High resolution urban air quality sensing at scale

Anastasios Noulas
Yasin Acikmese
Firefly
New York, USA
{tassos,yasin}@fireflyon.com

Renaud Lamitotte
University of Oxford
Oxford, UK
renaud.lambiotte@maths.ox.ac.uk

Marta Gonzalez
UC Berkeley
California, USA
martag@berkeley.edu

The proposition of exploiting taxi fleets as a platform to sense urban environments has come into prominence in recent years due to the many advantages it offers both in terms of the scale and the resolution that monitoring can take place, but also due to economic factors that can determine the feasibility and longevity of a sensing project [1]. Taxi units as well as ride-sharing vehicles are omnipresent across time and space within the territory of a city and collecting various forms of signals in their surroundings at high spatio-temporal granularity is possible. Moreover, unlike purpose-built monitoring vehicles that could be deployed by city authorities or private organizations interested in environmental monitoring for instance, taxis do not require explicit commissioning to move about a city, a process that is very expensive to execute presenting a major unit economics challenge. In the meantime, one of the most significant issues that has concerned citizens, urban authorities and other governing bodies over the past decades is the ability to consistently monitor urban air quality. Growing research evidence points that poor environmental and atmospheric conditions in urban environments are one of the lead causes of premature death and a number of age long conditions such as asthma or other respiratory diseases including lung cancer [2], negatively affecting billions of people worldwide that reside at urban and urban-proximate areas. One of the most resonating ideas to perform air quality monitoring at scale has been the initiation of citizen led projects [3] which have been inspired by numerous crowd-sourcing projects that have emerged in the last twenty years and which have enabled the successful collection of data on mapping, social activity and mobility amongst others. Those projects come however with their own challenges such as the need for wide citizen participation rates in addition to deployment related and operational obstacles. Ubiquitous sensing technologies on the other hand can be deployed in existing taxi fleets at scale in an economically viable manner. Over the recent years Firefly [1] has developed and deployed a nation-wide and internationally expanding technology platform the core components of which are a digital display deployed on top of taxis and ride sharing vehicles, and a cloud orchestrated edge-capable software system that communicates information across the display network in real time. The display installation has offered a unique opportunity to deploy a number of sensors that can collect contextual signals of a vehicle's environment as it navigates the city. A key sensor in the display panel is a dust sensor measuring the number of units of suspended particulate matter PM2.5 in the air volume (particle concentration with diameters that are generally 2.5 micrometers and smaller in $\mu g/m^3$). Each sensor measures PM2.5 particles approximately every 60 seconds. As Firefly taxis navigate the city covering over time a large part of the street network, these data is coupled with GPS sourced mobility information effectively offering a highly detailed view of pollution levels in the city. In Figure 1, we demonstrate the spatial resolution of the data by visualizing the maximum PM2.5 concentration recorded at the street segment level. Immediately two key observations can be made. Firstly, the large heterogeneity of pollution levels emerging in the various locations of the city, highlighting how novel analytical insights can be obtained when diving beyond the aggregate view of a large geographic area. Secondly, higher pollution levels are recorded around the center, which hosts the denser parts of the build environment, as well as the main street network arteries which form the traffic backbone of the city. In Figure 2 we plot the probability distributions of PM2.5 readings in histogram form across the eight cities we study in the dataset (value range 0-30). While all cities follow a common pattern with the probability dropping significantly as PM2.5 values rise, there are notable variations across cities with some having their probability mass shifted more towards higher values. Important questions in this setting revolve around the understanding of what urban structural as well as environmental and population activity characteristics raise the probability of higher pollution concentrations in a city. We also compare the data collected by Firefly with air quality monitors by citizen led projects as well as government agencies. In Table 1 we report basic statistical properties of the data collected by Firefly taxis, ranking cities according to mean pollution level. We note how dense cities known

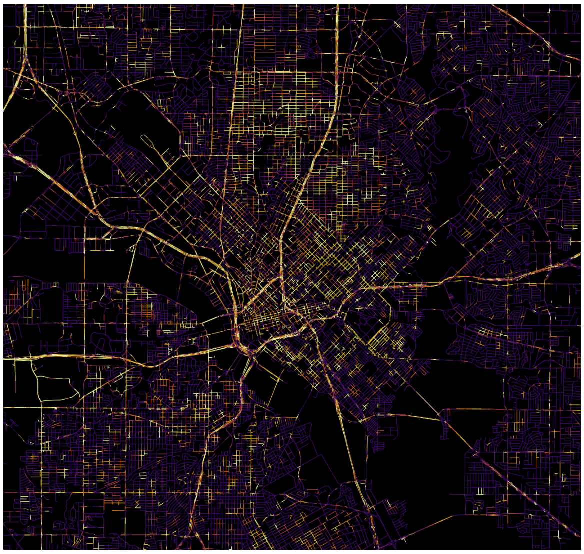

Figure 1: Dallas Street Network Colored according to max PM2.5 levels recorded (brighter color signifies higher levels).

---

[1] www.fireflyon.com

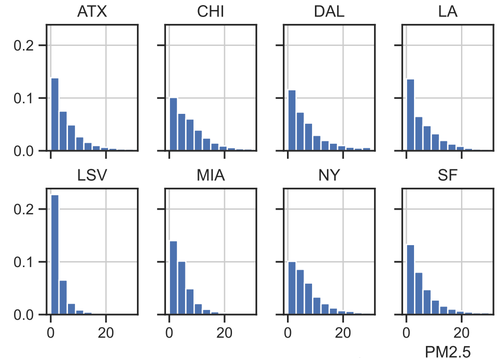

Figure 2: Per City PM2.5 Densities (0-30 value range).

City	Data points	Mean	SD	Rank
Dallas	3,337,678.00	8.11	12.81	1.00
New York	42,272,355.00	7.58	11.82	2.00
Chicago	16,557,911.00	7.45	8.73	3.00
Los Angeles	15,761,095.00	6.38	8.71	4.00
San Francisco	13,418,809.00	6.22	8.07	5.00
Austin	4,246,054.00	6.11	8.69	6.00
Miami	12,067,138.00	4.73	6.20	7.00
Las Vegas	37,189,840.00	2.64	6.83	8.00

Table 1: City level PM2.5 basic statistics, Firefly Taxis

City	PM2.5
Los Angeles-Long Beach-Anaheim, CA	12.1
Las Vegas-Henderson-Paradise, NV	10.5
Chicago-Naperville-Elgin, IL-IN-WI	10
Dallas-Fort Worth-Arlington, TX	9.8
San Francisco-Oakland-Hayward, CA	9.9
Austin-Round Rock, TX	9.5
Miami-Fort Lauderdale-West Palm Beach, FL	9.4
New York-Newark-Jersey City, NY-NJ-PA	8.7

Table 2: Annual mean values reported by the US EPA (https://www.epa.gov/) in similar regions (measured in µg/m3).

for their car-centric infrastructure and intense traffic conditions feature lower air quality standards. Similarly, in Table 2 we provide a ranking according to the annual mean values reported by the U.S. Environmental Protection Agency for similar regions. While there is a general agreement between the two sources of air quality measurement, disagreements are also apparent. We discuss potential sources of this discord including differences in measurement infrastructure as well as the regions across which measurement is carried out. As an example, in New York Firefly taxis are primarily active in the highly populous boroughs of Manhattan, Brooklyn and Queens whereas the US EPA reports measurement across a very wide region around New York City, which includes New Jersey. We discuss key sources of bias in the data, where taxi-based measurement is naturally skewed towards areas that taxis go to. They come however with the advantage of a very high spatial resolution view of the city when stations typically installed by government agencies are more sparsely dispersed across geographies. We argue that different sources of air quality monitoring are not only complementary, but essential, considering the environmental emergency our planet is going through and the epidemiological importance of air quality for urban populations. Another insightful perspective on air quality monitoring is the study of the temporal variability of pollution levels over time. In Figure 3 (top) we report mean pollution levels recorded by Firefly taxis in New York City for each hour of the week, where each data point corresponds to the mean pollution level observed at that hour. We normalize each data point with respect to the max observation during the 168-hour time window of a week and compare this signal with population fluctuations, taxi activity as well as traffic congestion signals. Despite the fact that pollutant particle diffusion and concentration patterns in the atmospheric realm of a city vary also due to weather conditions (e.g. wind patterns, humidity and temperature), here we concentrate our investigation on the link between pollution and human activity dynamics since the latter is the primary source of urban pollutants and a defining factor of their concentration dynamics. We also discuss seasonal variations in air quality patterns. In this context, we present evidence which suggests a connection between spikes in air pollution levels and significant climatic or social events. In Figure 3 (bottom) we present weekly mean PM2.5 levels in the New York region across the course of the year featuring an unusual increase in air pollution levels in early June 2023, a phenomenon known to have been induced by forest fires in Canada.

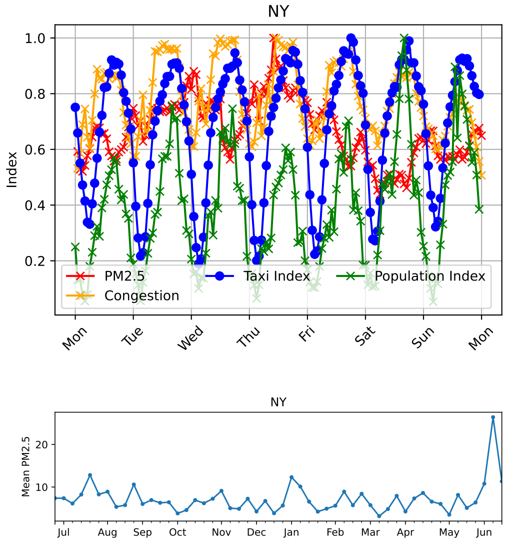

Figure 3: Weekly patterns of activity signals (top) and seasonal mean PM2.5 view (bottom) in the New York area.

REFERENCES

[1] K. O'Keeffe, A. Anjomshoaa, S Strogatz, P. Santi, and C. Ratti. Quantifying the sensing power of vehicle fleets. *PNAS*, 2019.

[2] R. Li, R. Zhou, and J. Zhang. Function of PM2.5 in the pathogenesis of lung cancer and chronic airway inflammatory diseases. *Oncology letters*, 2018.

[3] N. Castell, FR Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment international*, 2017.