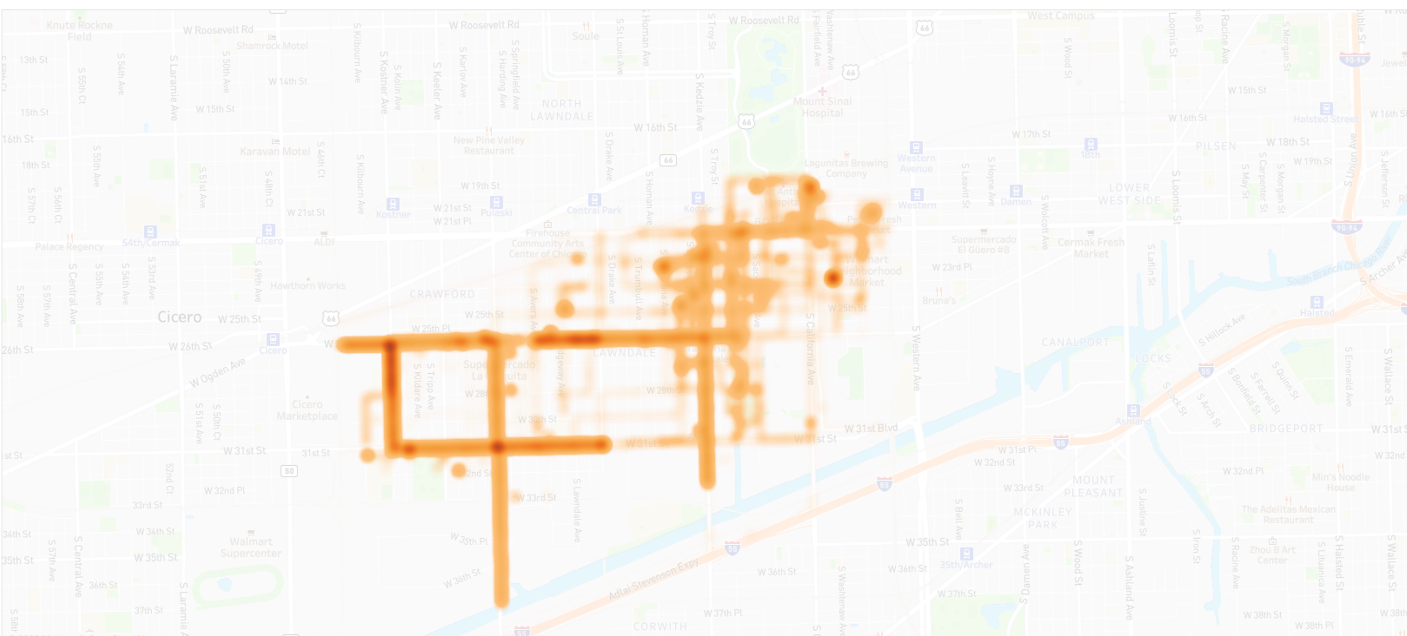

Figure 7: Map of Little Village neighborhood shows the density of Airbeam data collected by the ELPC team, showing the most data was collected on Pulaski, Kostner, 31st and 26th streets, along with a cluster just south of Douglass Park.

Microsoft Stationary Sensors Data

It is necessary to continuously monitor the air quality in order to comprehend the temporal variability of the air quality. This is obviously a difficult task for mobile air quality monitoring. Due to the fact that ELPC's air quality program was completely voluntary, there were also greater limits on consistently implementing a routine plan to cover the location and time of sampling. As a result, the required number of repetitions to accurately gauge the air quality at the hyperlocal level and identify hotspots was not always reached.

Fortunately, since July 2021, the Project Eclipse Network (Daepp et al., 2022), owned and operated by Microsoft, has been continuously monitoring the air quality for three pollutants, $PM_{2.5}$, O_3 and NO_2 in Chicago with over 100 stationary sensors. The initiative makes use of a low-cost urban air quality sensing network and is through a partnership with the City of Chicago, the Array of Things Project, JCDecaux Chicago, the Environmental Law and Policy Center, and regional environmental justice organizations in the area. Figure 8 shows a map of the stations captured on March 30, 2022, at 6:57 PM.

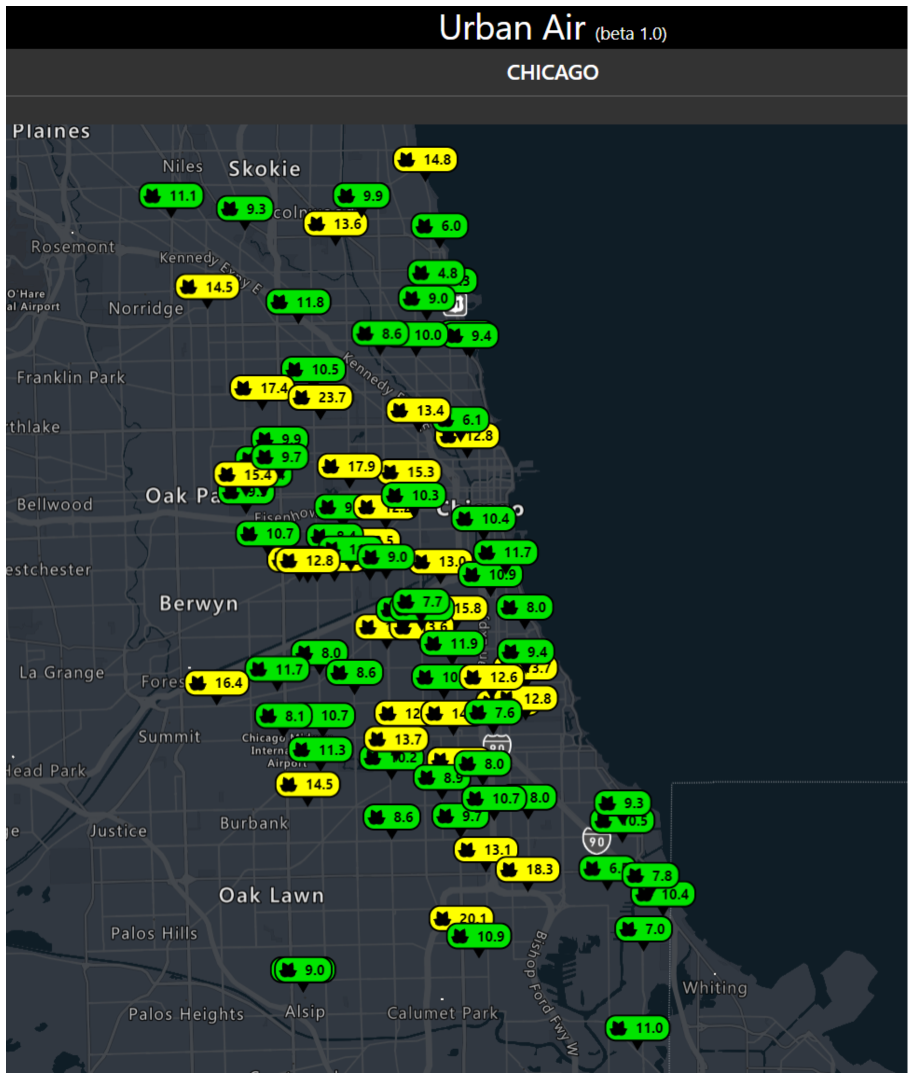

Figure 8: Stationary monitoring network, managed by Microsoft Research's Project Eclipse, in partnership with the City of Chicago, the Array of Things Project, JCDecaux Chicago, and local environmental organizations. Screenshot captured March 30, 2022, at 6:57pm.

The recorded data from Microsoft stations in this analysis helps capture the temporal variability of air quality in Chicago across different time scales, including daily and annually. As of December 31, 2022, over 15 million data points have been recorded by the stations.

Spatiotemporal Variability Assessment

The Microsoft data is available via the Spatiotemporal Asset Catalog (STAC) API provided by Planetary Computer. We selected the entire dataset of 122 stations in 2022, where 10.23 million data points were recorded. However, not all stations operated 100% of the time in 2022, and the percentages varied significantly. According to statistical guidance articles, analyses that have over 10% missing data are likely to exhibit bias (Dong & Peng, 2013; Jakobsen et al., 2017). To minimize bias in our analysis, we retained only the data from the stations with operating hours over 90% in 2022, leaving us with 67 stations. We divided the analysis of these 67 stations into daily and annual periods.

For the daily analysis, we calculated the daily $PM_{2.5}$ of all the stations and filtered those above the current daily threshold of 35 µg/m³. Our results show that in 2022, three stations located in South Side Chicago exceeded this threshold on 50 days, all of which occurred in September and October. The map below shows the locations of these stations. Of concern is that on these 50 days, the $PM_{2.5}$ concentrations ranged from 68.9 to 226.4 µg/m³, with an average of 175.75 µg/m³, which is significantly above the daily threshold of 35 µg/m³.

It should be noted the EPA defines its daily PM2.5 standard as the 98th percentile of 24-hour concentrations, averaged over a 3-year period, with a threshold of 35 µg/m^3. In the context of this study, due to the availability of only one year of air quality data from Microsoft stations, a simplified approach was adopted. Instead of applying the 98th percentile method, each day's average daily PM2.5 concentration that exceeded 35 µg/m^3 was reported for each site. It is important to emphasize that while this method captures every exceedance of daily PM2.5 levels, it does not align directly with the EPA's traditional compliance assessment method. The choice was made in light of data constraints. This study was primarily aimed at raising awareness about the instances of elevated daily PM2.5 levels during the monitored year.

Figure 9: Map of Chicago shows three stationary monitoring stations exceeded the EPA daily standard 35 µg/m³ on Chicago's South Side, meaning these areas are not in compliance with federal regulatory standards.

For the annual analysis, we calculated the $PM_{2.5}$ levels for all 67 stations with more than 90% operating hours. None of these stations had an annual $PM_{2.5}$ above the current threshold of 12 µg/m³.

Furthermore, we examined the monthly $PM_{2.5}$ variation across different stations to identify any specific areas of the city with heightened monthly pollution levels. Figure 10 displays the monthly $PM_{2.5}$ levels exceeding 12 µg/m³ at various stations. The absence of the first three months of the year in the figure suggests that these months are generally cleaner. However, as the weather warms up, more stations begin to exhibit monthly $PM_{2.5}$ medians above 12 µg/m³. Table 1 reveals that 33 stations had a monthly $PM_{2.5}$ median exceeding 12 µg/m³ in July. As depicted in Figure 11, the dispersion of these stations throughout the city implies that air quality across the entire city in July might be somewhat worse than during the rest of the year. With 21 stations, August is the second-most polluted month. It is also worth noting from the map in Figure 11 that the majority of the stations are situated on the South and West Sides of Chicago.

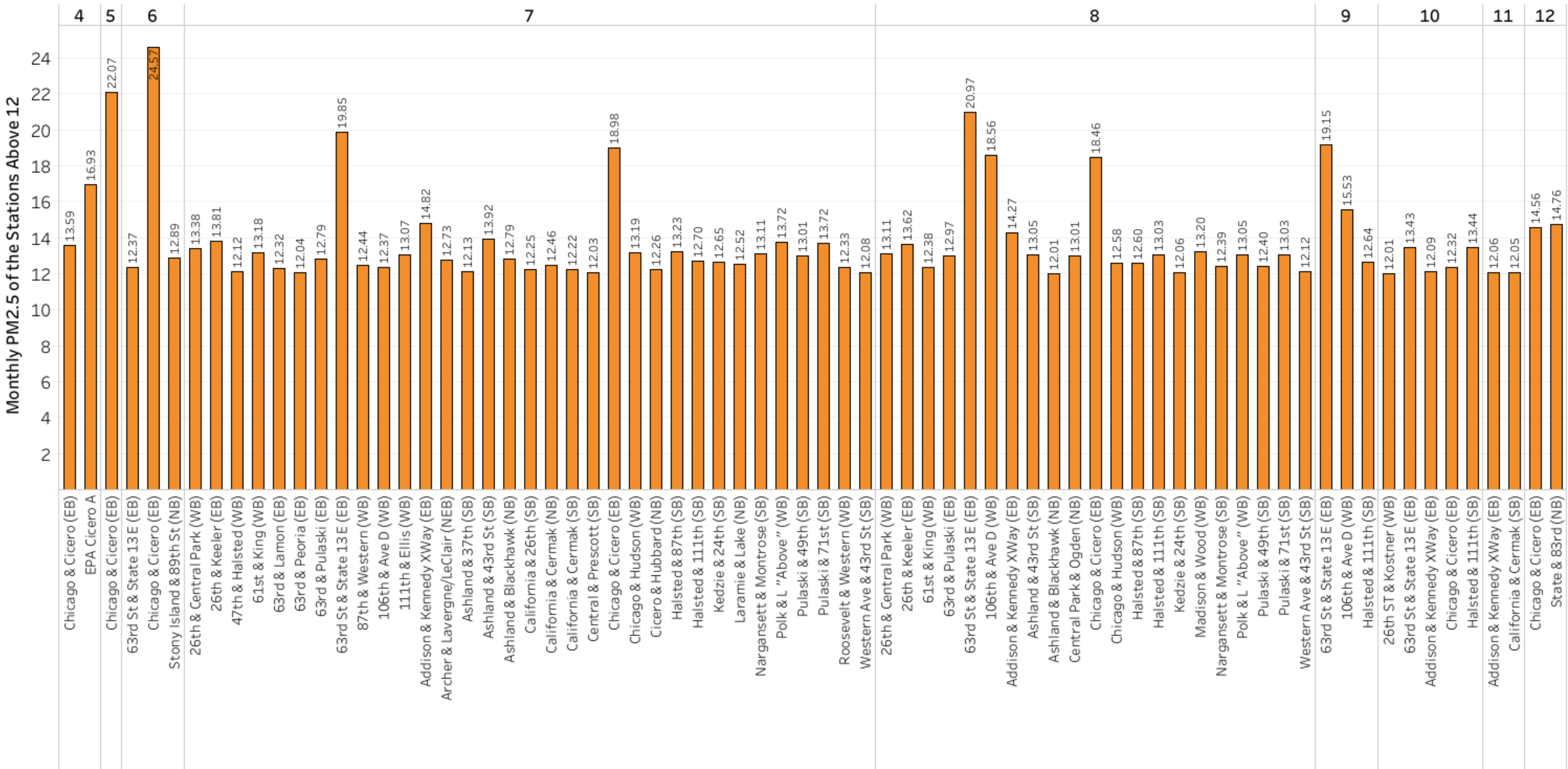

Figure 10: Stationary monitor locations where particulate matter readings exceeded 12 μg/m³, per month. Months are listed by numbers across the top, bus stops where the monitors were installed are listed along the bottom. Shows major exceedances at Chicago & Cicero, 63rd & State, and 106th & Ave D in particular.

Station Name/Month	April	May	June	July	August	September	October	November	December	Count
106th & Ave D (WB)				1	1	1		1		4
111th & Ellis (WB)				1						1
26th & Central Park (WB)				1	1					2
26th & Keeler (EB)				1	1					2
26th ST & Kostner (WB)							1			1
47th & Halsted (WB)				1						1
61st & King (WB)				1	1					2
63rd & Lamon (EB)				1						1
63rd & Peoria (EB)				1						1
63rd & Pulaski (EB)				1	1					2
63rd St & State 13 E (EB)			1	1	1	1	1			5
87th & Western (WB)				1						1
Addison & Kennedy XWay (EB)				1	1		1	1		4
Archer & Lavergne/LeClair (NEB)				1						1
Ashland & 37th (SB)				1						1
Ashland & 43rd St (SB)				1	1					2
Ashland & Blackhawk (NB)				1	1					2
California & 26th (SB)				1						1
California & Cermak (NB)				1						1
California & Cermak (SB)				1				1		2
Central & Prescott (SB)				1						1
Central Park & Ogden (NB)					1					1
Chicago & Cicero (EB)	1	1	1	1	1		1		1	7
Chicago & Hudson (WB)				1	1					2
Cicero & Hubbard (NB)				1						1
Cottage Grove & 49th St (SB)							1			1
EPA Cicero A	1									1
Halsted & 111th (SB)				1	1	1	1	1		5
Halsted & 87th (SB)				1	1					2
Kedzie & 24th (SB)				1	1					2
Laramie & Lake (NB)				1						1
Madison & Wood (WB)					1					1
Nargansett & Montrose (SB)				1	1					2
Polk & L "Above" (WB)				1	1					2
Pulaski & 49th (SB)				1	1					2
Pulaski & 71st (SB)				1	1					2
Roosevelt & Western (WB)				1						1
State & 83rd (NB)									1	1
Stony Island & 89th St (NB)			1							1
Western Ave & 43rd St (SB)				1	1					2
Total	**2**	**1**	**3**	**33**	**21**	**3**	**6**	**4**	**2**	**75**

Table 1: Stationary monitor locations where particulate matter readings exceeded 12 µg/m³ throughout the month.

Shows the most locations had poor air quality in July. Also shows that there were consistent problems at Chicago & Cicero, where air levels were high for seven months of the year. Stations where PM exceeded 12 µg/m³ for five months include 63rd & State and 111th & Halsted. Stations where PM exceeded for four months include 106th & Ave D and Addison & Kennedy XWay.

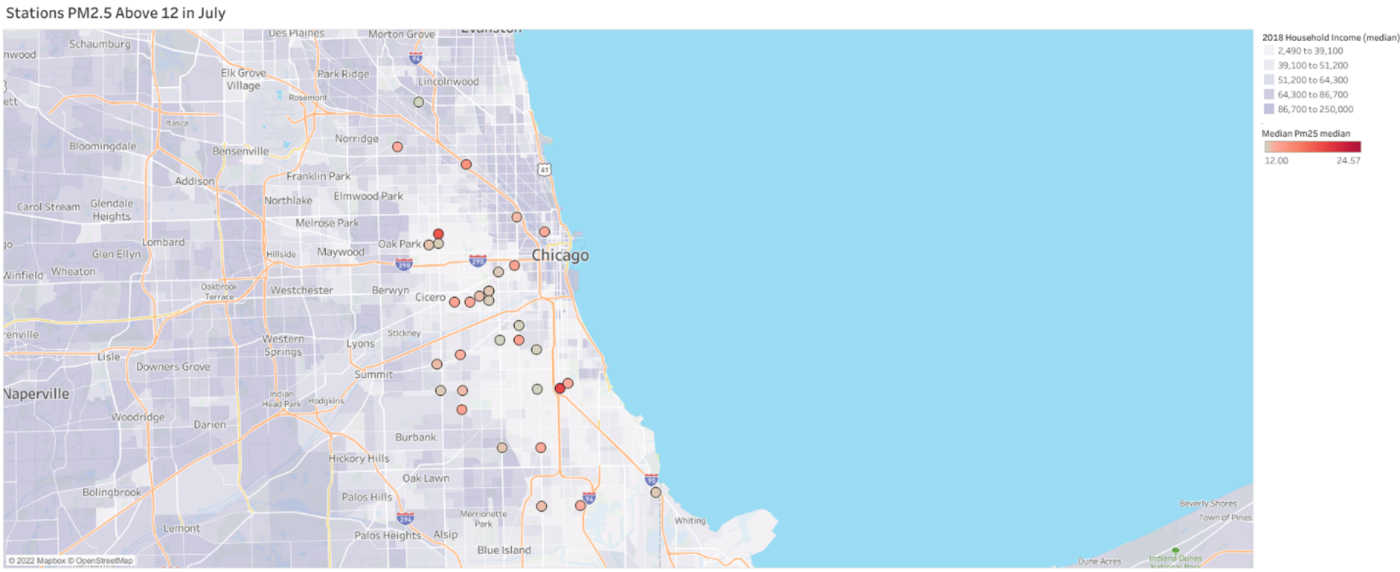

Figure 11: Chicago map shows all the stations that exceeded 12 µg/m³ in July, with brighter red dots indicating higher levels than grey dots. Purple layer across Chicagoland indicates median household income in that area.

We also analyzed the spatial and temporal variability of $PM_{2.5}$ across all data collected in Chicago at diverse time intervals, such as hours of the day, days of the week, and months of the year. This encompasses the entirety of measurements taken within the city during specific time intervals. This analysis will enable us to design a more effective mobile air quality monitoring system and furnish us with a comprehensive understanding of how pollution fluctuates across various time frames and locations. Figure 12 displays the hours-of-the-day variation of $PM_{2.5}$ for the entire dataset, encompassing all stations collectively. There is no substantial distinction in the median values across different hours; however, the period from 12 PM to 2 PM exhibits slightly elevated pollution levels. Figure 13 presents the day-of-the-week variation of $PM_{2.5}$. Similar to the hours-of-the-day variations, no significant distinctions in the values are observed; however, Tuesdays and Wednesdays tend to experience higher pollution levels. Figure 14 illustrates the variations in $PM_{2.5}$ levels across different months of the year. Notably, the two months with the highest pollution levels are July and August.

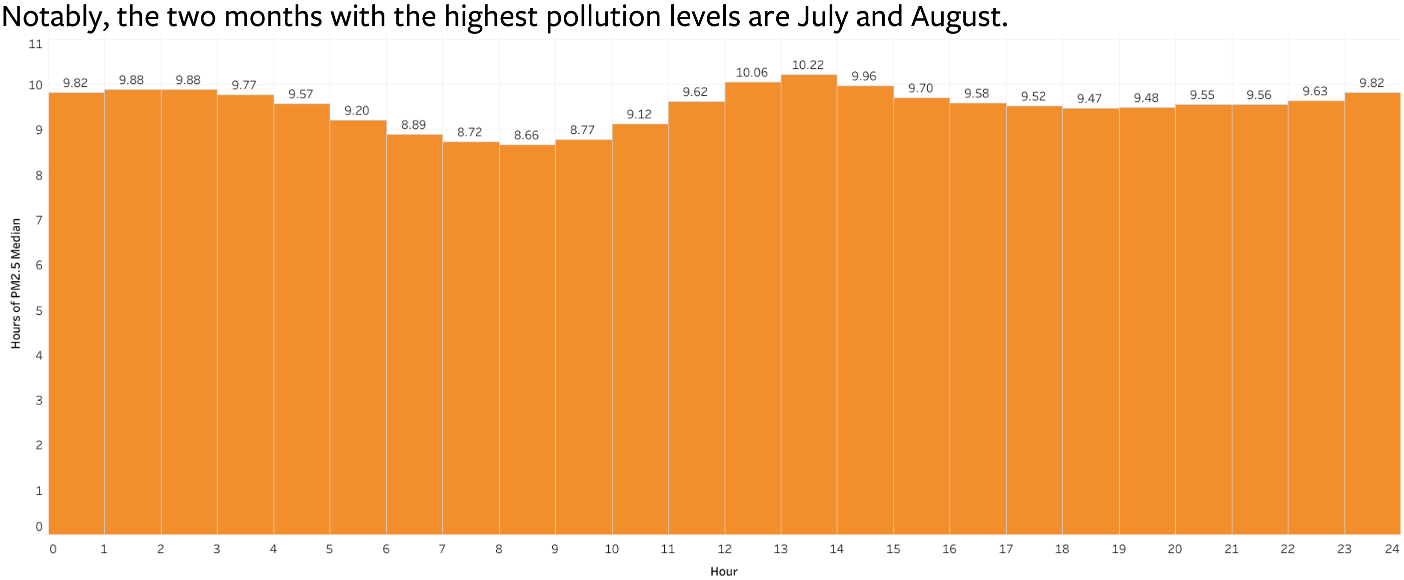

Figure 12: Median PM2.5 readings during each hour of the day across the stationary monitor network. Shows no substantial distinction between hours, but there is a slight improvement in air quality between 5am-11am, and a slight rise in the early afternoon.

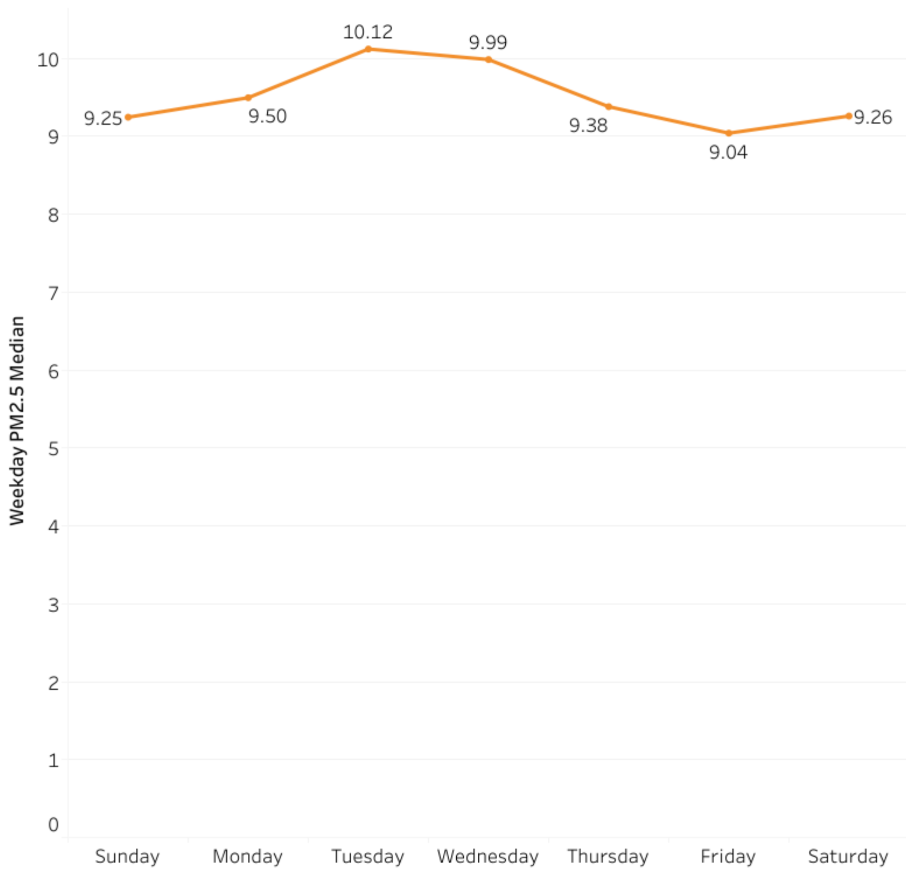

Figure 13: Median PM2.5 readings during each day of the week, across the stationary monitor network. Shows no substantial distinction between days, but there's a slight improvement on Sunday and Friday, while Tuesday and Wednesday have a bit more PM.

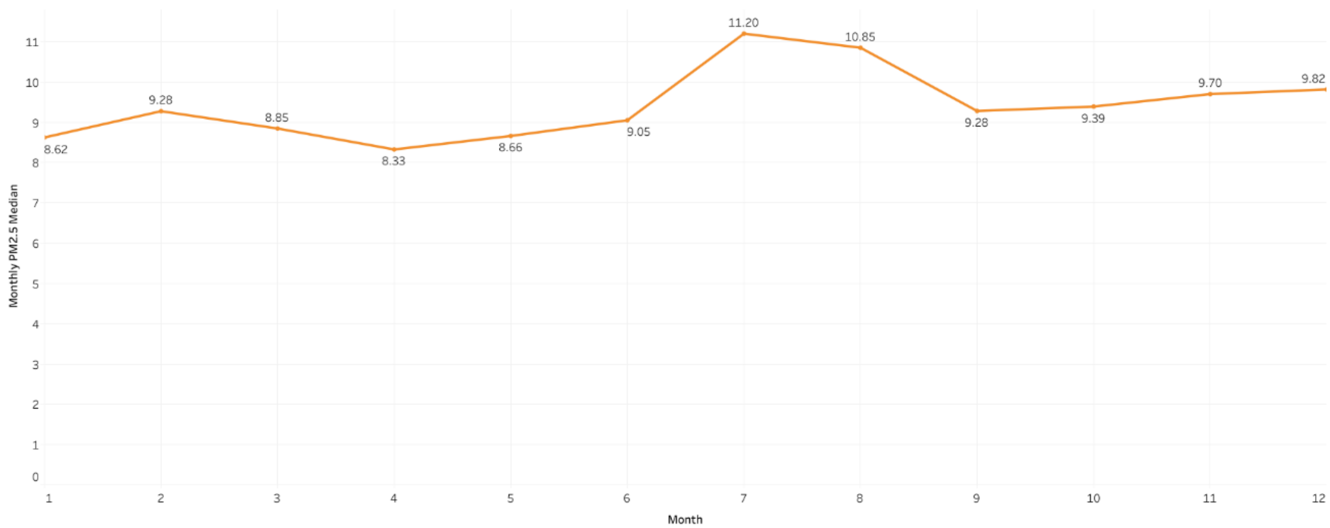

Figure 14: *Median PM2.5 readings during each month of the year, across the stationary monitor network. Shows no substantial distinction between months, but slightly higher PM levels in July and August.*

Random Data Sampling: A Method to Quantify the Number of Sampling Repeats

As previously discussed, determining the number of repeat visits needed to achieve a meaningful approximation of the spatial air quality pattern in a specific area is a crucial consideration for mobile monitoring. The challenge arises from the fact that the target air pollution variable and locational factors can influence the required number of repetitions. As a result, Chicago may not consistently benefit from suggestions made by studies conducted in other regions.

In this section, we describe the development of a method to calculate the number of Chicago-specific sampling repetitions for mobile monitoring, utilizing the data from Microsoft stations. As previously mentioned, stationary sensors are ideal for measuring the temporal variability and long-term patterns of air pollution, which allows us to learn from the station data and develop an efficient and effective approach for mobile monitoring. We started by assuming that each station represents the air quality of a block with dimensions of 420 feet by 680 feet, equivalent to the average size of a city block in Chicago. For each station, we selected a range of one to a hundred hours of data, ensuring that no two hours of data originated from the same day of the year. The primary goal of this approach is to determine the point at which we can be confident that we have achieved a certain level of accuracy regarding the annual $PM_{2.5}$ at a given location in mobile monitoring, by comparing the samples in that block to the annual $PM_{2.5}$ of that block. This process involves incrementally sampling one hour a day for up to a hundred days and evaluating the progression of the data collected. As a reminder, we utilized data from 67 stations that had over 90% operation time in 2022 to minimize bias in measuring their annual $PM_{2.5}$.

Moreover, mobile monitoring at the Environmental Law & Policy Center (ELPC) has typically been conducted during the warmer months of the year, mainly due to factors such as volunteer availability and the limited functionality of AirBeam devices under extreme weather conditions. To account for this, we divided random sampling into two separate yearly subsets of data: one spanning from May to October, and the other from June to August, both restricted to the hours between 8 AM and 8 PM to closely align with the actual implementation of mobile monitoring. Subsequently, we investigated the point of repetition at which sampling accuracy was most reliable within these timeframes.

To assess the performance and proximity of the sampling repeats to the annual $PM_{2.5}$, we employed three different metrics: Firstly, for each sample of every station, we measured the absolute error percentage between the sample median $PM_{2.5}$ and the station's annual $PM_{2.5}$. Next, we calculated the average absolute error percentage for all 67 stations for each sample size, ranging from one to a hundred.

$$E_i(x) = |(s_i - a_i) / a_i| \times 100$$
$$A(x) = (1/n) \times \Sigma E_i(x) \text{ for } i=1 \text{ to } n$$

Where:

- $E_i(x)$: Absolute error percentage for station i with sample size x
- s_i: Sample median $PM_{2.5}$ for station i
- a_i: Annual $PM_{2.5}$ for station i
- x: Sample size (ranging from 1 to 100)
- A(x): Average absolute error percentage for all stations with sample size x
- n: Number of stations (in this case, n = 67)

In addition to the annual $PM_{2.5}$ of the stations, another piece of information to consider is their standard deviation. Standard deviation is an important statistical measure because it provides insight into the dispersion or spread of a data set. By quantifying the degree of variation from the mean, it helps in assessing the data's consistency, reliability, and potential outliers. The annual $PM_{2.5}$ values for all the stations range from 7.52 to 11.15 $\mu g/m^3$. Apart from two stations with notably large standard deviations of 37.85 and 42.65, the remaining 65 stations exhibit standard deviations between 2.27 and 4.07, which still constitute a significant proportion of their annual $PM_{2.5}$ values. Consequently, the first metric used to compare the samples and annual $PM_{2.5}$ levels is insufficient for a comprehensive analysis of sampling accuracy. Therefore, we also employ the second metric, the coefficient of variation (CV), to further assess the data.

The coefficient of variation (CV) (Brown, 1998) is a valuable statistical measure that quantifies the dispersion of a data set relative to its mean. Expressed as a percentage, CV is particularly useful when comparing the variability of data sets with different units of measurement or disparate means. By standardizing variability as a proportion of the mean, the coefficient of variation enables meaningful comparisons between data sets that may otherwise be difficult to assess. Lastly, the third metric, Quartile Coefficient of Dispersion (QCD) (Bonett, 2006) a measure of relative dispersion, specifically for non-parametric data or data sets with outliers. It is based on the interquartile range (IQR), which represents the range between the first quartile (25th percentile) and the third quartile (75th percentile) of a data set. QCD is calculated by dividing the IQR by the sum of the first and third quartiles and is expressed as a decimal value or a percentage. Both CV and QCD are measures of relative dispersion. We follow the same method for calculating the mean absolute error percentage for the coefficient of variation (CV) and the quartile coefficient of dispersion (QCD) as we did for the median. For CV:

$$CV_i(x) = (\sigma_i(x) / \mu_i(x)) \times 100$$

$$E\_CV_i(x) = | (CV_i(x) - CV\_annual_i) / CV\_annual_i | \times 100$$

$$CV\_avg(x) = (1/n) \times \Sigma\, E\_CV_i(x) \text{ for } i=1 \text{ to } n$$

Where:

- $CV_i(x)$: Coefficient of variation for station i with sample size x
- $E\_CV_i(x)$: Absolute error percentage for station i with sample size x
- $\sigma_i(x)$: Standard deviation of $PM_{2.5}$ for station i with sample size x
- $\mu_i(x)$: Mean of $PM_{2.5}$ for station i with sample size x
- $CV\_avg(x)$: Average coefficient of variation for all stations with sample size x
- n: Number of stations (in this case, n = 67)

and for QCD:

$$QCD_i(x) = (Q_{3i}(x) - Q_{1i}(x)) / (Q_{3i}(x) + Q_{1i}(x))$$

$$E\_QCD_i(x) = | (QCD_i(x) - QCD\_annual_i) / QCD\_annual_i | \times 100$$

$$QCD\_avg(x) = (1/n) \times \Sigma\, E\_QCD_i(x) \text{ for } i=1 \text{ to } n$$

Where:

- $QCD_i(x)$: Quartile Coefficient of Dispersion for station i with sample size x
- $E\_QCD_i(x)$: Absolute error percentage for station i with sample size x
- $Q_{1i}(x)$: First quartile of $PM_{2.5}$ for station i with sample size x
- $Q_{3i}(x)$: Third quartile of $PM_{2.5}$ for station i with sample size x
- $QCD\_avg(x)$: Average Quartile Coefficient of Dispersion for all stations with sample size x
- n: Number of stations (in this case, n = 67)

The performance assessment results, based on three metrics for both data subsets, are illustrated in Figures 15, 16, and 17. It is important to note that the number of unique days for the June to August subset is capped at 92, while the May to October subset has sufficient days to reach 100. When comparing the mean absolute error percentages of the medians, a clear separation between them is not evident as sample sizes increase, until nearly 50 where the blue line begins to display smaller errors. In contrast, the other two metrics—CV and QCD, which indicate dispersion—reveal a different narrative. As observed, the longer the data subset (i.e., May to October), the less dispersion there is between the samples and the annual $PM_{2.5}$. For CV, the May to October subset consistently falls below 20% error at approximately 25 samples, while for the June to August subset, this occurs at roughly 45 samples. Regarding QCD, the May to October subset drops below 20% at nearly 20 samples, whereas for the June to August subset, considerably more sampling, close to 60, is required.

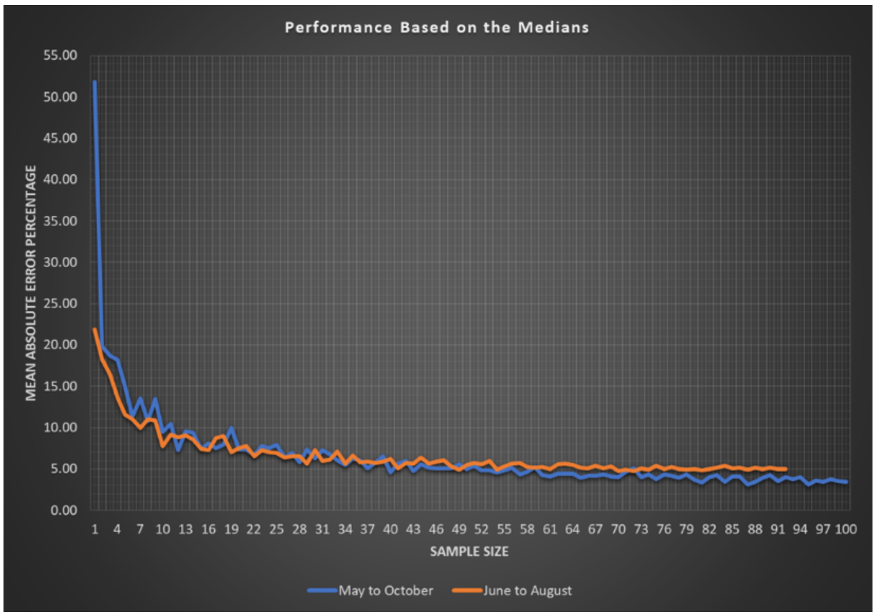

Figure 15: Performance assessment results, based on the median

For the May to October subset, the average error percentage experiences a significant decrease after approximately five samples, with a $PM_{2.5}$ error of nearly 15% (accompanied by a 30% dispersion error). As the number of repetitions increases, the error further reduces: between 10 and 15 repetitions (with dispersion errors ranging from 26% to 20%), the error remains around 10%; between 15 and 30 repetitions (with dispersion errors from 20% to 15%), it averages about 7.5%; and from 30 to 100 repetitions, the error varies from 6% to 3.5% (with dispersion errors between 15% and 10%).

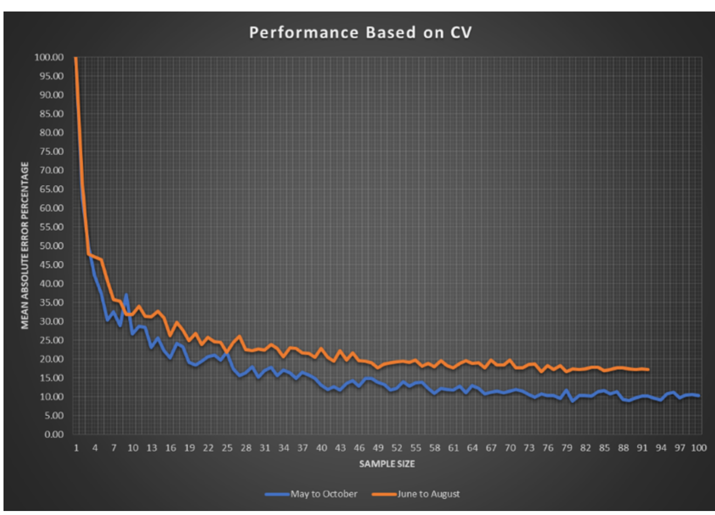

Figure 16: Performance assessment results, based on the coefficient of variation (CV)

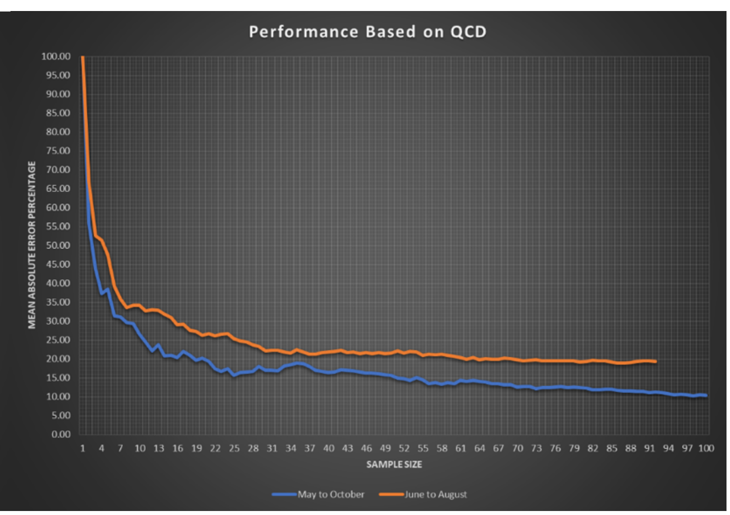

Figure 17: Performance assessment results, based on the quartile coefficient of dispersion

Exploring the Impact of City Zoning on $PM_{2.5}$ Concentrations: The Role of Parks and Open Spaces

In this air quality research project, we investigate the variations in particulate matter ($PM_{2.5}$) levels across different city zoning. The study encompasses 67 sensor stations, each providing annual $PM_{2.5}$ data, and examines the spatial distribution of nine zoning types: Business, Commercial, Downtown, Residential, Manufacturing, Planned Development (PD), Planned Manufacturing Development (PMD), Parks and Open Spaces (POS), and Transportation. We generated buffers at radii of 1/4 mile, 1/2-mile, 1 mile, and 2 miles around each sensor station, subsequently calculating the area percentages of each zoning category within these buffers.

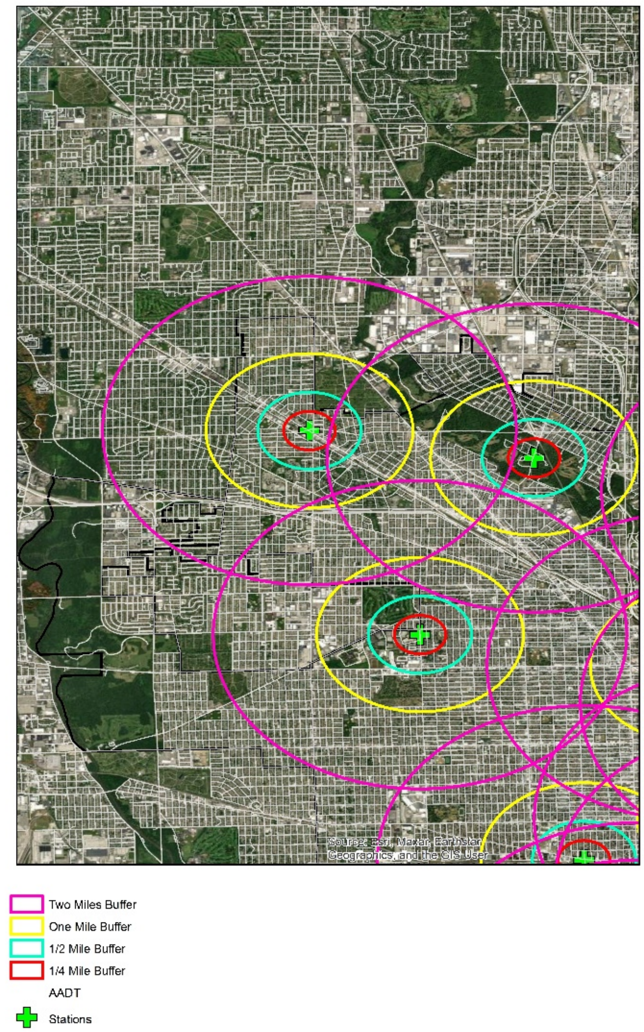

Figure 18: Screenshot of a city map showing stationary monitors as green plus signs, with buffer rings drawn around each. Within these rings, we calculated the proportion of each zoning type and measured the correlation with the air quality recorded at that station.

To establish potential correlations between $PM_{2.5}$ concentrations and zoning areas, we employed Pearson's correlation coefficient as our statistical measure, using a significance level of P=0.05. This approach enabled us to quantitatively assess the strength and direction of the relationships between the variables under consideration.

Our findings provide valuable insights into the spatial distribution of $PM_{2.5}$ in relation to city zoning and traffic patterns. We found that annual $PM_{2.5}$ levels were significantly correlated only with the Parks and Open Spaces (POS) zoning category. At a 1/4-mile radius, the correlation coefficient was -0.384, indicating a moderately negative relationship between $PM_{2.5}$ levels and the percentage of POS areas. Interestingly, this correlation became slightly stronger at a 1/2-mile radius with a coefficient of -0.427, suggesting that the influence of POS on $PM_{2.5}$ concentrations was more pronounced within this distance. However, as the radius increased to 1 mile, the correlation weakened to -0.358.

Although correlation doesn't imply causation, these findings imply that the presence of Parks and Open Space zoning might be a contributing factor in reducing PM2.5 concentrations especially within a half-mile radius. Beyond this distance, the correlation begins to weaken, suggesting that the benefits of POS in mitigating air pollution diminish as the radius expands. This observation highlights the importance of strategically placing parks and open spaces in urban areas to maximize their potential for improving air quality.

Analysis of Mobile Air Quality Monitoring Data: A Confidence Map

We utilized the findings from the previous sections to analyze data obtained through our mobile air quality monitoring efforts. To accomplish this, we first generated rectangular grid cells, measuring 420 by 680 feet, across the Chicago area using the ArcGIS Fishnet Tool. Subsequently, we employed a spatial join to match the points from the air quality measurement layer to the target grid cell polygon output layer. For each grid cell, we counted the number of unique measurement days and calculated the median $PM_{2.5}$ value of all the recorded measurements. Grid cells with fewer than five unique days were excluded, as the limited number of measurements was insufficient to accurately represent the air quality within those cells. The remaining grid cells were then categorized as follows:

- **5 to 9: Unconfident** – A good start, but additional sampling days are needed
- **10 to 14: Relatively Confident** – Very close, but a few more sampling days are needed
- **15 to 29: Confident** - No immediate need for more sampling, but additional data could increase confidence
- **30 and above: Very Confident** – No further sampling is necessary

Figure 19 presents a grid cell in which we have high confidence in the measured $PM_{2.5}$ value of 19 μg/m³. Figure 20 depicts the results of the analysis in which we have covered a 35.7-square-mile area within Chicago, which represents 15.2% of the city's total land area of 234.5 square miles. For a more detailed examination, you can access the interactive map here: [Interactive Map.]

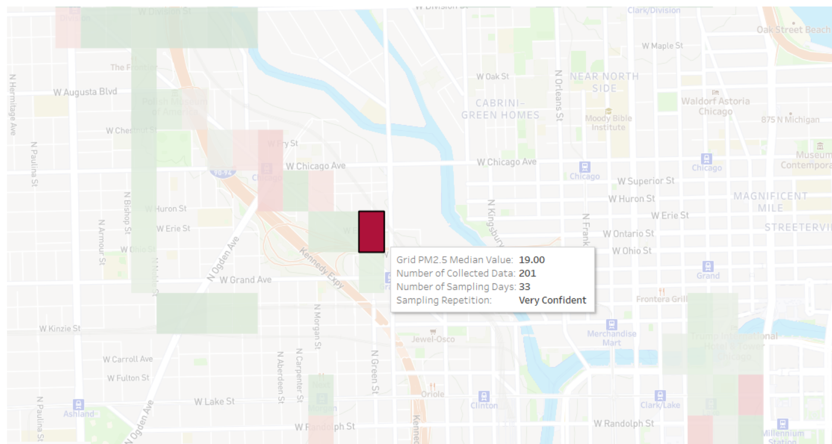

Figure 19: Map indicates a grid cell above 19 μg/m³ west of Halsted Ave, between Ohio and Erie

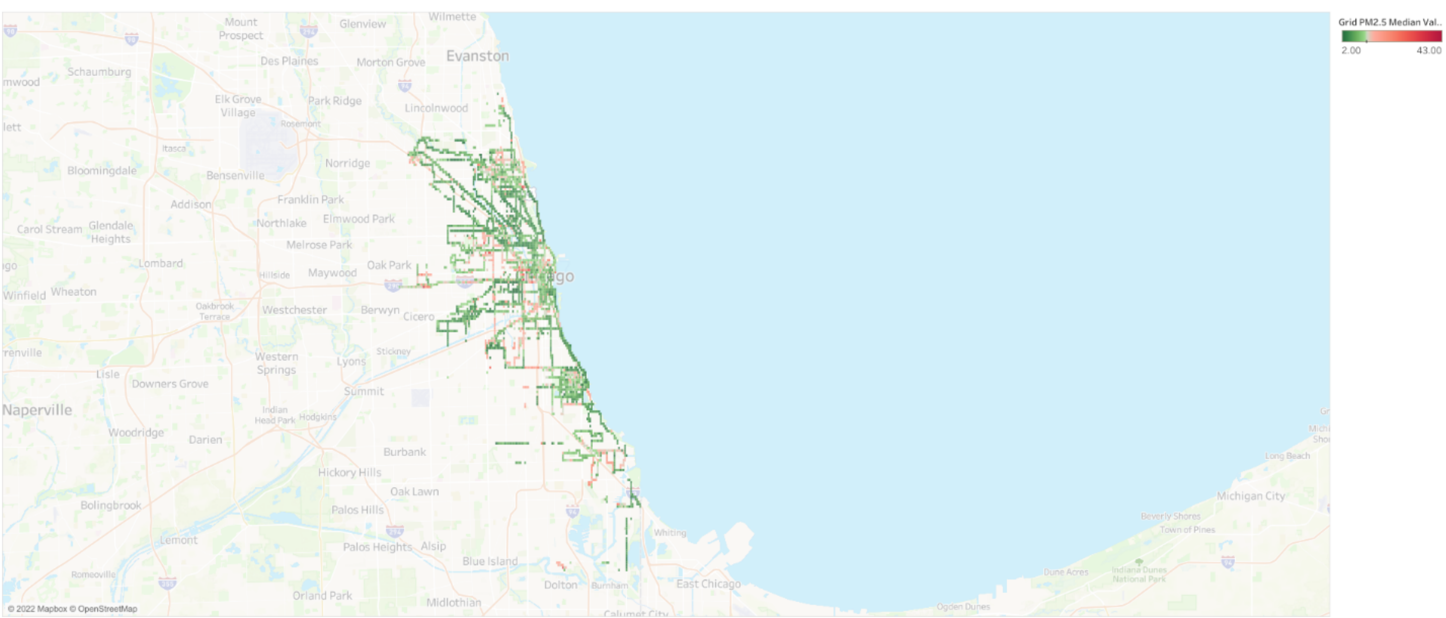

Figure 20: Full grid cell map of Chicagoland confidence map. Red indicates unhealthy air. See full [interactive map] here.

Recommendations:

Here are some suggestions to improve mobile air quality monitoring based on the insights gained from this study:

- To ensure a consistent sampling rate, conduct monitoring only while **walking**.
- **Number of repeats:** For annual $PM_{2.5}$ levels, **a minimum of 30 hours per route** is recommended, spread over 30 different days of the year. Divide the route into two halves, with each half covered in 30 minutes, while the second 30 minutes is spent returning to the starting point. Choose a route so that every point is measured twice, regardless of the route length.
- **Hours:** Sampling can be done at any hour, but it should not be limited to a single hour of the day. Spread sampling between 8 A.M. and 8 P.M.
- **Days:** Sampling can be conducted on any day of the week, but it should not be limited to a single day of the week.
- **Months:** Spreading data collection between May and October increases the likelihood of obtaining results closer to the annual $PM_{2.5}$ levels at the locations.
- To identify shorter term pollution levels such as daily $PM_{2.5}$ spikes above the EPA threshold (35 μg/m^3), it is more likely to observe such days in July, during Tuesdays and Wednesdays, around noon to afternoon. Spatially, there is a higher likelihood of finding hotspots on the West Side and South Side of Chicago. However, the South Side has also exhibited daily spikes in September and October.
- **Location:** Hotspots are more likely to be found in areas without **parks** within a **half-mile radius.**
- Data from the station at the **intersection of W Chicago Ave and N Cicero Ave** indicates higher pollution levels than at other stations. We recommend further mobile monitoring in this area to gain additional insights.
- From the provided **Interactive Map**, based on the provided map, continue monitoring the ″Unconfident″, ″Relatively Confident″, "Confident" areas with $PM_{2.5}$ levels above 12 μg/m^3. Ensure that the minimum requirement of 30 hours-30 days is met for each of these locations. By doing so, you can increase the accuracy and confidence of the measurements in these areas, leading to a better understanding of the air quality and potential hotspots.

References

Aclima. (2020). *New Research Reveals Hyperlocal Air Pollution Varies Over Space & Time*. https://www.aclima.io/blog/new-research-expands-understanding-of-how-hyperlocal-air-pollution-varies-over-space-and-time-78e82168bd52

Anjomshoaa, Amin et al. "City Scanner: Building and Scheduling a Mobile Sensing Platform for Smart City Services." IEEE internet of things journal 5.6 (2018): 4567–4579. Web.

Apte, Joshua S et al. "High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data." Environmental science & technology 51.12 (2017): 6999–7008. Web.

Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. *Computational statistics & data analysis*, *50*(11), 2953-2957.

Brown, C. E. (1998). Coefficient of variation. In *Applied multivariate statistics in geohydrology and related sciences* (pp. 155-157). Springer.

Cabral, Alex, Asta Roseway, and Paul Johns. "Design and Implementation of A Mobile Urban Low-Cost Environmental Sensor Network." 2021 IEEE Sensors. IEEE, 2021. 1–4. Web.

Cohen, Aaron J et al. "Estimates and 25-Year Trends of the Global Burden of Disease Attributable to Ambient Air Pollution: An Analysis of Data from the Global Burden of Diseases Study 2015." The Lancet (British edition) 389.10082 (2017): 1907–1918. Web.

Chen, Yanju et al. "A New Mobile Monitoring Approach to Characterize Community-Scale Air Pollution Patterns and Identify Local High Pollution Zones." Atmospheric environment (1994) 272 (2022): 118936–. Web.

Chen, Y., Gu, P., Schulte, N., Zhou, X., Mara, S., Croes, B. E., Herner, J. D., & Vijayan, A. (2022). A new mobile monitoring approach to characterize community-scale air pollution patterns and identify local high pollution zones. *Atmospheric Environment*, *272*, 118936.

Church, Zach. "The Psychological, Economic, and Social Costs of Air Pollution." MIT Sloan, MIT, 3 Feb. 2020, mitsloan.mit.edu/ideas-made-to-matter/psychological-economic-and-social-costs-air-pollution.

Cummings, Lucas E. et al. "Mobile Monitoring of Air Pollution Reveals Spatial and Temporal Variation in an Urban Landscape." Frontiers in built environment 7 (2021): n. pag. Web.

Daepp, M. I., Cabral, A., Ranganathan, V., Iyer, V., Counts, S., Johns, P., Roseway, A., Catlett, C., Jancke, G., & Gehring, D. (2022). Eclipse: an end-to-end platform for low-cost, hyperlocal environmental sensing in cities. 2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN),

deSouza, Priyanka et al. "Air Quality Monitoring Using Mobile Low-Cost Sensors Mounted on Trash-Trucks: Methods Development and Lessons Learned." Sustainable cities and society 60 (2020): 102239–. Web.

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*, 1-17.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, *17*(1), 1-10.

Fruin, S. et al. "Measurements and Predictors of on-Road Ultrafine Particle Concentrations and Associated Pollutants in Los Angeles." Atmospheric environment (1994) 42.2 (2008): 207–219. Web.

Gupta, Smarth, et al. "The Dangers in Our Air: Mapping Chicago's Air Pollution Hotspots." Chicago Sun-Times, 16 May 2022, chicago.suntimes.com/2022/5/16/23069860/dangerous-air-we-breathe-mapping-chicago-air-pollution-hotspots-pm2-5-particulate-matter.

Hankey, Steve, and Julian D. Marshall. "On-Bicycle Exposure to Particulate Air Pollution: Particle Number, Black Carbon, PM2.5, and Particle Size." Atmospheric environment (1994) 122 (2015): 65–73. Web.

Hudda, Neelakshi et al. "Emissions from an International Airport Increase Particle Number Concentrations 4 fold at 10 Km Downwind." Environmental science & technology 48.12 (2014): 6628–6635. Web.

Jiang, Xu-Qin, Xiao-Dong Mei, and Di Feng. "Air Pollution and Chronic Airway Diseases: What Should People Know and Do?" Journal of thoracic disease 8.1 (2016): E31–E40. Web.

Li, H. et al. "Spatially Dense Air Pollutant Sampling: Implications of Spatial Variability on the Representativeness of Stationary Air Pollutant Monitors." Atmospheric Environment: X 2 (2019): n. pag. Print.

"NAAQS Table." EPA, Environmental Protection Agency, 5 Apr. 2022, www.epa.gov/criteria-air-pollutants/naaqs-table.

Peltier, Richard E. "Top EPA Advisers Challenge Long-Standing Air Pollution Science, Threatening Americans' Health." PBS, 21 Apr. 2019, www.pbs.org/newshour/science/top-epa-advisers-challenge-long-standing-air-pollution-science-threatening-americans-health.

Shakya, Kabindra M. et al. "Mobile Monitoring of Air and Noise Pollution in Philadelphia Neighborhoods During Summer 2017." Environmental pollution (1987) 255.Pt 1 (2019): 113195–. Web.

"Ambient (Outdoor) Air Pollution." World Health Organization, World Health Organization, 22 Sept. 2021, www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.