Real-time Air Quality Monitoring Through Mobile Sensing in Metropolitan Areas

Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, Badri Nath
Department of Computer Science
Rutgers University
Piscataway, NJ 08854-8091
{skd70, parveen, antony26, rl475, iftode, badri}@cs.rutgers.edu

ABSTRACT
Traditionally, pollution measurements are performed using expensive equipment at fixed locations or dedicated mobile equipment laboratories. This is a coarse-grained and expensive approach where the pollution measurements are few and far in-between. In this paper, we present a vehicular-based mobile approach for measuring fine-grained air quality in real-time. We propose two cost effective data farming models – one that can be deployed on public transportation and the second a personal sensing device. We present preliminary prototypes and discuss implementation challenges and early experiments.

Categories and Subject Descriptors
C.5.3 [**Computer System Implementation**]: Microcomputers – *portable devices, microprocessors.*

General Terms
Measurement, Design, Experimentation, Human Factors.

Keywords
Air Quality, Pollution, Urban Sensing, Mobile Sensing, Social Networks, Participatory Sensing.

1. INTRODUCTION
As urbanization causes the growth of suburban communities, the existing transportation infrastructure dependent on fossil fuels must expand. Increase in vehicle use gives rise to an increase in traffic related pollutant emissions. According to census data, about 79% of the US population lives in urban areas [1]. As per 2010 Highway Statistics there are 242 million vehicles in the US alone [2]. To track the effect of this large fleet of urban vehicles on the environment and on the health of individuals, it is imperative to track pollutant levels in the urban and suburban settings. According to the US EPA [3], the six common air pollutants are particulate matter, ground-level ozone, carbon monoxide, sulfur oxides, nitrogen oxides, and lead. These are called the criteria pollutants and thus are required to be measured to tell us how healthy the air is to breathe [4, 5]. Among these, vehicular emissions contribute carbon monoxide, carbon dioxide and nitrogen oxides to the air pollution [6].

UrbComp'13, August 11–14, 2013, Chicago, Illinois, USA.

The current pollution measurement methodology uses expensive equipment at fixed locations or dedicated mobile equipment. The raw data obtained in this manner is used to further extrapolate the extent and concentration of pollution through dispersion models. This is a coarse-grained system where the pollution measurements are few and far in-between. Widespread deployment of this measurement paradigm is constrained by its prohibitive cost. In addition, it is desirable to have access to real-time measurements to be able to quickly analyze and identify alarming levels of pollutants. Currently, access to such data is limited [7] if not absent. It is available to and discernable by only a few who are well informed on the subject of pollution.

As opposed to a coarse-grained sensing system, a fine-grained approach would provide more frequent and spatially dense pollutant measurements. A scalable sensing platform could effectively disseminate pollution information to users in need. Today, the scarcity of fine-grained air quality information is hindering public awareness of health issues arising from pollution. Studies suggest that the health effects among asthmatics from short-term changes in air pollution levels are an important public health problem [8]. We anticipate that, with the help of fine-grained air quality measurements, people could be advised to take actions based on real time pollution levels to accommodate individual health needs.

The availability of real-time air quality data could make drivers better educated about driving patterns and how it impacts the environment and increases pollution. Better driving habits will lead to reduced pollution. Also, more health conscious citizens may choose alternate "healthy" routes based on the pollution information. It will benefit them as well as others by reducing pollution concentration in peak roadways so everybody breathes cleaner air.

At the same time, the emergence of cheap commodity air pollution sensors and the increase of cellular bandwidth have made mobile sensing platforms capable of real-time air quality data collection increasingly feasible. Several manufacturers such as Aeroqual or Variable Technologies have recently introduced handheld pollution measurement devices. These devices are small enough to be carried by walking people for personal use and measure all the criteria pollutants contributed by vehicle emissions [9,10,11,12]. But none of these off-the-shelf devices has been evaluated with respect to their real-time sensing performance when installed on mobile platforms such as vehicles. To the best of our knowledge, we have not come across any work that study the long-term stability, reliability and impact of real-time pollution monitoring systems using commodity sensors and the problems associated in deploying such systems.

In this paper, we present a vehicular-based approach of measuring fine-grained air quality in real-time. We propose two cost effective data farming models – one that can be deployed on public transportation and the second a personal sensing model. We present preliminary prototypes and discuss implementation challenges and experiments. In particular, we found that a personal sensing device conveniently mounted inside a vehicle in front of the vent can measure carbon monoxide levels that correlate well with outdoor values.

We start out with the mobile sensing schema, in which we talk about multiple mobile sensing models as well as their data processing in cloud. Then implementation is described in Section 3. Section 4 discusses the preliminary results and some challenges we identified based on our work. Section 5 gives related work. Finally, we close the paper with our conclusion in Section 6.

2. MOBILE SENSING SCHEMA

In our proposed mobile pollution-sensing schema (shown in Figure 1), a variety of mobile sensing models can be used to collect data from different scenarios. The sensing models measure the concentration of pollutants, tag the pollution data with relevant information, such as time, speed and GPS location[1], and send the data over a cellular data link to the cloud server. Raw pollution data is then processed and aggregated by the server to make it available as a pollution map. Various devices should be supported to access the map through browsers and mobile apps. Users would be able to view illustration of real time pollution data overlaid on map. This would enable users to get fine-grained street level air quality report.

We discuss the mobile sensing models, and data management in cloud in this section.

2.1 Mobile Sensing Models

The first sensing model is designed for deployment on **Public Transportation Infrastructure** such as buses, which have fixed and reliable routes along high volume corridors. For this model, we propose a custom-made Mobile Sensing Box (MSB) that includes a microcontroller board with add-on sensors, a peripheral GPS receiver and a cellular modem. Connecting to the bus battery would provide the power supply needed to operate this model. Since sensor bulk is not a primary design constraint in this case, it allows us to pack enough sensors per unit to measure all criteria pollutants. In our current prototype (see Section 3.1), we used two pollution sensors to measure carbon monoxide and particulate matter concentrations. However, in this paper we only focus on carbon monoxide.

The second sensing model relies on air quality-aware drivers who install a Personal Sensing Device (PSD) in their cars, connected over Bluetooth to their smart phone. Drivers can use this setting to measure the air quality for themselves, or they can register to participate in a **social community-based sensing**. The pollution data is geo tagged and posted to the central server over cellular network. In this paper, we focus on carbon monoxide sensor but the interchangeable sensors on wireless device can measure various other pollutants.

[1] Bluetooth enabled OBD-II Scan Tool can provide additional vehicle state information.

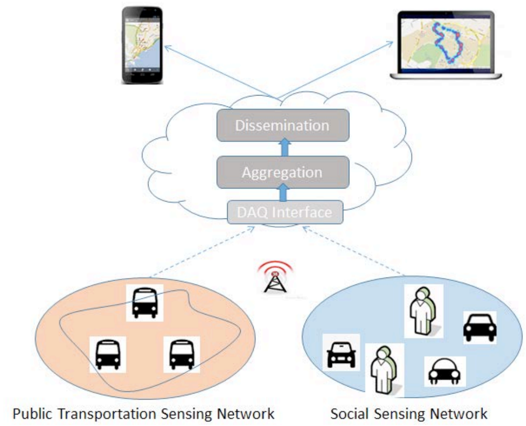

Figure 1 System Overview

2.2 DATA MANAGEMENT

The cloud service must be capable of collecting data from different types of sensors, which may sample different pollutants. In Section 2.2.1, we design the data collection so that the cloud can handle different data formats from various participating sensors. As this cloud service also provides community feeds to subscribing consumers, it must be capable of providing spatially or temporally aggregated feeds based on consumer-defined granularity. Besides, the service must be extensible and must be capable of supporting additional data dissemination patterns. We will discuss the data dissemination design in Section 2.2.2.

2.2.1 Data Collection

The server must provide a unified interface for sensors to communicate with. Different sensors produce different data formats, and this variation falls into two categories. (1) Variation in measured content - sensors have different pollution measurement capabilities. As a result, they produce different pollutant measurements with different levels of accuracy and variation. (2) Variations in data representation - location, for instance, can be represented in various formats.

We designed a protocol to handle such variations. The idea is to group different fields according to their role with respect to server's post processing step. Sensors are enabled to design their own data formats without adversely affecting the server processing. Any new data format is required to be registered with server a priori.

2.2.2 Data Dissemination

We provide a web portal where users can view real time pollution data. This is implemented as a new map layer, which we call Air Quality Index layer (AQI). There are two types of AQI layers available for different use cases – Marker map and Heat map. Marker map consists of data markers – when the user clicks on them will display all the information associated with each data point like – the time the data point was generated, GPS location, all the pollutant concentrations sent by the sensor, etc. Heat map shows all available measurements in a heat map style gradient color display where higher pollution is represented by higher ranked color in the color spectrum.

To provide efficient visualization, we adopt the visualization support from cloud storage services. We store common requests and their associated results as tables in server's domain. When we detect a match between a request and a table, we can return it immediately. This new design outperforms traditional massive data visualization tools in the sense that data filtering, sorting, aggregation and visualization are either pre-computed, or processed in the cloud.

The Heat map and the Marker map serve as basic visualization tools. Unfortunately, color gradients and markers do not help in data analysis and cannot be meaningfully consumed by downstream consumer applications. We accommodate these needs by extending our storage model to store temporarily generated tables and share them with consumer applications.

3. IMPLEMENTATION

In the following three subsections, we discuss the implementation details of two air quality-sensing models and of the server.

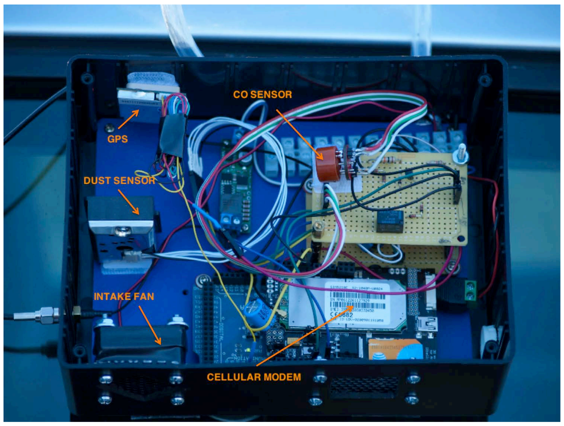

Figure 2 Mobile Sensing Box

3.1 Mobile Sensing Box

We assembled a Mobile Sensing Box (called MSB for short) as shown in Figure 2. It consists of a microcontroller, dust & carbon monoxide sensors, GPS and a cellular modem. The assembled unit can be mounted on any vehicle and can be powered by the vehicle's battery.

3.1.1 Details

We chose the Arduino Mega128 microcontroller for our prototype implementation [13]. The Arduino platform, with its large developer community and its reusable open-source libraries, provides a versatile microcontroller platform for rapid prototyping. We used the SIM5218 3G/GPRS cellular shield with AT command support for data transmission over HTTP [14]. The large number of I/O pins on the microcontroller facilitates the inclusion of many add-on sensors and peripherals to collect a wide variety of data. For the prototype implementation, we chose to measure position, velocity, carbon monoxide concentration and particulate matter concentration. The PMB 648 GPS receiver allows reading of position and velocity with high accuracy and can easily interface with the Arduino platform [15]. A Sharp Dust Sensor is used for dust concentration measurements. Carbon monoxide concentrations are measured using the MQ-7 carbon monoxide sensor from Hanwei Electronics [16].

3.1.2 Cost

The cost of assembling one unit came to about $700. In addition, we signed up for a $25 per month prepaid data plan with 1.5 GB data cap per month. The MSB generates 1600 bytes/minute of data when sampling every 5 seconds. With an average driving time of 2 hours per day for 30 days, the MSB produces about 5.5 MB of data per month, which is a small fraction of a typical low-end data plan.

3.1.3 Software

Our software on the Arduino uses a 'software serial' library to control and communicate with the carbon monoxide sensor, dust sensor, GPS chip and the cellular modem. Pollutant measurements are read from the analog to digital converter output. The MQ-7 sensor has a 30 second sensor heating cycle and a 60 second sampling cycle. GPS, carbon monoxide and dust readings are sampled periodically and transmitted to our cloud server. In order to prevent data loss on loss of data connectivity, a store & forward mechanism is included in our implementation.

3.2 Personal Sensing Device

The main components of the personal sensing device (called PSD for short) include a mobile air quality sensor and a smart phone to act as an interface with the central repository hosted on a cloud server. Our system uses a NODE Wireless Sensor platform available for smart devices from Variable Technologies [17]. The device is shown in Figure 3.

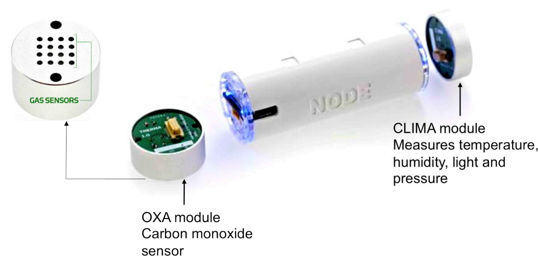

Figure 3 Personal Sensing Device [17]

3.2.1 Details

The NODE sensor platform is customizable with add-on sensor modules. Each device can accommodate two sensors on either end of the device. We selected OXA and CLIMA modules to measure carbon monoxide, humidity, temperature, ambient light and barometric pressure. Only carbon monoxide sensor was available when we started using this device for our experiments. The mobile sensor for our social model is intended for use anytime, anywhere irrespective of the mode of transport.

NODE uses Bluetooth connectivity to interface with users' smartphones to transmit the pollution levels in the environment.

3.2.2 Cost

The NODE device along with the OXA and CLIMA sensor modules costs about $400. In addition, user's existing iPhone device and data plan will be used to transmit data periodically to server. The personal sensing device generates 1536 bytes/minute of data when sampling every 5 seconds. With an average driving time of 2 hours per day for 30 days the MSB produces about 5.27 MB of data per month.

3.2.3 Smart Phone Application

An iPhone application uses the NODE iOS Framework to scan, connect and communicate with the NODE OXA and CLIMA sensor modules. We developed a custom iOS application using the

NODE Open API to read sensor data, tag the data with location information and send the data to the central server. The application allows the user to pair the smartphone with a sensor device of their choice over Bluetooth.

3.3 Cloud Server

Our cloud server is deployed on Amazon EC2 with classic LAMP settings. It solves most of the problems mentioned in Section 2.2, yet does not support some advanced features such as smart aggregation and automatic device registration. First, we describe some essential technical features, and then, we describe a typical workflow in order to show the data processing procedure.

3.3.1 Google Fusion Tables

Among the popular cloud storage services, Google Fusion Tables (GFT) has proven to be the best fit. It is designed as a new file type within Google Drive, with all the capabilities associated with a compact database. It supports a special data type for location storage, and supports various visualization tools for large data sets. This provides a convenient data storage in the cloud with Google's cloud visualization support.

The share feature proposed in Section 2.2 is implemented using Google Drive API. This API provides the ability to add, modify, or delete permissions for a file that resides in Google Drive. To perform these actions, the user simply needs to authorize requests using OAuth 2.0 and provide the email address of subscribers. In our case, server can create tables for users when necessary and then transfer ownership to the users. In this way, data producers retain full control of their raw data, by taking the role of the *owner*.

As mentioned above, we also adopt Google Fusion Tables API to utilize its visualization tools. The results are shown in Section 4.

3.3.2 GFT Repository

GFT Repository is a MySQL database, which we use to keep track of the Google Fusion Tables.

Besides normal aggregation tables, this repository also manages all the temporary aggregation tables. For example, "Busch" can be the name of the region containing all the route segments within the Busch Campus of Rutgers University. All queries with locations belonging to this area will be mapped to this region. As a result, a table storing the campus pollution data is created for smoother data retrieval.

Figure 4 shows a typical GFT item in the repository. Aggregation Table in Figure 4.a keeps track of the aggregations performed, owner who initiated the aggregation, the period for which data aggregation was performed, whether spatial and temporal aggregations are performed. In the table above, Start and End indicate the time span of data used for aggregation. A value of 1 int SpatialAgg indicates that a spatial aggregation has been performed and a value of 1 in TimeAgg indicates that a one-minute temporal aggregation has been performed. The Region Table in Figure 4.b indicates the region for which the aggregation was performed. In the example above, Rutgers campus was the region chosen.

Table in Figure 4.c – the region to road segment mapping table - consists of the road segments pertaining to a given region. So, the aggregation performed for the Rutgers campus is also applicable to the road segments that belong to the campus region.

a. Aggregation table

TableId	OwnerId	Shared	Start	End	RegionId	SpatialAgg	TimeAgg
1DnLldQXal	antonylhz	1	2013-03-27 21:44:23	2013-03-27 21:45:22	15VXJwMo	1	1

RegionId	Region Name
15VXJwMo	Busch

b. Region table

RoadSegId	RegionId
Ck9J95y1bbxef	15VXJwMo
alwXXB6oH13fY	15VXJwMo
VqbyvycGmaV0	15VXJwMo

c. Region – Road Segment Mapping table

Figure 4 GFT Repository Example

3.3.3 Data Process Pipeline

In this section, we present the data process pipeline through which we convert raw data from our two types of sensors into aggregated data stored in GFT Repository. We first give two sample measurements and then explain all the processing that takes place.

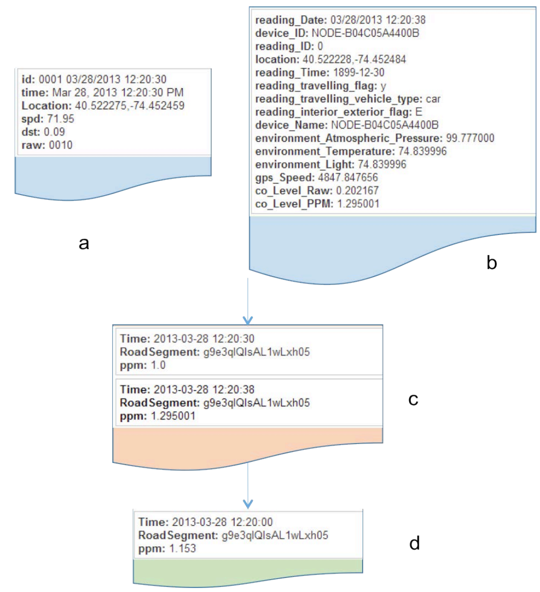

Figure 5 Aggregation Example

Figure 5 gives the two sample data points, which are of different sizes and contain different fields. Data point 'a' is from the MSB and data point 'b' is sent by the NODE sensor. The first step is the removal of the device identifiers, which are id, device_ID and device_Name in these two cases. Then, initial geographical aggregation is applied, causing the translation of the location field to road segments and then to the merging of the measurements as two data points for a road segment; shown in Figure 5 as 'c'. Finally, temporal aggregation is applied. Since these two records are assigned the same road segment and they are in the same time frame, they will be joined as shown in Figure 5 as 'd'. During this step other post-processing can take place such as the computation of the mean of the measurements or the Air Quality Index value. For example, the mean is calculated in the example shown in Figure 5.

4. EXPERIMENTS

In this section, we describe preliminary experiments we performed with the two platforms on highways in New Jersey and New York. In the first experiment, both sensing devices are placed outside the car. In the second experiment, NODE is mounted inside the car while MSB is still placed outside.

4.1 MSB Outside vs. PSD Outside

We deployed the Personal Sensing Device (NODE) inside Mobile Sensing Box (MSB) so that both the devices can measure the pollution simultaneously and under the same conditions. Mounting MSB outside the vehicle would be the typical deployment scenario on public transportation infrastructure. Figure 6 shows the MSB mounted outside the car. The inset picture shows the layout inside the MSB.

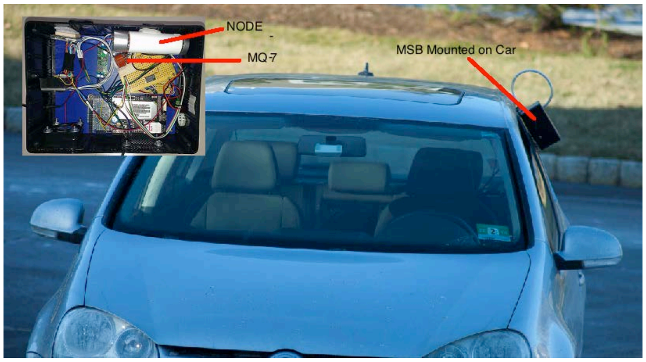

Figure 6 MSB Mounted on Car

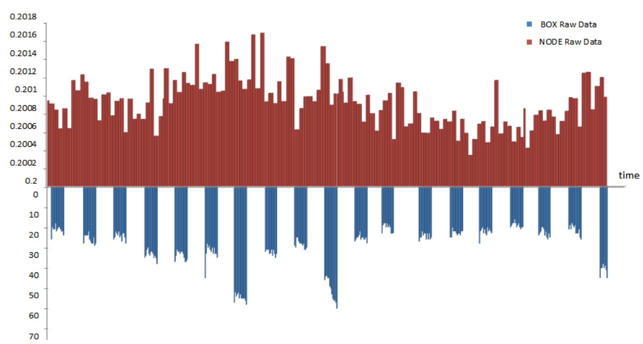

Figure 7 Pollution Data from NODE and MSB on NJ Turnpike

Figure 7 shows the correlation between the two platforms in a typical experiment. Each vertical represents one sample. Data in red is from the NODE and data in blue is from the MSB. In MSB, there is a 30 second heating cycle between two sampling cycles, so we can observe gaps in the data from the MSB.

We observed similar variations between the data from the two platforms. For linear regression, we calculated the average of each cluster of MSB readings and the average of the corresponding cluster from the NODE values. Using these pairs of values, we performed the linear regression analysis. From this analysis (shown in Figure 8), we found there is a significant positive linear relationship between the two measurements ($p<0.001$). In addition, the spearman correlation coefficient between these two was 0.85, suggesting that the corresponding pollution data from both platforms is highly correlated. This is an experiment conducted on the highway and in suburban areas, which have relatively steady pollution distribution. Whether this correlation applies to an urban environment needs to be studied.

Figure 9, shows the pollution level from this experiment. The Heat map shows that the highly polluted areas are located 2 miles north of exit 8 of NJ Turnpike due to congestion and the intersection of Route 18 and Main Street due to merging traffic. In addition, the Turnpike entrance and exit areas have slightly high pollution levels compared to other local roads because of the relatively high traffic going to and from the Turnpike. This is consistent with the hypothesis that high traffic density is correlated to high levels of pollution.

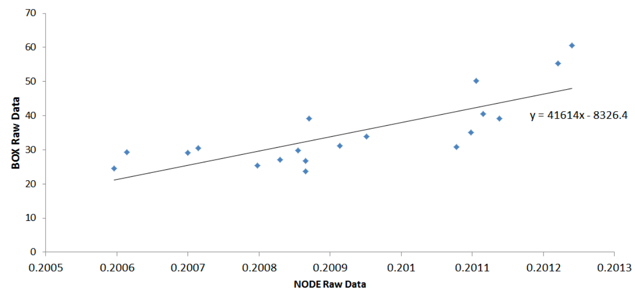

Figure 8 Linear Regression in Pollution Measurements on NJ Turnpike

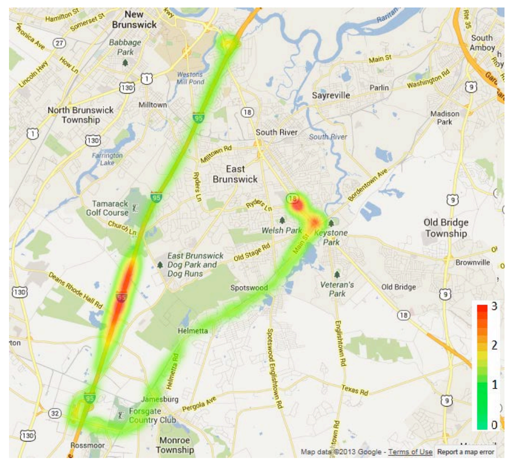

Figure 9 Heat map of Carbon Monoxide Concentrations on NJ Turnpike

4.2 MSB Outside vs. PSD Inside

In this section, we describe the experiment and the results when Personal Sensing Device (NODE) is deployed inside the car, which would be typical in personal sensing scenario. Figure 10 shows the PSD (NODE) mounted inside the car near the vent, while the MSB is fixed outside the car (as shown in Figure 6). During the experiment we kept the fans open to maintain the airflow from outside.

As shown in Figure 11, the test started from Route 440, Exit 4 and went through Staten Island Express Way (I-278), some local roads and finally returned to the starting point. Due to the influence of road constructions, the carbon monoxide concentrations on Staten Island Express Way are higher than other places in this experiment.

Using the same format as shown in Figure 7, we show data from this experiment in Figure 12. We can see that the variations of pollution levels from the two platforms are similar. We performed regression analysis as described in Section 4.1 (shown in Figure 13), which supports the positive linear relationship between the two platforms. However, spearman correlation coefficient is 0.5931, which is lower than the result from Section 4.1. More work is required to identify the reasons behind this.

Figure 10. Usage of NODE inside car

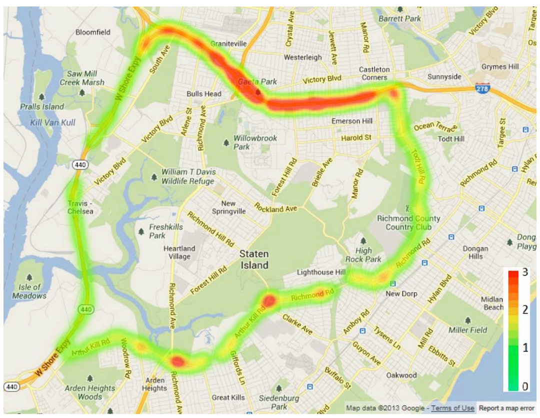

Figure 11 Heat map of Carbon Monoxide Concentrations on Staten Island

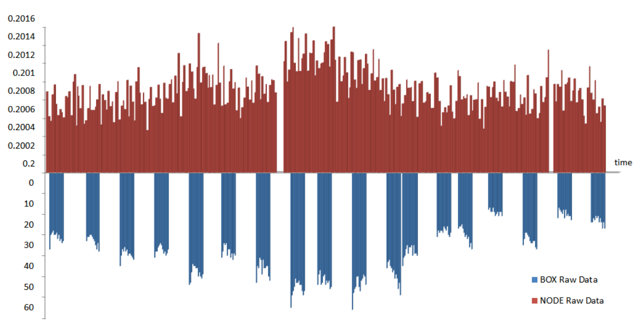

Figure 12 Pollution Data from NODE and MSB on Staten Island

4.3 Challenges

Our work revealed several challenges, which we plan to address in our future work.

Sensor Location. We observed that the orientation of the sensor relative to the vehicle movement influences the measurements. The inline deployment allows air to flow into the sensor body. Intuitively, this should provide true ground level measurements as opposed to the transversal deployment of the sensor. However, in cases where sensors use a sensor preheat cycle before sampling, this will affect the measurements.

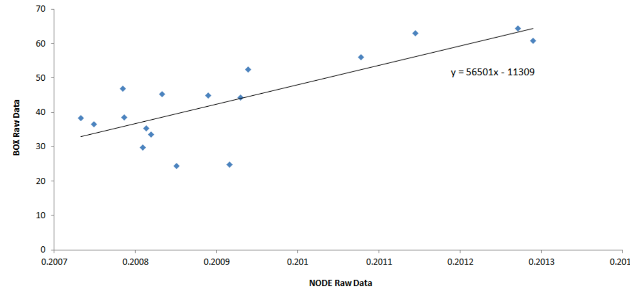

Figure 13 Linear Regression in Pollution Measurement on Staten Island

In MSB design, sensors are mounted in a box and a small fan draws air into the box for sampling. We plan to evaluate the measurements using the multiple orientations described above to understand how orientation influences the measurements. In addition, we have conducted tests with the personal sensing device placed inside and outside the car. Our preliminary studies indicate that pollution readings inside vehicles correlated with those taken outside vehicle on freeways and suburban roads.

Sampling Strategy. The frequency at which measurements are done should relate to the vehicle speed and to the spatial gradient of pollution. For instance, at highway speeds, a frequency of one sample per minute means the measurements are one mile apart (assuming a vehicle speed of 60 MPH). Is this measurement frequency really necessary given the rate at which pollution level changes relative to distance? If not, what factors should the client-sensing app consider to decide the best sampling time for accuracy? Geography of the road, speed, and changes in traffic are the obvious candidates.

The sampling strategy also needs to define the minimal subset of criteria pollutants that must be measured on a mobile platform to reflect the air quality. This is essential to arrive at a cost-effective mobile platform.

Server Aggregation. The server needs to implement a smart data aggregation strategy on data provided by multiple cars so that the air quality reports correlate with the ground truth pollution.

Temporary pollution surge is an additional phenomenon we observe in our measurements. A passing truck that produces high pollution around itself usually causes this; but this effect is temporary and does not reflect the ground truth. Algorithms to find and eliminate these outliers need to be designed.

We also found that pollution on the road is highly sensitive to weather. Sensors are also susceptible to produce varying readings during abnormal weather conditions. Currently, we segregate pollution measurements under different weather conditions and plan to further investigate the correlation between weather and the ground-truth pollution measurements.

Incentives and Applications. Given these inevitable sources of inaccuracy, we must rely on developing a large pollution sensing community and more advanced pollution sensor technology. How to incentivize users to produce more measurements is yet another challenge. We plan to provide value added services derived from our system to volunteers as a complimentary reward for their participation.

Furthermore, it will be interesting to study how general drivers behave when they are presented with fine-grained pollution data collected by our two models. For example, given the information on highly polluted routes, how many drivers will choose to take cleaner route at the cost of longer driving times?

5. RELATED WORK

Due to the huge gaps in ground-based networks of air pollution monitors, there is a necessity to obtain fine-grained air quality data. Various attempts have been made to employ mobile sensors in order to achieve this goal. The School Bus Monitoring Study [31] conducted at University of California along with NRDC (National Resources Defense Council) highlights the health hazards posed to school children by their exposure to diesel pollutants. It also emphasizes the urgent need for mobile monitoring of air quality because diesel exhaust is a known carcinogen and a cause of respiratory illnesses. An interesting study was conducted by EPA [24] to measure air pollutant concentrations inside and outside of a truck cabs. The study however used measurement techniques that involved collecting air samples in the truck and later analyzing them in a lab to derive actual air quality values. Wireless sensor networks for monitoring personal pollutant exposure [25], indoor air quality [23] and hazardous sites [30] have also been proposed.

In order to bridge the gap between the sampling phase and the analysis phase, researchers introduced monitoring approaches using commodity sensors, which can provide real time pollution data. N-smarts [26] and CommonSense research conducted jointly by UC Berkeley and Intel focused on collecting air quality data by attaching sensors to GPS enabled cell phones. It also highlights various challenges with the quality of sensor data from networked mobile sensing units such as interference of user behavior, location coverage, calibration accuracy and social aspects of mobile sensing and impact on citizen behavior.

Work has also been done to evaluate the design issues of sensor boards for air quality monitoring [22]. The challenges in preserving privacy of participants of personal sensing have been studied [28, 29]. A software framework for data gathering using smart phones has been presented in [27]. Air Quality Egg [33], a project hosted on Xively (formerly Pachube/COSM) has introduced a personal pollution-sensing platform.

OpenSense [18], a project run by EPFL and ETH Zurich, Switzerland, aims to study the feasibility of installing sensors on the roofs of buses and trams, taking advantage of existing public transportation vehicles to form an extensive network of mobile air quality data collection sites. Similar pollution sensing network has been tested on the buses in the city of Sharjah, UAE [20]. The Air Project [19] is a public, social experiment in which people are invited to use portable air monitoring devices to explore their neighborhoods and urban environments for pollution and fossil fuel burning hotspots. Teco Envboard [21] focused on design of sensing platform with commercial off the shelf sensors for carbon monoxide, carbon dioxide, ozone and nitrous oxide for urban/participatory sensing projects. Another interesting approach discussed in [32], wherein; the historical and real-time air quality measurements are used to infer the fine-grained air quality in a city.

6. CONCLUSION

This paper presents two mobile platforms for fine-grained real-time pollution measurement, mobile sensing box, deployable on public transportation infrastructure and a personal sensing device (NODE) that can be used to create a social pollution sensing. We conclude that both approaches are feasible. We also show that a personal sensing device can be conveniently used inside the car, yet still producing meaningful pollution measurements. However, more work is needed to arrive at a model that reflects the ground truth pollution values.

The two models presented in this paper, though built with the common goal of air quality monitoring, present varied advantages and challenges. Since the public infrastructure model uses buses travelling on fixed routes at scheduled times, it ensures a constant reliable stream of pollution measurements. The social community-sensing model, on the other hand, relies entirely on the participants to generate data. Hence, the number of individuals actively participating will determine the breadth of pollution information obtained. But, unlike the bus scenario, where data will always be pertaining to certain fixed routes everyday, the social scenario would collect air quality measurements for multitude of routes across the country and enable wider coverage and provide redundancy.

The air quality data obtained using such sensing models could serve various applications. Patients with respiratory or cardiovascular diseases would find our results valuable to determine less polluted routes. Health aware individuals could also take advantage of this form of cleaner route navigation. Individuals using our system will become more knowledgeable about the extent of pollution and be motivated to follow better driving patterns, such as not allowing their vehicle to idle for long periods or to drive more environmentally friendly cars. Apart from these applications at an individual level, this data could be used as an additional input to large-scale policy making. For example, public health officials and policymakers could use our results to predict potential health impacts based on air quality across various regions to make decisions such as possible locations for a future school or residential community.

7. ACKNOWLEDGEMENTS

This work has been supported in part by NSF Grant CNS 1111811.

REFERENCES

[1] 2000 US Census Data Retrieved May 2013 from http://www.census.gov/compendia/statab/2012/tables/12s0029.pdf

[2] Federal Highway Administration Highway Statistics Retrieved May 2013 from http://www.fhwa.dot.gov/policyinformation/statistics/2010/mv1.cfm

[3] What Are the Six Common Air Pollutants Retrieved May 2013 from http://www.epa.gov/air/urbanair/

[4] Criteria Pollutants Retrieved May 2013 from http://www.epa.gov/airquality/montring.html#criteria

[5] Clean Air Act Retrieved May 2013 from http://www.epa.gov/air/caa/

[6] Automobile Emissions-An Overview Retrieved May 2013 from http://www.epa.gov/otaq/consumer/05-autos.pdf

[7] AirData web portal Retrieved May 2013 from http://www.epa.gov/airdata/

[8] Effects of Ambient Air Pollution on Symptoms of Asthma in Seattle-Area Children Enrolled in the CAMP Study Retrieved May 2013 from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1240204/pdf/ehp0108-001209.pdfAirData web portal Retrieved May 2013 from http://www.epa.gov/airdata/

[9] CO Sensor. Retrieved May 2013 from https://www.sparkfun.com/products/9403

[10] CO_2 Sensor. Retrieved May 2013 from http://sandboxelectronics.com/store/index.php?main_page=product_info&products_id=197

[11] NO_2 Sensor. Retrieved May 2013 from http://wickeddevice.com/index.php?main_page=product_info&products_id=111

[12] Optical Dust Sensor. Retrieved May 203 from https://www.sparkfun.com/products/9689

[13] ArduinoMega online specification. Retrieved May 2013 from http://arduino.cc/en/Main/arduinoBoardMega

[14] SIM5218 online reference guide. Retrieved May 2013 from http://www.cooking-hacks.com/index.php/documentation/tutorials/arduino-3g-gprs-gsm-gps

[15] PMB-648 GPS Receiver Module Specification. Retrieved May 2013 from http://www.parallax.com/Portals/0/Downloads/docs/prod/sens/PMB-648_Specification_V0.1.pdf

[16] MQ-7 CO Sensor datasheet. Retrieved May 2013 from https://www.sparkfun.com/datasheets/Sensors/Biometric/MQ-7.pdf

[17] (2013) Main page - Variable Technologies (Online) available at http://nodesensors.com

[18] Karl Aberer, Saket Sathe, Dipanjan Chakraborty, Alcherio Martinoli, Guillermo Barrenetxea, Boi Faltings, and Lothar Thiele. 2010. OpenSense: open community driven sensing of environment. In Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS '10), Mohamed Ali, Erik Hoel, and Cyrus Shahabi (Eds.). ACM, New York, NY, USA, 39-42. DOI=http://doi.acm.org/10.1145/1878500.1878509

[19] AIR: Area's Immediate Reading (September 2006) Retrieved May 2013 from http://www.pm-air.net

[20] Al-Ali, A. R.; Zualkernan, I.; Aloul, F., "A Mobile GPRS-Sensors Array for Air Pollution Monitoring," Sensors Journal, IEEE, vol.10, no.10, pp.1666, 1671, Oct. 2010. DOI=http://dx.doi.org/10.1109/JSEN.2010.2045890

[21] Budde, M.; Berning, M.; Busse, M.; Miyaki, T.; Beigl, M., "The TECO Envboard: A mobile sensor platform for accurate urban sensing — And more," Networked Sensing Systems (INSS), 2012 Ninth International Conference on, vol., no., pp.1, 2, 11-14 June 2012. DOI= http://dx.doi.org/10.1109/INSS.2012.62405 (2013) N-Smarts available at http://www.paulos.net/papers/2008/NSDR08.pdf

[22] S. Choi, N. Kim, H. Cha, and R. Ha. Micro sensor node for air pollutant monitoring: Hardware and software issues. In Sensors MEMS, 2009. DOI=www.mdpi.com/1424-8220/9/10/7970/pdf

[23] Wan-Young Chung, Sung-Ju Oh, Remote monitoring system with wireless sensors module for room environment, Sensors and Actuators B: Chemical, Volume 113, Issue 1, 17 January 2006, Pages 64-70, ISSN 0925-4005, 10.1016/j.snb.2005.02.023.

[24] Prakash Doraiswamy, Wayne T. Davis, Terry L. Miller, Joshua S. Fu and Yun-Fat Lam. 2005. *Measuring Air Pollution Inside And Outside of Diesel Truck Cabs*. Report prepared for US EPA by University of Tennessee, Knoxville, TN.http://www.epa.gov/smartway/documents/publications/incabairquality-110405.pdf

[25] Tsow, F.; Forzani, E.; Rai, A.; Rui Wang; Tsui, R.; Mastroianni, S.; Knobbe, C.; Gandolfi, A.J.; Tao, N.J., "A Wearable and Wireless Sensor System for Real-Time Monitoring of Toxic Environmental Volatile Organic Compounds," Sensors Journal, IEEE , vol.9, no.12, pp.1734,1740, Dec. 2009. DOI= 10.1109/JSEN.2009.2030747

[26] Richard Honicky, Eric A. Brewer, Eric Paulos, and Richard White. 2008. N-smarts: networked suite of mobile atmospheric real-time sensors. In Proceedings of the second ACM SIGCOMM workshop on Networked systems for developing regions (NSDR '08). ACM, New York, NY, USA, 25-30. DOI=http://doi.acm.org/10.1145/1397705.1397713

[27] A. Joki, J. Burke, and D. Estrin. Campaignr-a framework for participatory data collection on mobile phones. Technical report, CENS, UCLA, 2007

[28] Shilton, K., Burke J., Estrin D., Hansen M., Gorvindan R., and Kang J. 2009. Designing the Personal Data Stream: Enabling Participatory Privacy in Mobile Personal Sensing. The 37th Research Conference on Communication, Information and Internet Policy (TPRC). DOI=http://escholarship.org/uc/item/4sn741ns

[29] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. 2009. Exploring Privacy Concerns about Personal Sensing. In Proceedings of the 7th International Conference on Pervasive Computing (Pervasive '09). DOI=http://dx.doi.org/10.1007/978-3-642-01516-8_13

[30] Gianfranco Manes, Giovanni Collodi, Rosanna Fusco, Leonardo Gelpi, and Antonio Manes. 2012. A Wireless Sensor Network for Precise Volatile Organic Compound Monitoring. International Journal of Distributed Sensor Networks Volume 2012, Article ID 820716, 13 pages DOI= 10.1155/2012/820716

[31] Gina M. Solomon, Todd R. Campbell, Gail Ruderman Feuer, Julie Masters, Artineh Samkian, Kavita Ann Paul. 2001. *NO BREATHING IN THE AISLES, Diesel Exhaust Inside School Buses*. Report prepared by the Natural Resources Defense Council and Coalition for Clean Air. http://www.nrdc.org/air/transportation/schoolbus/schoolbus.pdf

[32] Yu Zheng, et al. U-Air: When Urban Air Quality Inference Meets Big Data. 19th KDD 2013.

[33] Air Quality Egg – http://airqualityegg.com