Air Pollution Trends in the San Francisco Bay Area

Rivkah Gardner-Frolick

GIS in Water Resources
Fall 2017

Introduction

This GIS project focuses on spatial analysis of data gathered during a campaign to map air pollution using Google Street View cars as mobile air quality monitors. The campaign is a result of collaboration between Google, Aclima, Environmental Defense Fund, and Dr. Joshua Apte's group at University of Texas, Austin in the Civil, Environmental, and Architectural Engineering Department. Google provides the cars and hires the drivers, Aclima develops and maintains the instruments, and Environmental Defense Fund provides part of the funding and manages the project.

The project began in May 2015 and is currently ongoing. The first portion of sampling focused on Oakland, California and lasted a year. During this time, three main areas or "polygons" of Oakland were extensively sampled and intentionally oversampled. The purpose was to be able to use Oakland data to determine how to best design and analyze other studies with mobile monitoring, something that has not been done to this extent before. The second part of sampling is still taking place in the San Francisco Bay Area. Other polygons, representing a variety of land use and climate zones, were selected for driving (Figure 1). Each polygon is completely driven a set number of times to be considered "fully covered," ranging from about 15-40 complete drives. Eventually, the data will provide the basis for a land use regression model (LUR). LURs predict air pollution based on land use and the Bay Area model will be one of the few to use such fine scale spatial data.

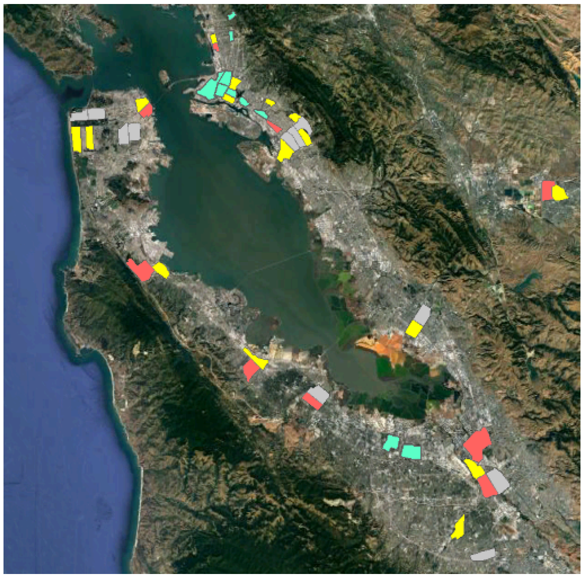


Figure 1: Group 1 (completed polygons) in light blue, (focus areas) in light red, Group 3 (possible additions) in yellow, and Group 4 (deprioritized) in gray

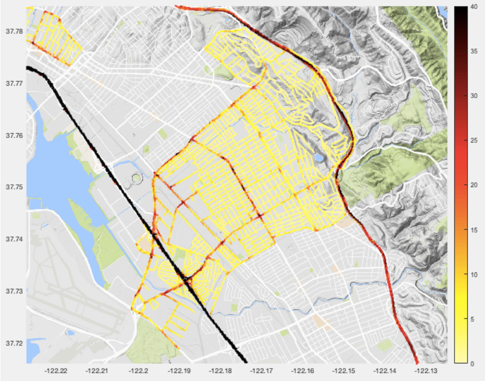


Figure 2: data points for one polygon Group 2

Land Use	Polygon Count
High-density commercial	3
High-density mixed residential/commercial	5
High-density residential	6
Medium-density residential	13
Medium-density commercial	4
Medium-density industrial	1
Medium-density mixed commercial/industrial	2
Medium-density mixed residential/commercial	10
Medium-density residential/industrial	2

Table 1: Polygon Land use

Climate Zone	Climate Description	Polygon Count
14	Northern California's inland areas with some ocean influence	2
15	Chilly winters along the Coast Range	9
16	Central and Northern California Coast thermal belts	2
17	Marine effects in Southern Oregon, Northern and Central California	33

Table 2: Polygon climate area

Project Overview

One feature of the dataset that jumped out upon initial visual analysis was that there are fine scale (~30m) air pollution hot spots, or areas of high concentrations relative to the surroundings. Hot spots at such a fine scale have not yet been thoroughly investigated and a systematic means of identification has not been developed. In the initial paper from the Google Street View mapping, Apte et al. identified hot spots based on a code written by Dr. Kyle Messier that identified statistically significant elevated concentrations.[1] A few points were then checked using Google Street View imagery of the road conditions and buildings to ascertain if the hot spot was reasonable. Hot spots are one of the interesting features that come from such fine scale pollution mapping. They could contribute greatly to human exposure and most air quality maps simply do not have enough resolution to pick up on the anomalies. As a result, epidemiology studies focused on air pollution have had to rely on sweeping estimates for exposure. Between mobile monitoring and other advances in technology like smartphones, estimation of air pollution exposure and its health effects could become much more accurate.

In this project, GIS is used to spatially analyze the Google Street View air quality data that has been collected. This analysis will focus on small and large spatial patterns. One of the most prominent spatial patterns is the fine scale hot spots discussed in the previous paragraph. Relying on manual checks of hot spots is not satisfactorily accurate and a new method of hot spot identification is needed. This is where ArcGIS Pro tools could be useful in developing a rigorous and less time-consuming method of identification. GIS tools will be used in conjunction with another program, SatScan, to identify spatial patterns. The focus is on black carbon and nitric oxide because these two species are the direct result of combustion activities, can be quite variable, and are important air pollutants for health concerns. It is important to analyze multiple pollutants because points with elevated concentrations of more than one pollutant are more likely to be a true air pollution hot spot. Given the availability of data from multiple pollutants (only good spatial coverage with black carbon, nitric oxide, and nitrogen dioxide) and time constraints of this project, two pollutants were chosen.

Methods

Initially, the Google Street view dataset contained 16 million points. Dr. Kyle Messier performed corrections on the raw data and “snapped” each measured value to a 30m set of points along the drive routes. The purpose of this was to make the data easier to analyze and reduce.

[1] J. Apte, et al. Environ. Sci. Technol., 51, 6999-7008 (2017)

Data Visualization

As a first-cut analysis, the concentrations of black carbon and nitric oxide that resulted from the data snapping above were mapped. Since there were many data points with multiple values at each geographic point, a MATLAB GUI created by Dr. Messier was used to visualize the data. His GUI takes all of the data points and applies a user-input function to reduce the data at each point down to one representative value that is then overlaid on a Google street map. A visualization of relative concentrations gives polygons or points to investigate further.

Neighborhoods to Know

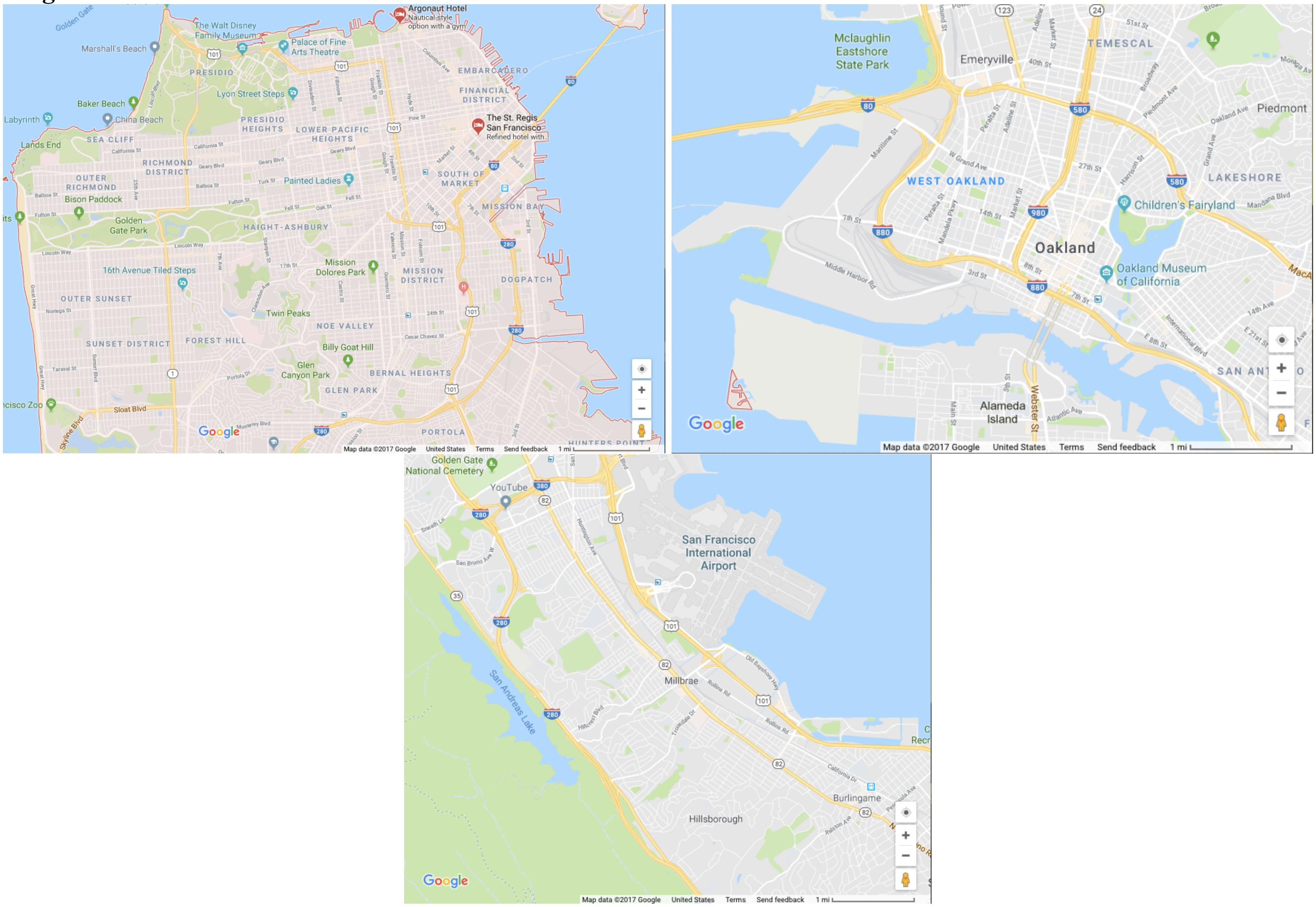


Figure 3: San Francisco Bay Area neighborhoods, will be referred to later

Black Carbon

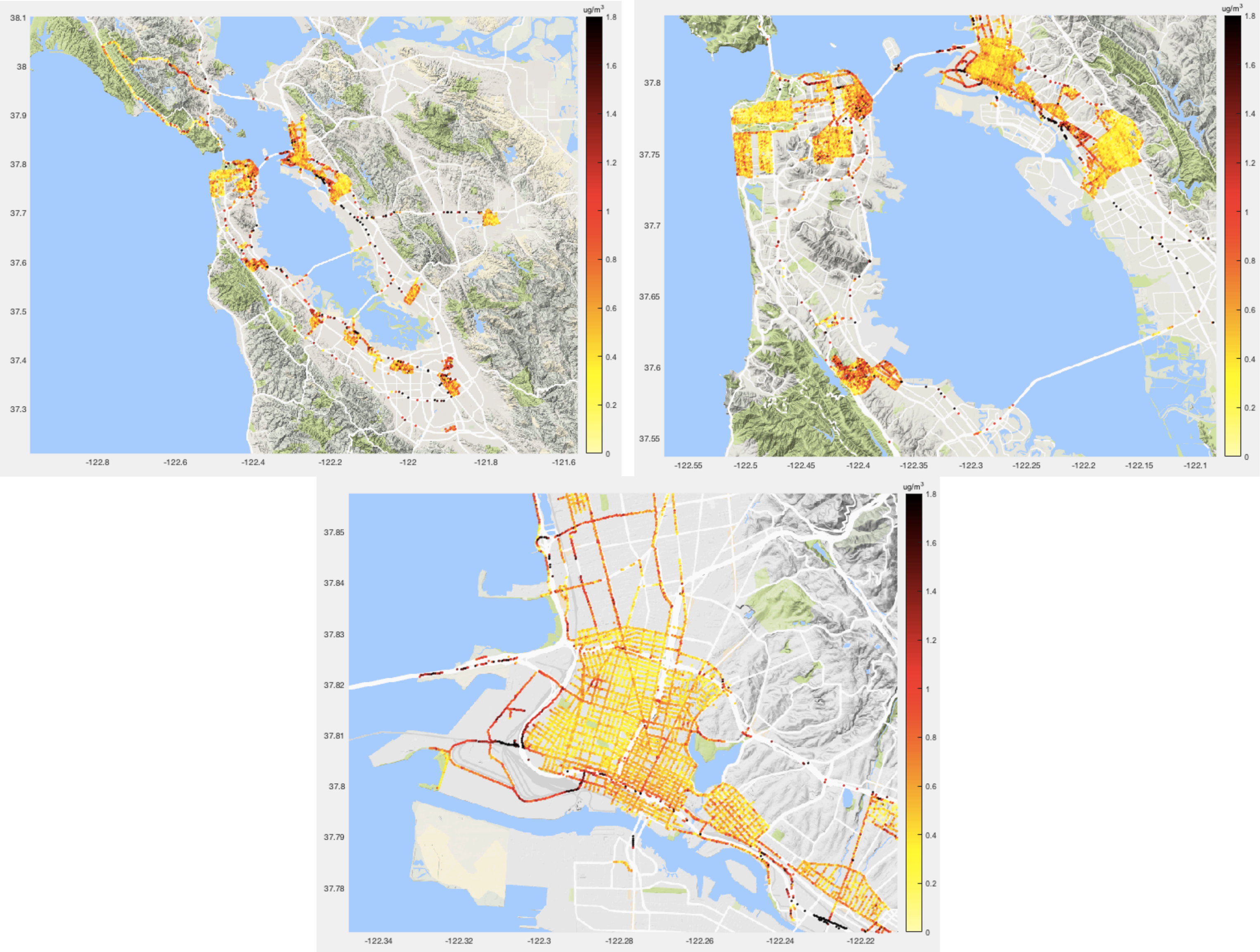


Figure 4: Black carbon concentrations

The data in the maps show some expected and some unexpected patterns. Obviously, areas near highways and major arterial roads have higher black carbon concentrations. Some areas of San Francisco that are expected to be relatively clean clearly have lower concentrations overall, such as Inner Sunset, Outer Sunset, and Richmond. The Financial District has relatively high concentrations. Interestingly, there are some unexpected spatial patterns, such as the high overall concentrations in Millbrae (a mainly residential area that should not have many point sources) and that Outer Sunset is dirtier than Inner Sunset (since Outer Sunset is Closer to the Pacific Ocean which should be bringing in clean air). These large-scale spatial patterns can be caused by different characteristics of each area, like traffic, land use, and predominant wind patterns. The wind pattern during this study will be analyzed to attempt to describe some patterns seen in this map.

Nitric Oxide

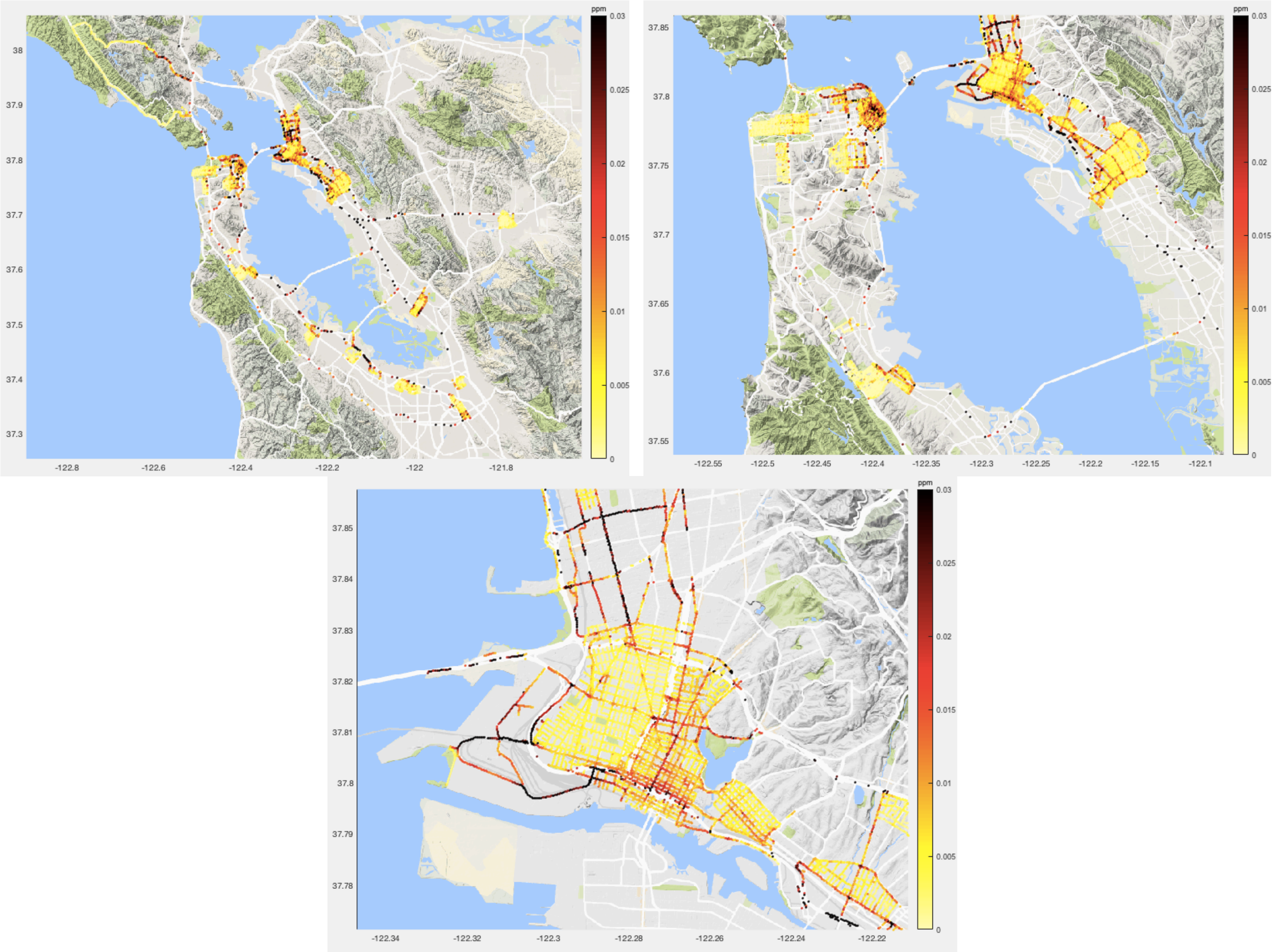


Figure 5: Nitric oxide concentrations

In the maps above it can be seen that nitric oxide is heavily linked to combustion emissions, mainly coming from traffic in many cases. For example, and area with a lot of traffic like the Financial District has high concentrations of nitric oxide. It can also be seen that it is a fairly short-lived pollutant. Large roads have concentrations that are clearly higher than the surrounding areas and there does not seem to be much dispersion from roadways as points adjacent to road with extremely high values can be quite clean themselves.

Data Reduction

In order to import the Google Street View into GIS, the data had to be reduced. A MATLAB code was written that took each snapped point and classified each measurement into a drive pass. Some locations have multiple data points from the car passing one time because the car may have been moving slowly or stopped at a light or stop sign. Since these data that are clustered in time are highly correlated and may skew the true mean of the measurements at a location, the mean of each drive pass was taken. Next, the median of all of the drive pass means at each snapped point was taken. The median usually gives a better representation of the average conditions with air pollution data such as this because one or two points could be abnormally high from passing by a transient point source, like a truck idling. However, if there are consistent point sources such as a truck, the high value will still be captured well with the median. To make sure the each snapped point analyzed is truly a representative measurement, all snapped points with less than six unique days of driving were taken out. Finally, all the highways from the dataset were removed because they were expected to have consistently high measurements and the main focus of this project is hot spots near places where people live or work. High values on highways next to residential areas could throw off the identification of hot spots on the neighborhood streets because they are high in comparison to the expected value but not in comparison to that of a highway.

ArcGIS Pro Tools

To investigate the hot spots in a systematic way, two GIS statistical tools were chosen. The Hot Spot Analysis (Getis-Ord Gi*) and the Cluster and Outlier Analysis (Anselin Local Moran's I) are two tools in ArcGIS Pro that are meant to take data sets and identify statistically significant hot and cold spots in spatial data.

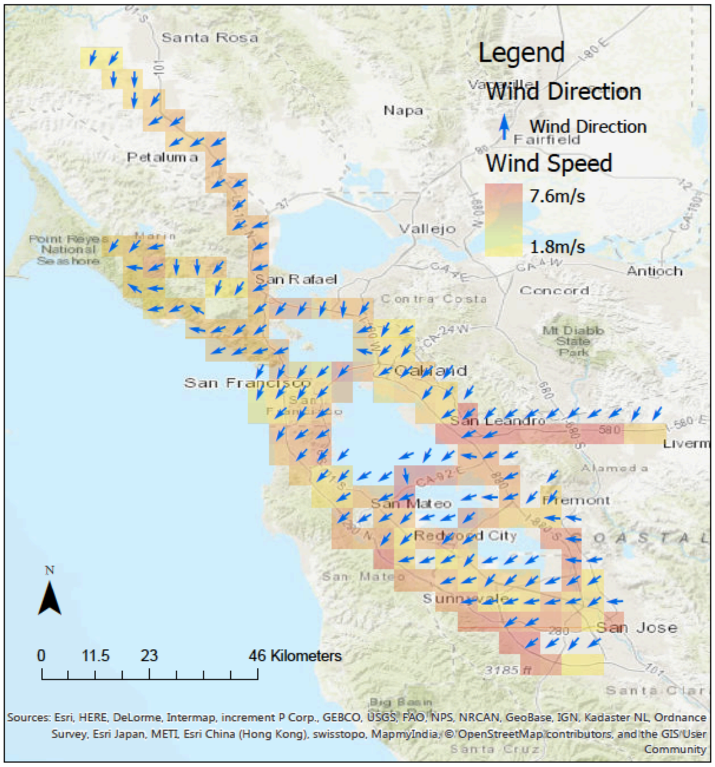
The Hot Spot Analysis tool uses the Getis-Ord local statistic to calculate a z-score and p-score for each point, which determines whether or not a point is in a hot spot. For this tool, a hot spot is a spatial cluster of high features in combination with low features of neighboring values. As inputs, this tool requires a distance band and spatial relationship to be applied. The distance band is the radius of a circle within which the adjacent points will be considered neighbors to the point being analyzed. The spatial relationship defines how the neighbors will be weighted in comparison to the point in question.

In this project, three different distance bands were used to identify local hot spots, polygon hot spots, and regional hot spots. For local hot spots, an inverse distance squared relationship (very strongly weighted close neighbors but weakly weighted far neighbors) was used as the spatial relationship. This was because the likely causes of local hot spots are local point sources like a lawnmower, which only influences the air in close proximity. To find the local hot spots a 200m radius was used. This circle would consist of about 12 neighbors if the point analyzed was on a street sampled outside of the driving polygons, such as a highway. A 200m band to identify local pollution points might be too large, hence the inverse distance squared weighting. The polygon and regional analyses were conducted with 1,000m and 3,000m distance bands with an inverse distance weighting. A different weighting was chosen because hyper-local pollution points were not the aim of hot spot identification. Larger distance bands, such as 10,000m, were also considered but when compared to the 3,000m distance band, it gave quite similar results and took much longer to run. The Hot Spot Analysis tool could potentially work better on the unsnapped data. However, this would prevent any data reduction and would likely crash ArcGIS with 16 million data points.

The other GIS tool chosen was the Cluster and Outlier Analysis (Anselin Local Moran's I). This tool finds spatial clusters of high or low values and outliers. It calculates the local Moran's I, a z-score, and a p-score to determine the clusters and outliers. Clusters are referred to as high (HH) or low (LL) and outliers are a high value surrounded by low values (HL) or a low value surrounded by high values (LH). The inputs are similar to those in the Getis-Ord-Gi* tool and an inverse distance squared distance relationship was chosen for an analysis of a 200m distance band. This distance band was chosen because of time constraints and where the most value in this analysis would lie.

Results and Discussion

Wind Map



Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 6: Wind speed and direction in study area

Wind patterns are an important part of air pollution transport and analysis of the wind patterns in this study could begin to explain some of the pollution patterns that are seen, or they could point to other pollution sources. To begin to get a sense of wind patterns in the Bay Area, a map was created to show the average wind speed and direction during driving. Each Google Street View car has wind measurement instruments attached. I took the average of the wind speed and wind direction measured at each snapped point. The arrows represent the wind direction and the color represents the wind speed for each raster cell. The wind direction and wind speed values at each snapped point were assigned to a raster cell and the average of each cell was calculated. The wind direction raster is represented as a vector field and the wind speed raster is represented as a continuous color scale.

Though the instruments were not highly reliable, they could give an estimation of the driving conditions. Additional skepticism is needed because the measurements were taken close to ground level and wind transport of pollutants occurs at many levels in the atmosphere. However, as a first-cut this map seems to be accurate enough for the purpose of this project. As a basic check, the general wind direction is toward the Pacific Ocean and on an average day in the Bay Area, this is indeed what happens. During the night, the wind usually switches direction and comes in from the Pacific. These two phenomena did not average out in the data collected because all driving is done during the day, for a variety of practical reasons.

This map is a rough explanation for unexpected pollution patterns seen in the data, namely the high concentrations in Millbrae. The wind, on average, appears to be blowing toward Millbrae from Oakland, and likely more importantly, from the Port of Oakland

and the San Francisco airport. Additionally, there is elevated topography to the West of the residential area that could serve to trap and accumulate this pollution. This also begs the question of why Inner Sunset and Outer Sunset are so clean given the predominant wind patterns. A reasonable explanation could be that the hills in the middle of San Francisco and the buildings work well to stop the transportation of pollution from the Financial District and other areas. There are many ground-level obstructions in this area of San Francisco, unlike the open waters to the East of Millbrae. From a first glance, the pollution patterns seen make sense.

Black Carbon

1) Using Hot Spot Analysis (Getis-Ord-Gi*):

Study Area

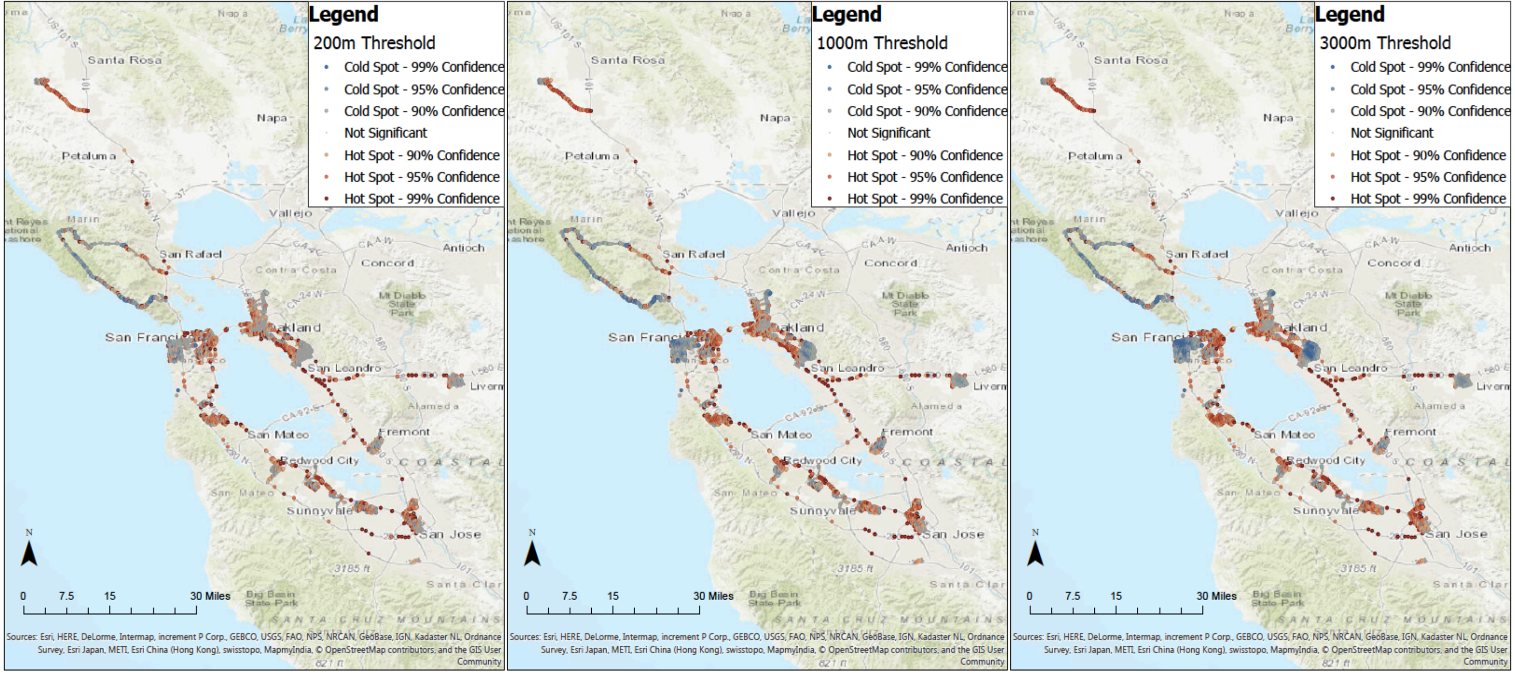

Legend
200m Threshold
1000m Threshold
3000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Santa Rosa
Napa
Petaluma
Marin
Vallejo
San Rafael
Antioch
Concord
Contra Costa
Mt Diablo State Park
San Francisco
Oakland
San Leandro
Livermore
Alameda
San Mateo
Fremont
Redwood City
Sunnyvale
San Jose
Santa Clara
Santa Cruz Mountains
N
0
7.5
15
30 Miles
Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 7: Hot Spot Analysis tool results in study area for black carbon

San Francisco and Surrounding Polygons

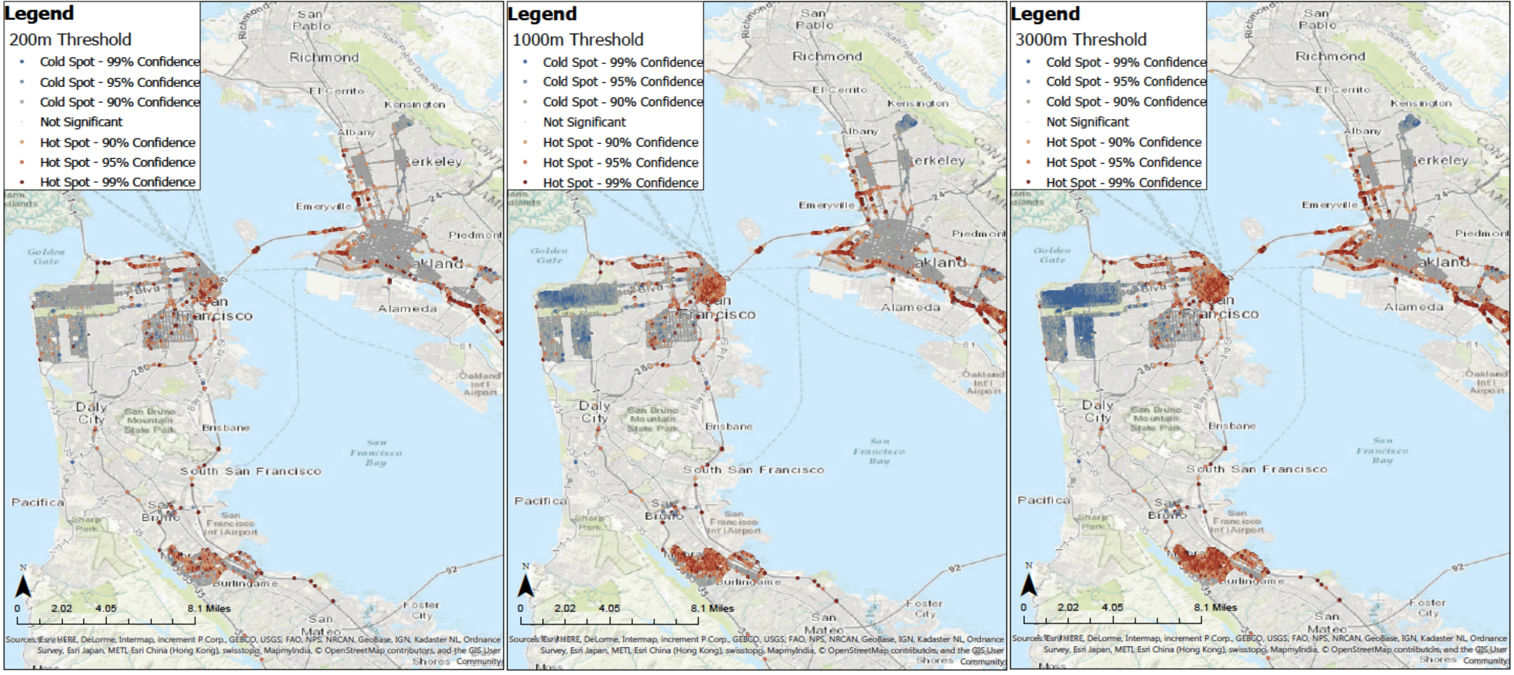

Legend
200m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
1000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
3000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
San Pablo
Richmond
El Cerrito
Kensington
Albany
Berkeley
Emeryville
Piedmont
Oakland
Alameda
Golden Gate
San Francisco
Daly City
San Bruno Mountain State Park
Brisbane
San Francisco Bay
South San Francisco
Pacifica
San Bruno
San Francisco Int'l Airport
Oakland Int'l Airport
Burlingame
Foster City
San Mateo
N
0 2.02 4.05 8.1 Miles
Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 8: Hot Spot Analysis tool results in San Francisco area for black carbon

Oakland

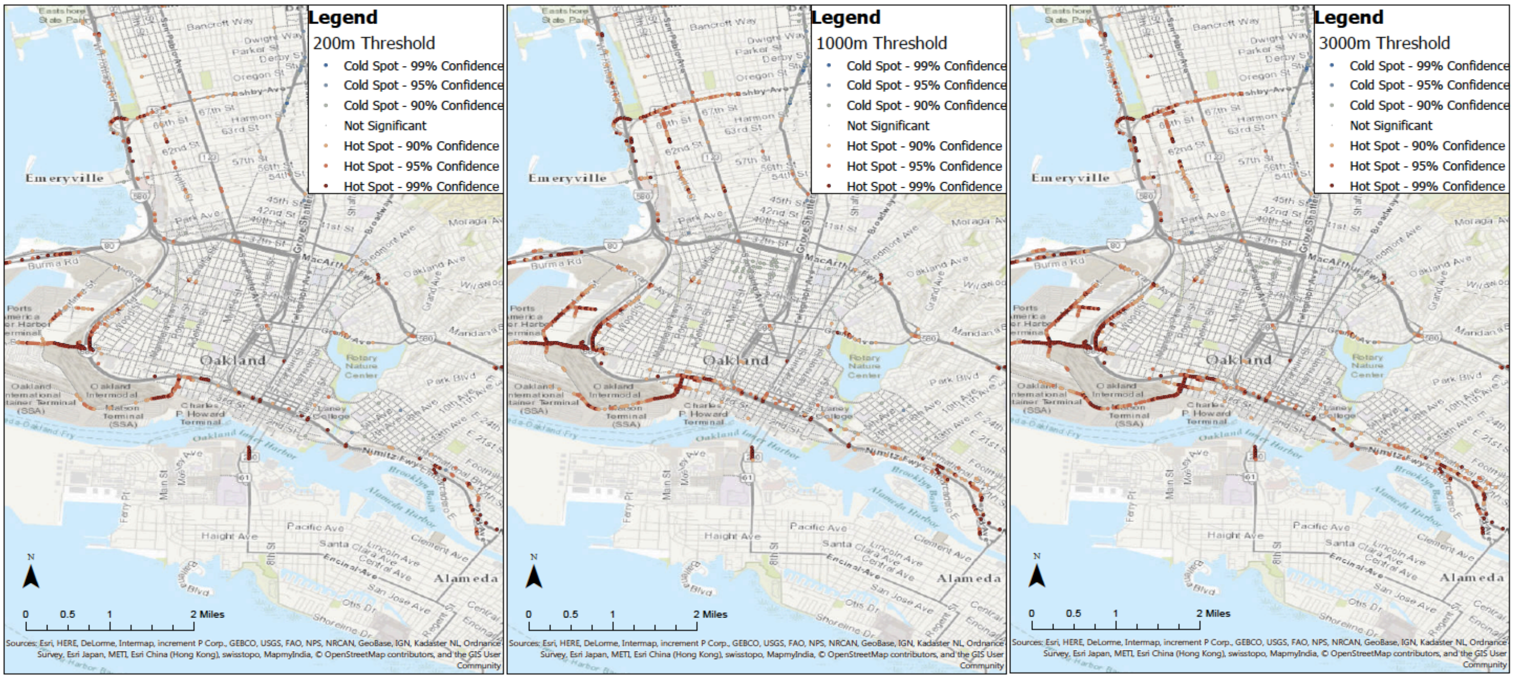
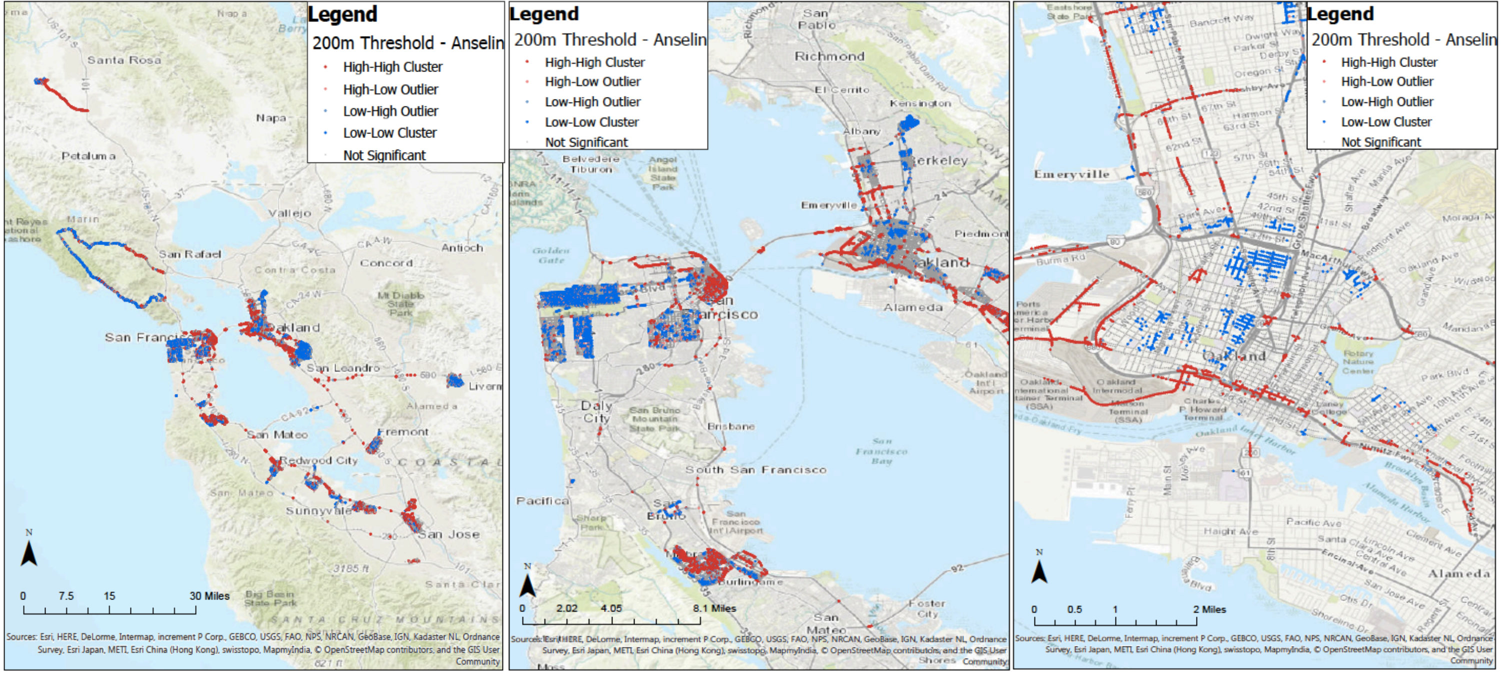

Legend
200m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
1000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
3000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Eastshore State Park
Bancroft Way
Emeryville
Oakland
Alameda
Oakland Inner Harbor
Charles P Howard Terminal
Rotary Nature Center
Pacific Ave
Haight Ave
Santa Clara Ave
Lincoln Ave
Central Ave
Encinal Ave
N
0 0.5 1 2 Miles
Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 9: Hot Spot Analysis tool results in Oakland for black carbon

From the above maps a couple trends appear. In the maps of the of the San Francisco area, as the distance band becomes larger, regional hot and cold spots indeed become more pronounced. The Financial district and Millbrae are obviously regions of high pollution relative to other measurements and Inner Sunset is a region of relatively low concentrations. Additionally, the Financial District contains hot spots at the smallest distance band, even though the whole area was determined to be a hot spot. This result makes sense because there are likely certain lights that consistently have traffic and produce emissions or other local point sources of pollution. The area just has many such points. For Oakland, some hot spots are identified, and visually seem to be hot spots based on the concentration map, though they are identified in analysis using all three distance bands. Most of West Oakland is not statistically significant in terms of points being hot or cold spots. One reason this could be is that Oakland was driven significantly more times than other polygons and the reduced data value at those points may be more representative of the true average. Six drive days may not be enough to identify stable local hot spots with this method.

2) Using Cluster and Outlier Analysis (Anselin Local Moran's I):

200m Distance Band Maps at Various Magnification Levels



Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 10: Cluster and Outlier Analysis tool results for black carbon

The Anselin Local Moran's I tool was used just for identifying local hot spots to compare these results to the Getis-Ord-Gi* tool. Many of the hot spots found with this tool were similar to those found with Getis-Ord-Gi* tool. The number of points each hot spot includes is larger. This could be because the Anselin Local Moran's I tool is designed to define more points as a cluster. Interestingly, a number of cold spot clusters were identified in this analysis that were not there previously.

Nitric Oxide

1) Using Hot Spot Analysis (Getis-Ord-Gi*):

Study Area

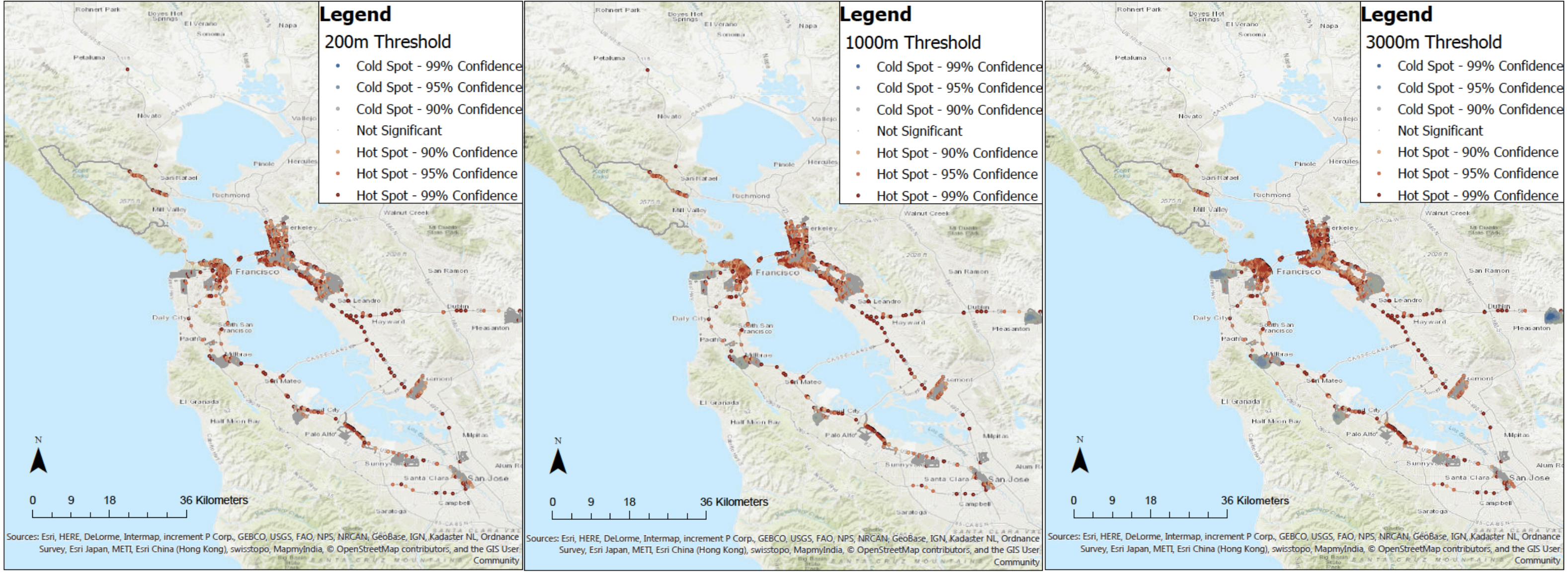

Legend
200m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
1000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
3000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
N
0 9 18 36 Kilometers
Petaluma
Novato
Vallejo
San Rafael
Richmond
Mill Valley
Berkeley
Francisco
San Leandro
Hayward
Daly City
Pacifica
San Mateo
Palo Alto
Fremont
Milpitas
Santa Clara
San Jose
Pleasanton
Dublin
San Ramon
Walnut Creek
Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 11: Hot Spot Analysis tool results in study area for nitric oxide

San Francisco and Surrounding Polygons

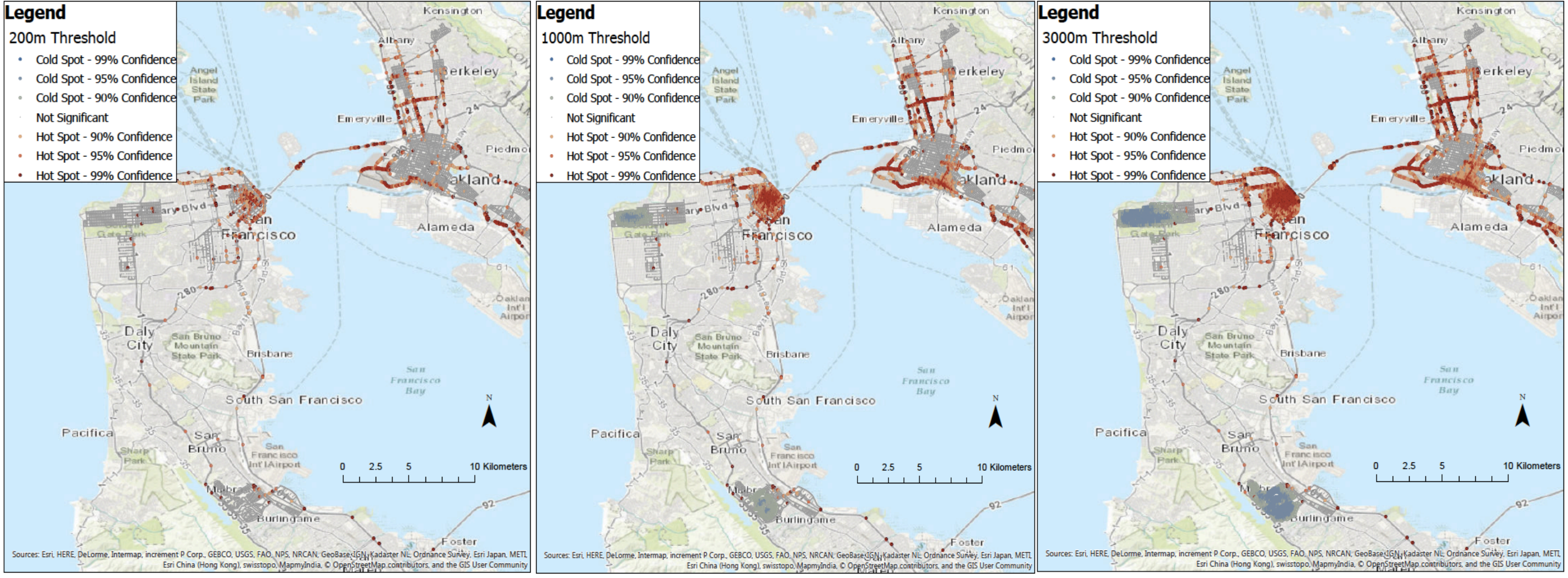

Legend
200m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
1000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
3000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
N
0 2.5 5 10 Kilometers
Kensington
Albany
Berkeley
Angel Island State Park
Emeryville
Piedmont
Oakland
Alameda
San Francisco
Daly City
San Bruno Mountain State Park
Brisbane
San Francisco Bay
South San Francisco
Pacifica
San Bruno
San Francisco Int'l Airport
Burlingame
Foster
Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 12: Hot Spot Analysis tool results in San Francisco area for nitric oxide

Oakland

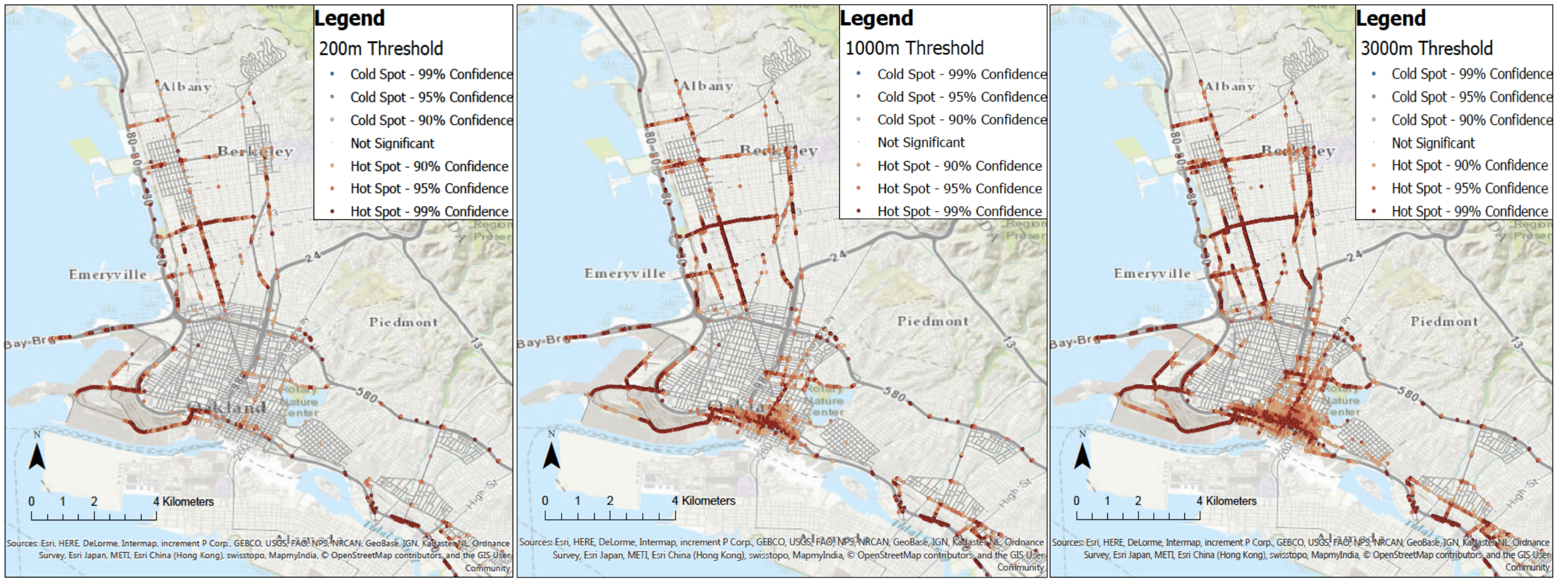

Legend
200m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
1000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Legend
3000m Threshold
Cold Spot - 99% Confidence
Cold Spot - 95% Confidence
Cold Spot - 90% Confidence
Not Significant
Hot Spot - 90% Confidence
Hot Spot - 95% Confidence
Hot Spot - 99% Confidence
Albany
Berkeley
Emeryville
Piedmont
Oakland
N
0 1 2 4 Kilometers
Sources: Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community

Figure 13: Hot Spot Analysis tool results in Oakland for nitric oxide

The results seen for nitric oxide are similar to those previously found for black carbon. Regional hot spots are well pulled out in the 3,000m distance band. For the same distance band, cold spots are less statistically significant than they were for black carbon. It also identifies quite a large hot spot in Oakland along the highway. This area makes sense as a regional hot spot, as highways are major emitters of nitric oxide. A few local hot spots in Oakland with a handful of points are identified and appear similar to the visually high values on the concentration maps.

2) Using Cluster and Outlier Analysis (Anselin Local Moran's I):

200m Distance Band Maps at Various Magnification Levels

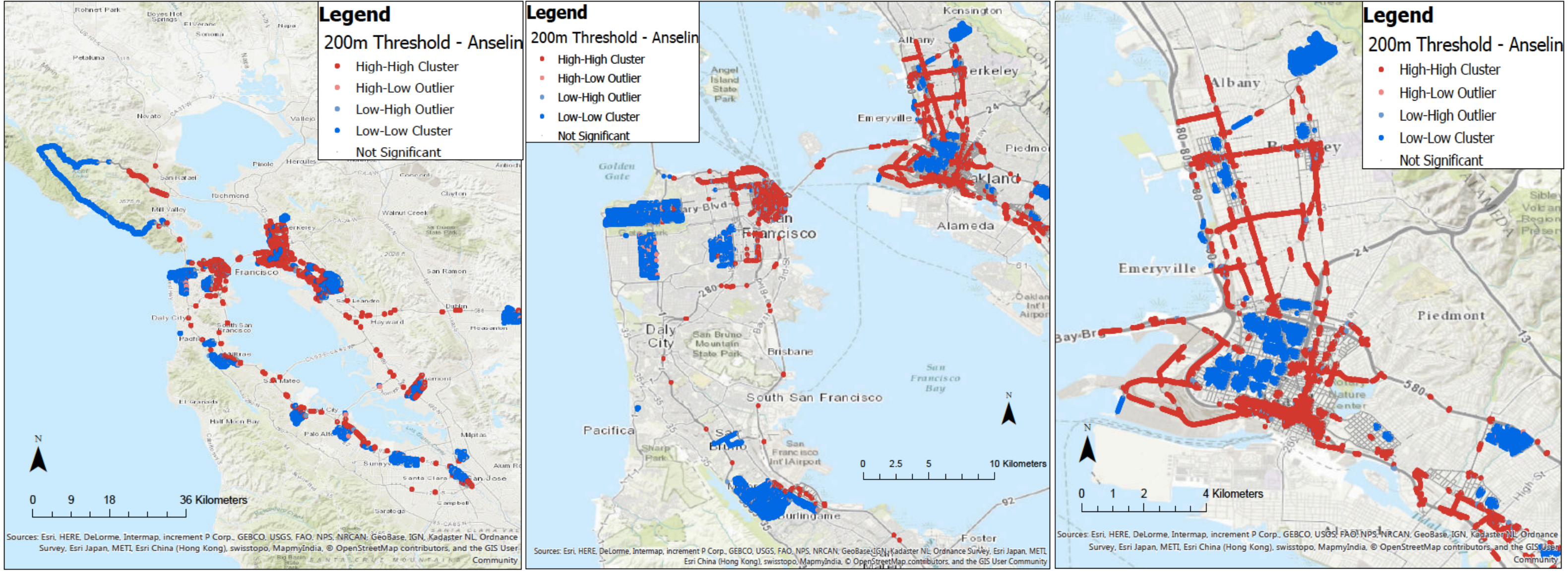


Figure 14: Cluster and Outlier Analysis tool results for nitric oxide

With the Anselin Local Moran's I, much more area has been found to be either a hot or cold spot. The identification of the highway area running through Oakland is similar, but much more area in the heart of Oakland is identified as a cold spot. This could be because nitric oxide is a short-lived air pollutant and areas adjacent to a highway or other major roads have very high concentrations but outside of traffic conditions there is little nitric oxide. Combining this knowledge with the possibility that this tool identifies larger clusters, it could be expected to identify larger areas as hot and cold spots.

Future Work

To make this analysis more robust, other methods will be tested to see if they give the same result. One such method is a program called SatScan, which is a software developed by the Centers for Disease Control and Prevention as well as other health-focused organizations to analyze patterns in spatiotemporal data by scanning different space-time parameters. For example, this could be used to find hot spots on different spatial scales in the Google Street View data by sweeping the radius of an influencing circle. If SatScan returns similar results given the same distance band considered in GIS, the conclusions from this project will be more robust. If unique methods return the same result, the result is more likely to be accurate.

Another investigation using different data reduction methods would provide more information on the sensitivity of the hot spots seen in the data. Maybe the mean of the drive passes or a nonparametric measurement like the ratio of the 10th percentile to the median could be used. The mean could show hot spot sensitivity to high measurements from a few drive passes. Similarly, a ratio of the 10th percentile to the median might be helpful because some of the air pollution instruments return negative values. Another way of measuring the data could be to consider the normalized difference to the mean. This could also lend itself to calculation of hot spots by comparing the normalized difference to the road type mean instead, so each potential hot spot would be compared only to measurements on similar roads. Even with my data reduction algorithm, there are a few highway points left even after removing highway road classes, likely because some points were mis-classed. These might have to be removed manually but it would make the analysis more robust.

Finally, since the data has rich spatial resolution but sparse temporal resolution, it would be helpful to look at the temporal distribution of a few of the hot spots that are identified. Temporal profiles of each point would show if the hot spot identified is consistently high or if the patterns suggest the high values are more transient and the hot spot identified would not be considered stable. Not many of the identified hot spots would need to be analyzed to get a good feel for what points have been identified but it would be incredibly useful information.

Conclusion

Both ArcGIS Pro tools yield good identification of regional hot spots. They perform similarly for each pollutant, an indication that the regional hot spots could indeed be considered hot spots because they are robust to multiple means of analysis. For the identification of local hot spots, both methods returned similar areas but the Anselin Local Moran's I gave larger hot spot areas. These two methods may not be the best way to find local hot spots without considering what changes are mentioned in Future Work. The best way to use Anselin Local Moran's I could be to find the outliers instead of the clusters. This likely would identify only a few spot that is high within an area of low spots, which is more consistent with how local hot spots are expected to look. More analysis for this idea is needed. Overall, Hot Spot Analysis (Getis-Ord-Gi*) and Cluster and Outlier Analysis (Anselin Local Moran's I) have potential as relatively simple methods to identify local and regional hot spots in the Google Street View dataset specifically and mobile monitoring data in general.