Mobile Monitoring of Air Quality in the Washington DC Region

Kuldeep Kumar Dixit

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Urban and Regional Planning

In

Urban and Regional Planning

Steven C. Hankey

Gabriel Isaacman-VanWertz

Theodore C. Lim

December 05, 2022

Blacksburg, Virginia

Keywords: particulate air quality, air pollution, land use planning,
Exposure assessment

Mobile Monitoring of Air Quality in the Washington DC Region

Kuldeep Kumar Dixit

ABSTRACT

Exposure assessment is a critical step in air quality-related epidemiological studies. Accurate estimates of exposure within urban areas are a vital input to models that aim to assess the health effects of air quality among populations of interest. In this study, I have derived and applied a novel approach for capturing the distribution of air quality in Arlington, VA and Washington DC using mobile monitoring. The main objectives of this study are:

1. Deploy a year-long sampling campaign in the Washington DC region to capture the within-city variability of air quality for Particle Number Concentration (PNC), fine particulate matter ($PM_{2.5}$), and Black Carbon (BC) using mobile monitoring.
2. Derive a method for selecting the best representative mobile monitoring routes to capture within-city spatial patterns of air quality. The end-use of the monitoring campaign described here is as an input for Land Use Regression (LUR) models.
3. Collect unconventional data to characterize the built environment, e.g., videos, sound, etc., that could be employed to improve the LUR models beyond conventional approaches.

This study describes the data collection effort that was deployed for a year to characterize annual average concentrations at different locations across the Washington DC region. My thesis describes the challenges experienced and lessons learned during the data collection phase. The

goal of this thesis is to describe the data collected and the methods used to sample the DC region. This effort is a component of a larger project that will later use these observations in LUR models.

The central site used for measurement of background concentration had a lower concentration median when compared with the median concentration measured on bike. The median $PM_{2.5}$ concentration at the central site was observed to be 5.2 $\mu g/m^3$ and the median PNC at the central site was observed to be 6,365 $\#/cm^3$. The Arlington $PM_{2.5}$ concentration was 1 $\mu g/m^3$ and the Washington DC concentration $PM_{2.5}$ was 0.3 $\mu g/m^3$ higher than the background median concentration. Also, the Particle Number Concentration (PNC) was 222 $\#/cm^3$ more and the Washington DC PNC was 2,139 $\#/cm^3$ higher than the median background concentration.

Mobile Monitoring of Air Quality in the Washington DC Region

Kuldeep Kumar Dixit

GENERAL AUDIENCE ABSTRACT

Many studies have shown that living in polluted air has long-term negative impacts on human health. These negative impacts include premature death, lung disease, heart disease, blood disease, and other complications. Due to these impacts, it is critical to know the level of air pollution within cities to identify areas that have elevated concentrations.

The measurement of air quality is challenging because of the low number of monitors available due to cost. Reference grade air quality monitors are often very costly. In this study, I have developed an approach for using a bike to collect mobile measurements of particulate air pollutants in the Washington DC area. I collected one year of data at a fixed site in Arlington and four seasons of data from the bike mobile monitoring campaign. After analyzing the data, I observed that the fixed station showed lower concentration when compared with the data collected by bicycle. I have also suggested improvements in the mobile monitoring method and developed an approach for joining these data with outputs from computer vision models to describe the built environment.

Acknowledgements

I thank Krishna and Hanuman, who blessed me with a good life and have been a continued source of my spiritual well-being. I would also like to thank my parents for raising me, getting me educated, and for their unending support. I am grateful to Steve for motivating me, advising me, and helping me navigate my journey at Virginia Tech. Steve's open-door policy has helped me countless times and I am also grateful to him for sending me to the AGU conference.

I thank Christopher Evans for helping me collect the bike samples and making the workplace fun and enjoyable. I also appreciate the open conversations with Chris that helped in understanding the cultural differences. I would also like to thank the Virginia Tech Research Center's building management staff for helping me with access to the rooftop and giving me support with central site management. Thank you, Kevin, for all the instrument transfers from three floors under the ground to seven floors up to the penthouse.

I would also like to thank all members of Steve's group for their help and for being available for consultation and helping me with technical issues. I am grateful to Luke and Meng for helping me with the data processing. Thanks, Sajad for always being available for consultation.

Also, a big thanks to the SPIA staff who helped me manage my student documentation on time. Thank you, Han, for all the purchase orders and reimbursement assistance. Thank you, Ryan, for helping with the P3 cage management and assisting with the building access.

Table of Contents

List of Figures

List of Tables

Abbreviations

GSV - Google Street View

VTRC - Virginia Tech Research Center

CPC - Condensation Particle Counter

UPAS - Ultrasonic Personal Air Sampler

LUR - Land Use Regression

PPA - PurpleAir

FRM - Federal Reference Method

FEM - Federal Equivalent Method

USB - Universal Serial Bus

PSPNet - Pyramid Scene Parsing Network

FOV - Field of View

1.0 Introduction

According to the Lancet commission (Fuller et al., 2022), air pollution causes 9 million deaths around the world annually; according to the global burden disease study (Cohen et al., 2017), fine particulate matter ($PM_{2.5}$) was the fifth leading global cause of death. This highlights the importance of $PM_{2.5}$ monitoring at ground level. However, due to the high cost of monitoring instruments, it is not always possible to deploy a large number of monitors, and scientists use methods like Land Use Regression (LUR) models to estimate exposure. Researchers have also employed satellite data-based models and chemical transport models to estimate various pollutants; but, chemical transport models are dependent on emission inventory data that are available at a very coarse resolution and are thus unable to capture the urban-level variation within the city.

1.1 Mobile Monitoring Method

One of the limitations of the Land Use Regression (LUR) method is that it needs ~100 fixed sites (Basagaña et al., 2012) to have good model performance and subsequently estimate exposure in an urban environment. To overcome these limitations mobile monitoring is one of the low-cost and viable alternatives. Mobile monitoring campaigns are much more affordable than buying multiple (Federal Reference Method) FRM monitors. For monitoring pollutants some studies have used data from available from fixed monitors for mapping, some have used the monitoring data as input for the LUR models and some try to assess exposure through different modes of transports like automobiles, bikes, pedestrian mode, etc. Some Studies have (Apte et al., 2017; Chambliss et al., 2021) used measurements from the Google street-view vehicle to build a LUR model. Some (Hankey et al., 2015; Samad et al., 2020) have used a bike to sample air quality and some (Alas et al., 2021) have used backpack on foot to measure air quality. The three methods have their advantage and disadvantages, with the most accurate being the pedestrian method of collecting as it measures the most-accurate exposure to the human body but, due to the slow speed of travel, it often lacks spatial coverage. The automobile mode has the most spatial coverage of all three modes but, it is limited to near-road exposure only as automobiles are not allowed on many walkways. The bike mode lies in between the pedestrian mode and the automobile mode as it can cover more spatial coverage than the pedestrian mode (and less than the automobile) but can sample most areas where pedestrians are allowed (e.g., trails or parks). In this study, I have used an approach extended previous study (Hankey et al., 2015) for measuring air quality. I have used miniaturized instruments like the sidepak AM520, Condensation Particle Counter (CPC) 3007, microaeth AE51 to collect air quality parameters like fine particulate matter ($PM_{2.5}$), particle number concentration (PNC) and black carbon (BC) with reasonable accuracy. Using the bike-based measurement platform I was able to collect measurements from trails, walkways, and parks in addition to a variety of roadways. I also tried to spread the route across the spatial and temporal domains to keep the sample normalized as much as possible in this study.

1.2 Use of Computer Vision in Air Quality Modelling

Recent studies (Qi et al., 2021; Qi et al., 2022) have demonstrated that Google Street View (GSV) images could be utilized to build a highly accurate air quality model for Particle Number Concentration (PNC), Black Carbon (BC) and Nitrogen Dioxide (NO_2). Studies (Ganji et al., 2020;

Dao et al., 2021) have shown that street view images could be used to enhance the accuracy of air quality estimates. In this study, I have also used this unconventional new approach and recorded a stable video for each bike run using two GoPro Hero 9 Black cameras to capture the built environment from each route. This thesis focuses on the data collection effort; However, this data could be utilized later to extract built environment features for better estimation in the LUR model. My approach is novel, as to the best of my knowledge, no one has tried to build a similar model using on-the-shelf cameras. The setup of the two bike cameras is explained more in the methodology section.

1.3 The Portable Filter Sampler

This study also shows low-cost sensor and sidepak performance in comparison to the filter-based measurement from the Ultrasonic Personal Air Sampler (UPAS). The UPAS is a relatively new mobile measurement instrument, and there are few studies available that compare UPAS performance in an outdoor environment. Some studies (Pillarisetti et al., 2018; Volckens et al., 2017; Burrowes et al., 2020; Arku et al., 2018) have compared the UPAS performance in indoor environments. Although, I have not used the Federal Reference Method (FRM) instrument in this study to compare UPAS performance but our study in agreement with one other study (Vernooij et al., 2022) show a very high correlation between the sidepak and gravimetric method in an outdoor environment. I have used the correlation between sidepak and UPAS as a metric of the performance of the two devices.

2.0 Methodology

I primarily collected data in this study from two sources 1) the central site and, 2) the two fixed bike routes. The Virginia Tech Research Center (VTRC) located at 900 N Glebe Rd, Arlington, Virginia was used as the central site to collect the background measurements. The location of the background site is shown in Figure 2.0.1. The CPC 3007, sidepak, and all four micro-athelometers were serviced before the start of the campaign. I also regularly changed the wick of CPC 3783 and CPC 3007 at regular intervals of one per month. Also, the sidepaks were zero-calibrated roughly every two months, and the flow rate was checked for the CPC 3007 at least three times during the one-year study.

The central site operated on the top of VTRC Arlington building from 15th September 2021 to 15th September 2022. There were less than seven days when the central site was not running because of maintenance and extreme weather conditions. The central site effectively has one year's worth of data. Figure 2.0.1 shows the location of the central site in Arlington.

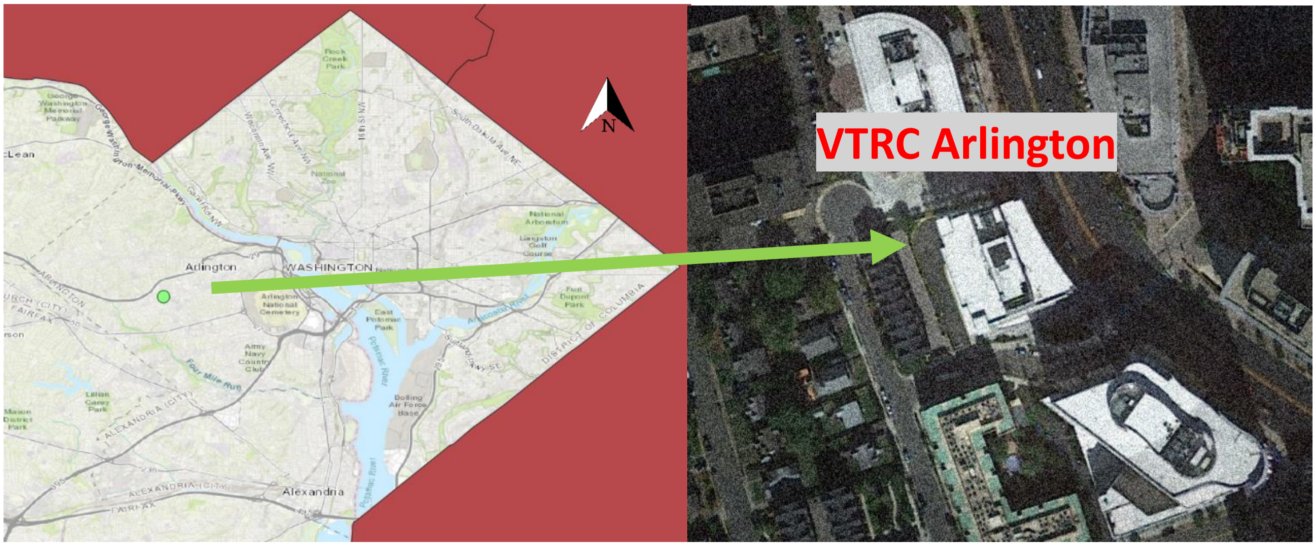


Figure 2.0.1 The location of the central site on a map and from satellite view in Arlington, Virginia

2.1 Background Site Setup

Instruments were located on the top floor penthouse to capture respective environmental parameters. The list of instruments used at the central site are given in Table 2.1.1.

It is also to be noted that while the location of the central site was fixed, instruments like the microaeth, sidepak, and purpleair (PPA) varied in number as some of these devices were rotated with the instruments installed on the bike and some of them were deployed to the other locations. We strived to rotate instruments to the central site regularly to ensure that all instruments were calibrated appropriately throughout the study period. The CPC 3783 and the UPAS were the only instruments that were not used on the bike for data collection.

Table 2.1.1 The instruments located at the central site

Device	Manufacturer	Measurement	Sampling Interval	Number of Instrument	Instrument flow rate	Instrument Range
Ultrasonic Personal Air Sampler	Access Sensor Technologies	$PM_{2.5}$ mass concentration	30 - 72 hours	3	1.0 L/min	-
CPC 3783	TSI, Inc.	Particle number concentration	1 second	1	3.0 L/min	>3 µm
Micro-aethalometer AE-51	AethLabs	Black carbon concentration	1 second	4	100-200 ml/min (50, 150 ml/min also possible)	0-1 mg/m3
SidePak AM520	TSI, Inc.	$PM_{2.5}$ mass concentration	1 second	3	0 - 1.8 liters/min	0.001 - 100 mg/m3
PurpleAir PA-II-SD	PurpleAir, LLC	$PM_{2.5}$ mass concentration, Temperature and Humidity	~2 minute	Up to 25	0.1 L/min	0–500 µg/m3

2.1.1 The Penthouse Setup

Most instruments were running inside the penthouse as the penthouse was air-conditioned and heated. This helped us to collect the data during the entire year that otherwise would not have been possible because of the instrument operational temperature ranges. A picture of the penthouse showing the instruments used at the central site is given in Figure 2.1.1. The instrument includes two CPC 3783, one CPC 3007, four microaeths, and two sidepaks. It is also to be noted that while Figure 2.1.1 show the two CPC 3783s, we used only one CPC 3783 at the central site and kept the other one as a reserve because of the supply chain issues at the time. Also, we used three sidepaks with a rotation of one sidepak to the bike from the penthouse.

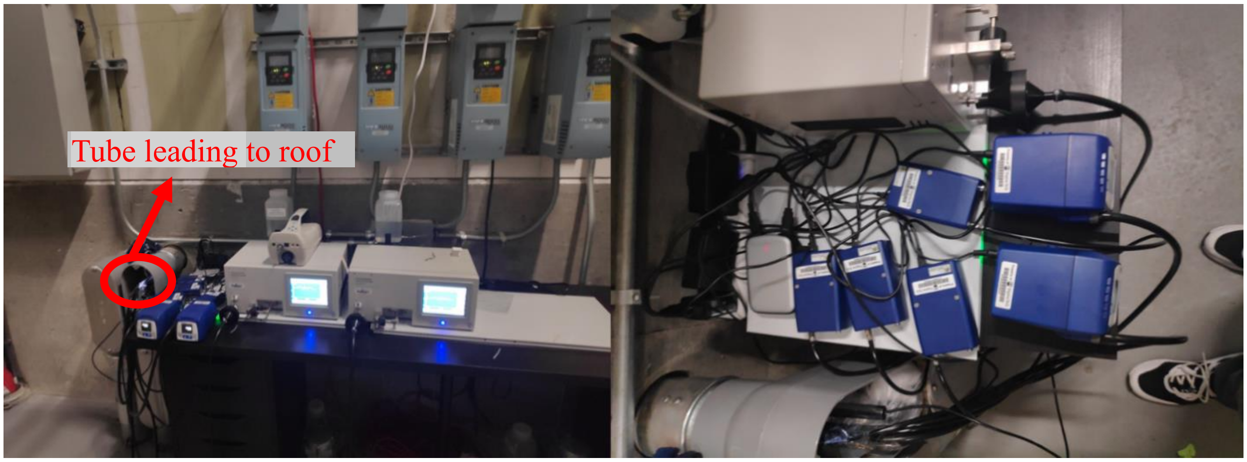


Figure 2.1.1 The setup of instruments at the penthouse of the VTRC building

2.1.2 The Rooftop Setup

The UPAS and the PPA were kept outside the penthouse because there were no tubing arrangement on these devices and they needed to be directly in contact with outdoor air. Figure 2.1.2 shows the outside setup of the PPA and the UPAS device.

Figure 2.1.2 The rooftop device setup for the PPA sensors and UPAS

As shown in Figure 2.1.2, the UPAS was used under a shade to protect the device from rain and snow. The UPAS was also connected with an IoT-compatible Voltaic V50 12,800mAh external battery pack to supply the UPAS with enough power to last ~48 hours. Due to the lack of enough power, we had to use the external power pack with the UPAS as we needed at least 24 hours runtime to collect enough $PM_{2.5}$ weight on the filter. Figure 2.1.3 shows the UPAS used for collecting filter samples.

Figure 2.1.2 The picture of UPAS used in our study

2.2 The Bike Setup

An e-bike was used following the previous methodology (Hankey et al., 2015). I placed instruments in a rear rack and strapped tightly with two layers of foam in between to cushion the shock from road impacts. I used the Specialized Turbo e-bike for sampling. The e-bike was equipped with front shock absorbers to dampen the sudden impact of potholes making the sample collection effort smoother. The position of the instruments installed on the bike is shown in Table 2.2.

Table 2.2 The list of instruments installed on the bike.

Device	**Manufacturer**	**Measurement**	**Sampling Interval**	**Instrument flow rate**	**Instrument Range**
Condensation Particle Counter 3007	TSI, Inc.	Particle number concentration	1 second	700 ml/min	0-100,000 pt/cc[a]
Micro-aethalometer AE-51	AethLabs	Black carbon concentration	1 second	100-200 ml/min (50, 150 ml/min also possible)	0-1 mg/m^3
SidePak AM520	TSI, Inc.	$PM_{2.5}$ mass concentration	1 second	0 - 1.8 liters/min	0.001 - 100 mg/m3
PurpleAir PA-II-SD	PurpleAir, LLC	$PM_{2.5}$ mass concentration, Temperature and Humidity	~2 minute	0.1 L/min	0–500 µg/m3
BU-353	GlobalSat WorldCom Corp.	Global coordinates	1 second (>1 micro-second)	-	-
Hero 9 Black	GoPro	Video	24 frames/sec	-	-

The bike carrier was padded with two layers of 2-inch soft foam to cushion the instruments from mechanical shock. The foam was strapped to the wooden base of the rear rack using parachute buckle straps. A picture of the bike carrier with the different instruments is shown in Figure 2.2.1. The bike carrier was strapped with two portable 10000-watt Anker power banks, one to power the raspberry pi and the other to power the PPA sensor. The list of other devices is given in Table 2.1 and the setup is highlighted in Figure 2.2.1.

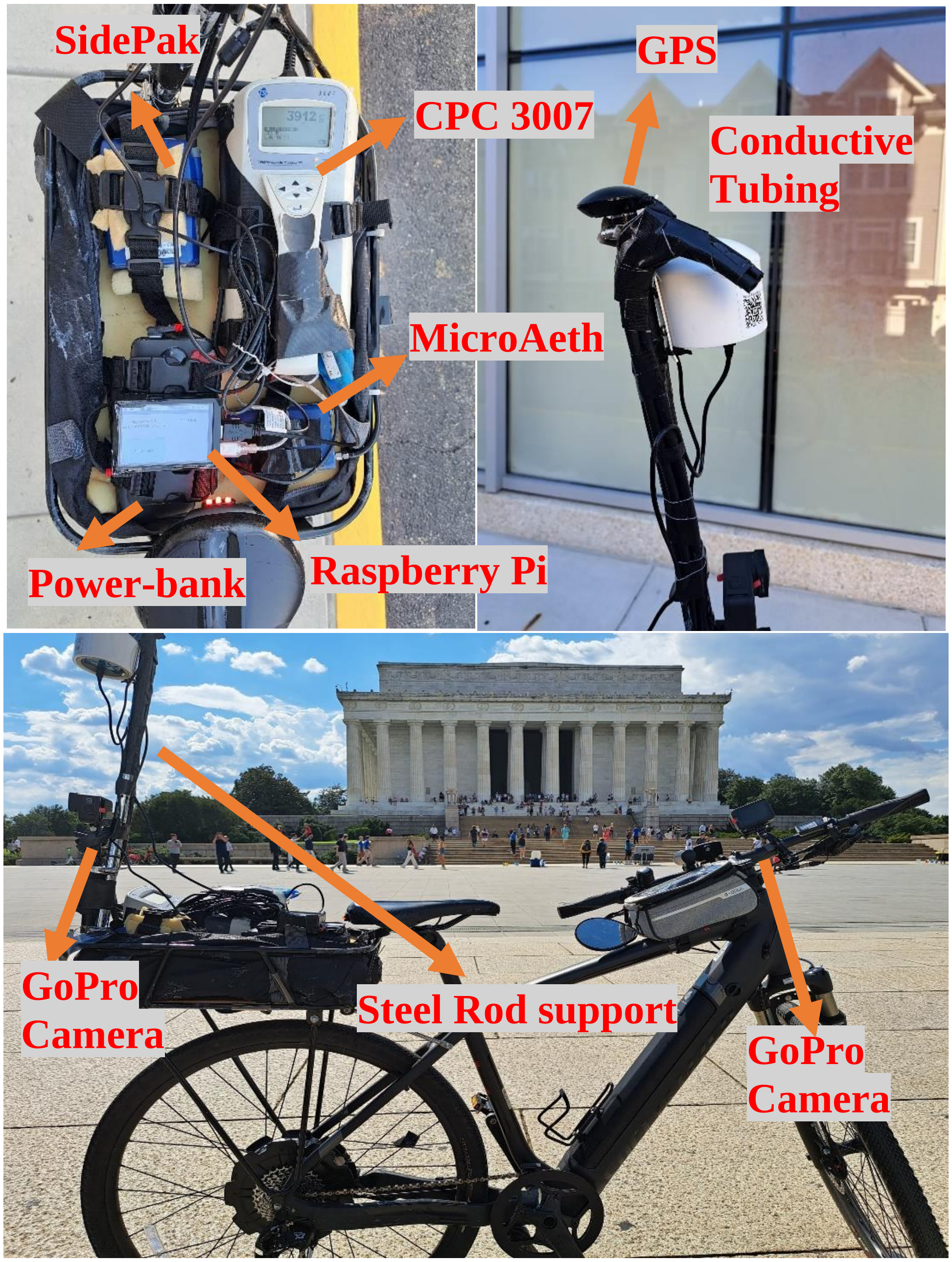


Figure 2.2.1 The instruments fitted on the e-bike for air quality measurement, geolocation and video logging

A square foam was glued above the sidepak AM520 to make up for loose grip because of the small size of the sidepak, the straps were tied carefully before every run to secure and keep clear the outlet for the proper functioning of airflow. A pole was selected that matched the height of the rider. The inlet tube of all three devices (i.e, microaeth, CPC, and the sidepak) was fixed near the top of the steel pole at about six feet from the ground to match the breathing height of the pedestrian and rider so that it matched closely with the actual exposure of the commuters. The GPS was glued over the top of the pole to receive the GPS signal accurately. The PPA sensor was fixed to the pole at the same rider height near the top of the pole. Also, both the tubing and the PPA air inlet were kept pointing downward to minimize wind interference while riding the bike. The CPC tilt alarm was also disabled by soldering and bypassing the circuit in CPC 3007.

2.2.1 The GoPro Setup

The GoPro Hero 9 was used to capture video and sound from the bike routes in conjunction with the air-quality measurements. As shown in Figure 2.2.1, the two opposite-facing cameras were used for recording video with one camera attached to the front handle and the other one attached to the rear pole. The video from both cameras was captured in a 2.7K resolution at a rate of 24 frames per second. The cameras were also fitted with a fisheye camera module from GoPro otherwise, called Max lens to maximize the Field of View (FOV) of the camera and, according to the GoPro community support page, the Hero 9 with max lens has a diagonal FOV of 144 degrees. Ideally, we intended to capture the full panoramic image but because of the structural limitations of the bike setup, we fitted the camera on the front and back of the bike to capture as much of the image as possible. Brief information on the image-based model is in section 2.6.

2.3 The Raspberry Pi Application

The raspberry pi application was developed by the Center for Geospatial Information Technology (CGIT), Virginia Tech team. The application was used to record the data from the CPC 3007, microaeth AE51, and the GPS BU-353 sensors through serial communication over the Universal Serial Bus (USB) cable connected to pi while collecting the data on the bike. The data was collected inside the pi memory card in form of a database file. The data was synced in real-time with the internal pi clock during bike runs. The data from the other instruments not connected to pi like the sidepak and PPA were merged with pi-collected data in the post-processing of the data. Figure 2.3.1 shows the image of the raspberry pi application running on the home screen. The raspberry pi was powered by an on-bike power bank during sample collection. The raspberry pi was also fitted with a touch screen to interact with the application.

Figure 2.3.1 The bike data application running over raspberry pi and showing the live values of CPC, MicroAeth and GPS

2.4 Bike Route Selection

To select the best routes, I primarily used three data sources to optimize the routes. The three data sources were:

1. The Google Street View images and segmentation by PSPNet model
2. Jurisdiction zoning data
3. The road classification

The objective for selecting the above data was to select routes that were the most representative of the DC region and the route selection was done to maximize the representation of each region along with balancing of different factors like safety and feasibility of the route. To obtain the optimum path, I first planned different routes using the Google bike routing tool in such a way that each route was roughly 2 hours long (i.e., Arlington route and Washington DC route). I planned ~15 routes for both the DC and Arlington regions. After selecting the route, I used the google directions API and gpsvisualizer tool to convert the google route links into kml shape files. These shapefiles were then used to extract features from Arlington and Washington DC's local government road data and the zoning data. Figure 2.4.1 shows the flow chart of the route selection process.

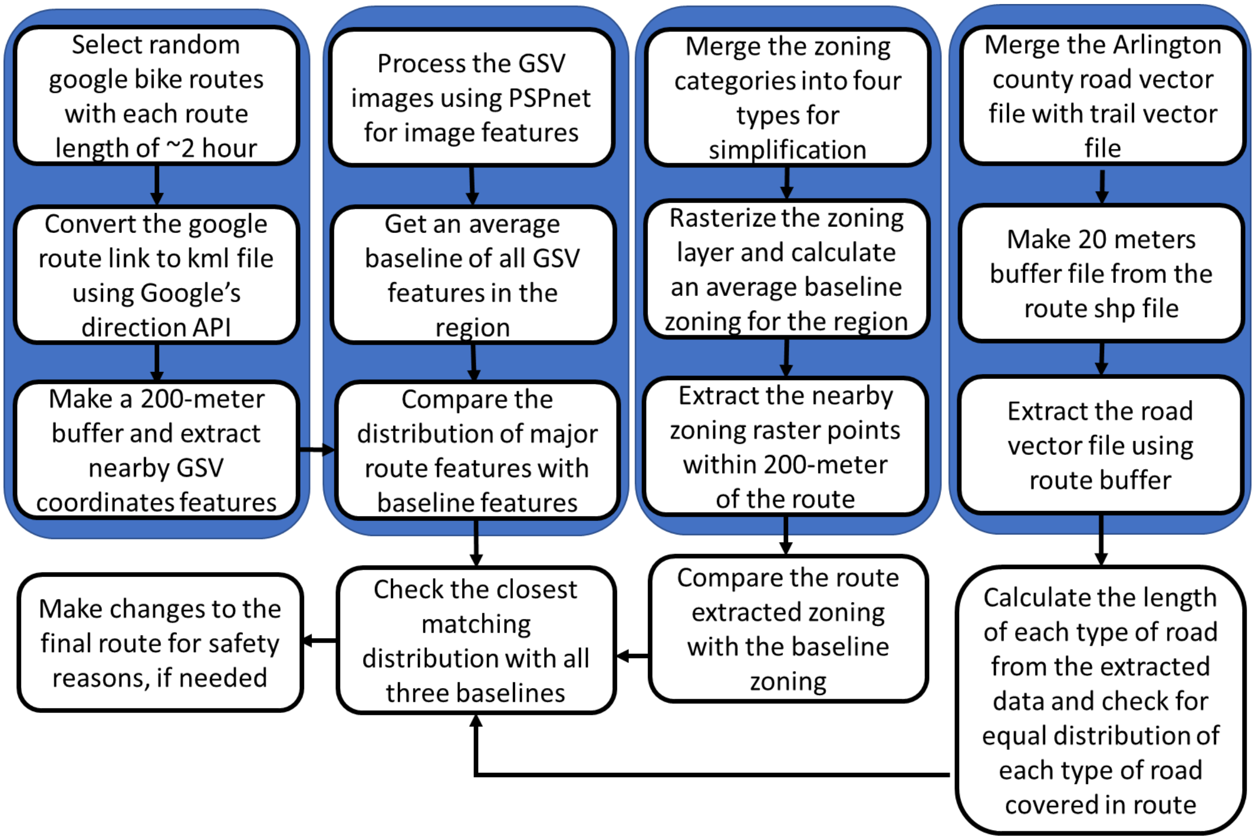


Figure 2.4.1 The flowchart of the route selection process

2.5 Selected Bike Routes

Utilizing the process described in section 2.4 the two routes selected are shown in Figure 2.5.1. It is to be noted that even though one of the routes is called as Washington DC route, it covers both the Arlington and Washington DC area because we used the central site for storing the bike and equipment at the basement parking level of the VTRC building.

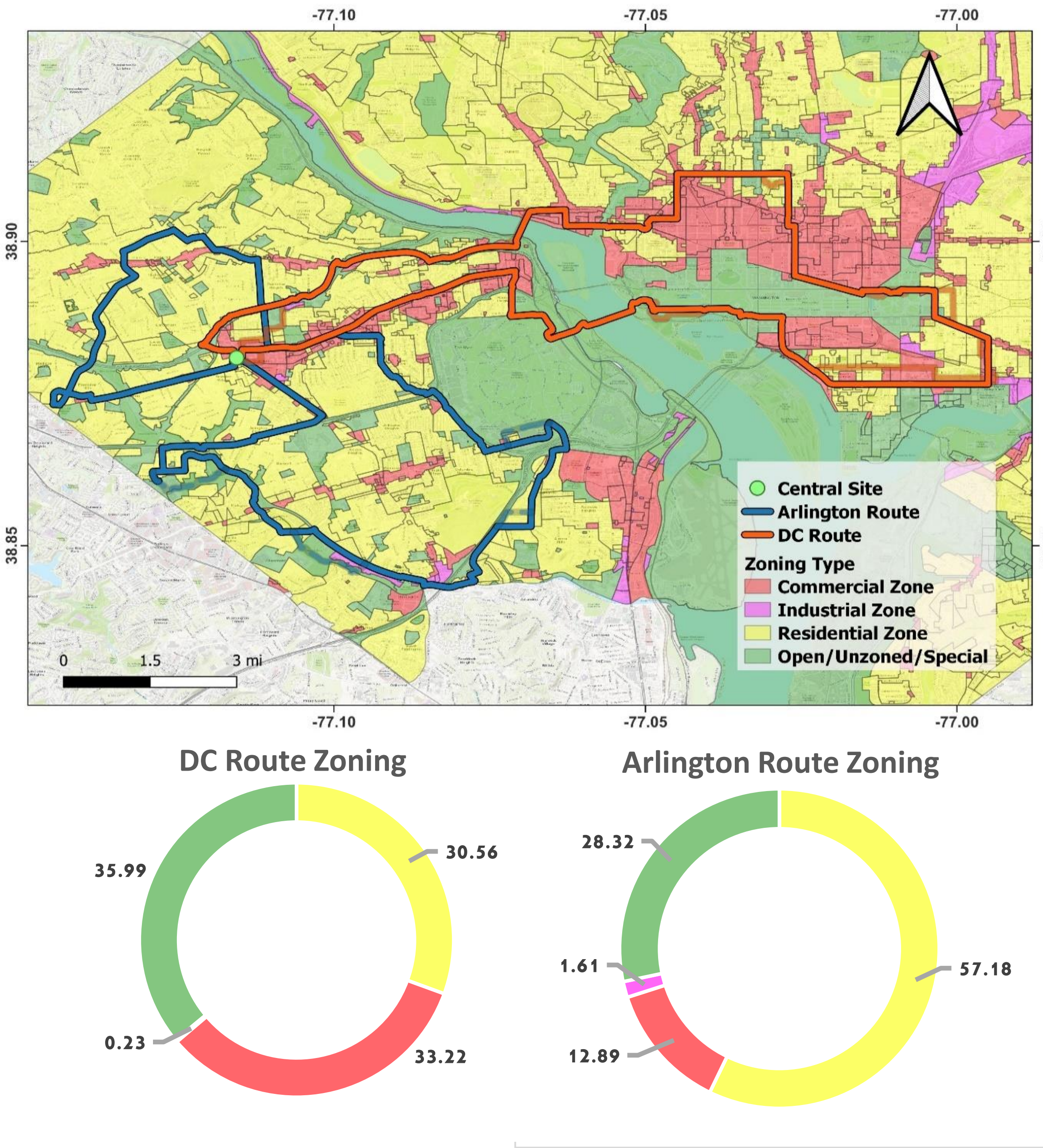


Figure 2.5.1 The zoning distribution of the selected Arlington and Washington DC route. The light colored lines show the frequent detours taken during sample collection

I also consolidated the zoning codes before the extraction of zoning by the routes. The motivation for aggregating the zoning categories is that most jurisdictions in the country have unique zoning ordinance. Thus, we needed to aggregate the ordinance into basic categories among the two jurisdictions: Washington DC and Arlington local government. I merged the residential apartment, and residential zone for the residential category; the downtown zone, mixed-use zone, neighborhood mixed-use zone, and commercial zone were merged into the commercial zone category; the industrial, production, distribution, and repair zones were merged inside the industrial zone category; and the unzoned, roads and parking zone special was merged inside the open zone category.

2.6 GSV Images and PSPNet

The study uses the Pyramid Scene Parsing Network otherwise known as PSPNet (Zaho et al., 2017) for the segmentation of the GSV images. GSV images were downloaded from the Google Streetview API and these images were segmented using the PSPNet model trained over the ADE20K dataset (Zhou et al., 2017). Although the dataset has 150 objects and classes, based on analysis done in GSV model study (Qi et al., 2021), I selected 16 categories that had an impact on the built environment and exposure. These 16 categories are 'wall', 'building', 'sky', 'tree', 'road', 'grass', 'earth', 'field', 'plant', 'water', 'fence', 'person', 'HMV', 'car', 'bikes', and 'ship'. There were 1151 GSV images within the 200-meter buffer for the Washington DC route and 992 GSV images for the Arlington route. The distribution of both routes was compared to the baseline features located in Arlington for the Arlington route and, Washington DC-Arlington combined features for the Washington DC route. The distribution of the 16 selected features is shown in Figure 2.6.2 and Figure 2.6,3. Also, Figure 2.6.1 shows the segmentation of an image captured near VTRC using the GoPro. The features identified by PSPNet are highlighted in discrete colors below the raw image. These features are aggregated in percentage cover in text data from each GSV image and were then utilized for comparing the distribution.

Figure 2.6.1 The image segmentation of a GoPro image by the PSPNet model. The image was captured using GoPro near the VTRC building

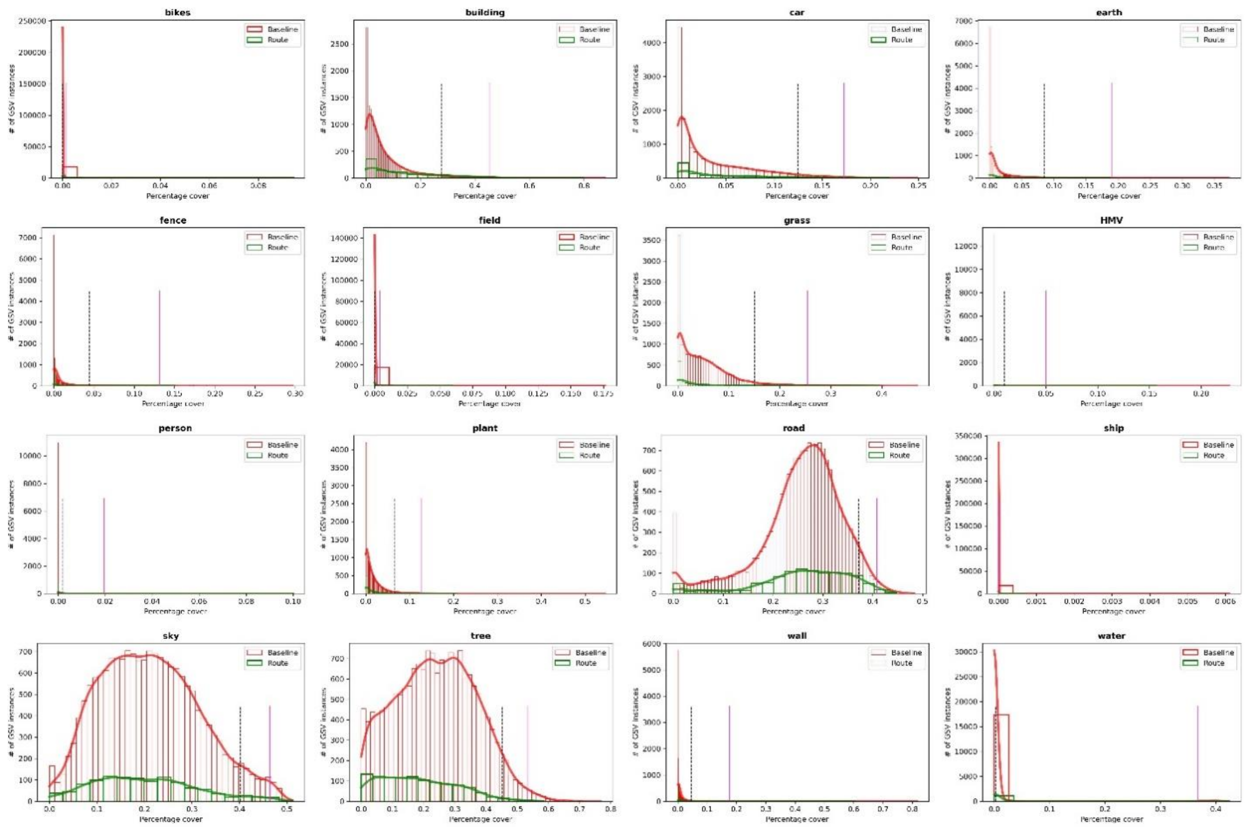


Figure 2.6.2 The distribution of 16 factors compared with the baseline distribution for Washington DC route. The dashed vertical line and solid line represents the 95% and 99% coverage of each class respectively

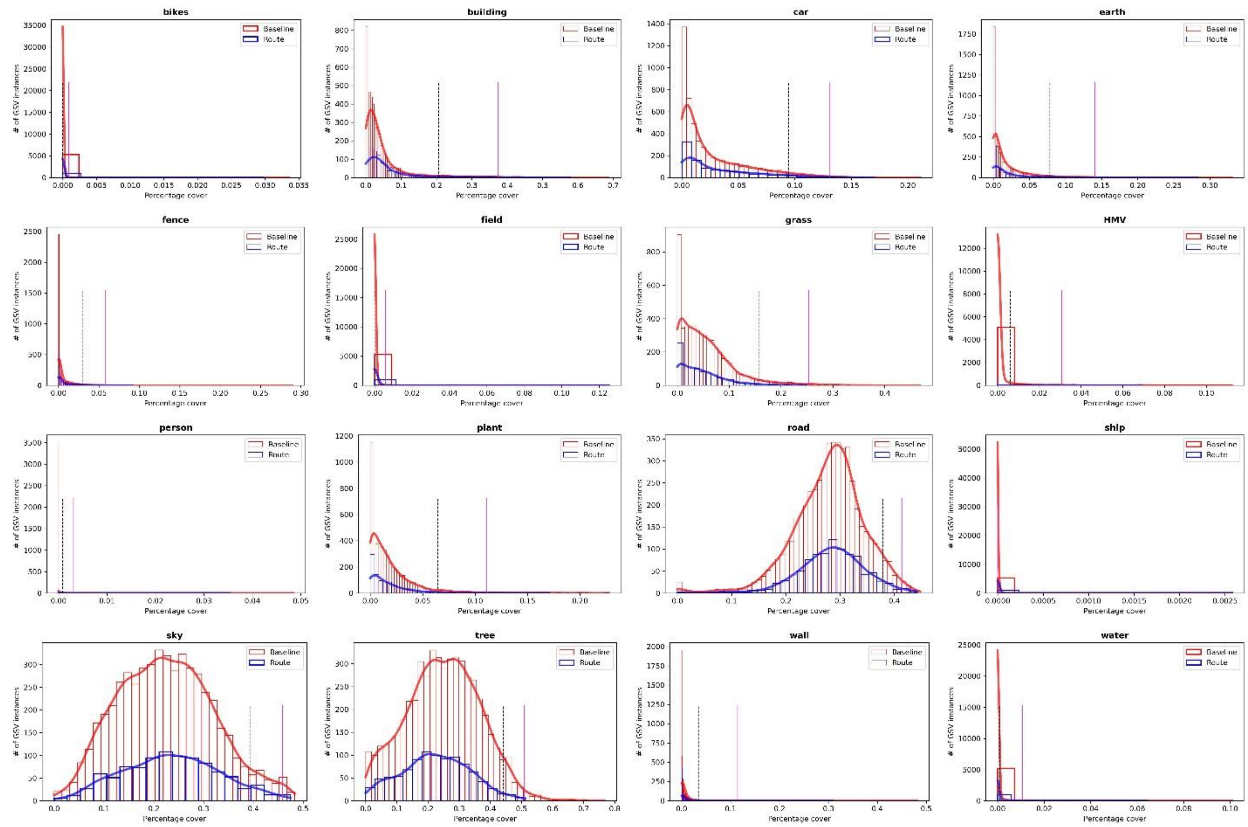


Figure 2.6.3 The distribution of 16 factors compared with the baseline distribution for Arlington route. The dashed vertical line and solid line represents the 95% and 99% coverage of each class respectively

2.7 Road Coverage of Bike Routes

I compared the type of road covered by both routes. The comparison of the routes for the type of roads is given in Figure 2.7.1. I tried to normalize the distribution of the road types in the route selection process but, the trails still represented a slightly higher percentage because of safety reasons as described in the last step of the route selection process in Figure 2.4.1. The road type shown in Figure 2.7.1 is extracted using an algorithm. It was extracted using the selected routes

20 meters buffer shape file. It essentially means that the road which lies within 20 meters of the trail would also be counted in the overall length of each type of road.

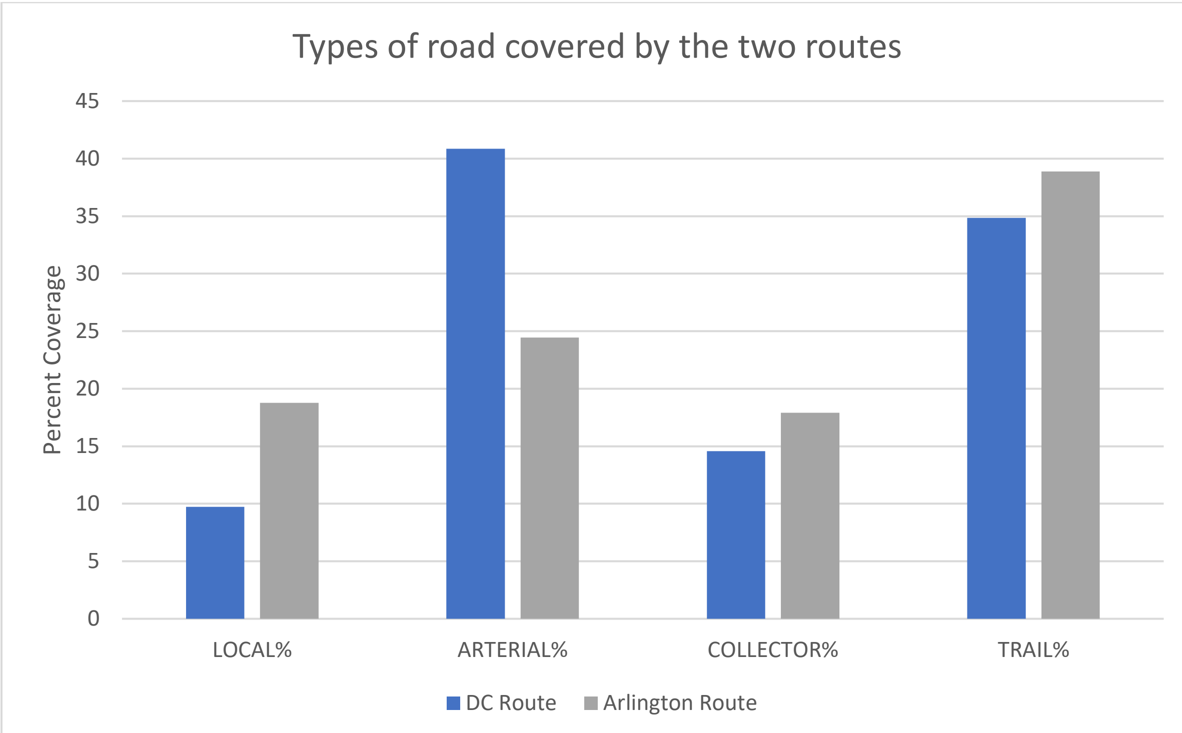


Figure 2.7.1 The type of road covered by each route

3.0 Data Description

This section describes and compares the data collected from the bike routes and the background central site. I was not able to process the data for the microaeth with limited time so, it was not included in this section.

3.1 Central Site Data

The data collected from the central site was continuous except for a few days. To process the central site data, I averaged at a one-hour resolution to keep it standardized for all the devices. The median CPC concentration was observed to be 6365 #/cm^3 and the median $PM_{2.5}$ observed to be 5.2 µg/m^3.

3.1.1 UPAS Data

We collected about 20 UPAS samples but used only 12 to correct the rest of the central site instruments. The UPAS filters were ignored for the days on which the precipitation was more than 0.002. I used ERA5 (Hersbash et al., 2020) reanalysis data to estimate the precipitation over the VTRC building location. We first corrected the sidepak SP_632 and then used SP_632 to correct the rest of the PPA and the sidepak devices. The R-squared between the SP_632 and the UPAS filter was found to be 0.942 with a standard error of 0.03 mg/m3. The intercept was fixed at zero for the OLS linear regression. The scatter plot for the correction of sidepak SP_632 and the UPAS is shown in fig 3.1.1. The correction table for the rest of the $PM_{2.5}$ devices used on the bike is given in Table 3.1.1. The intercept was fixed at zero for all the correction equations and so the intercept is zero for all the devices.

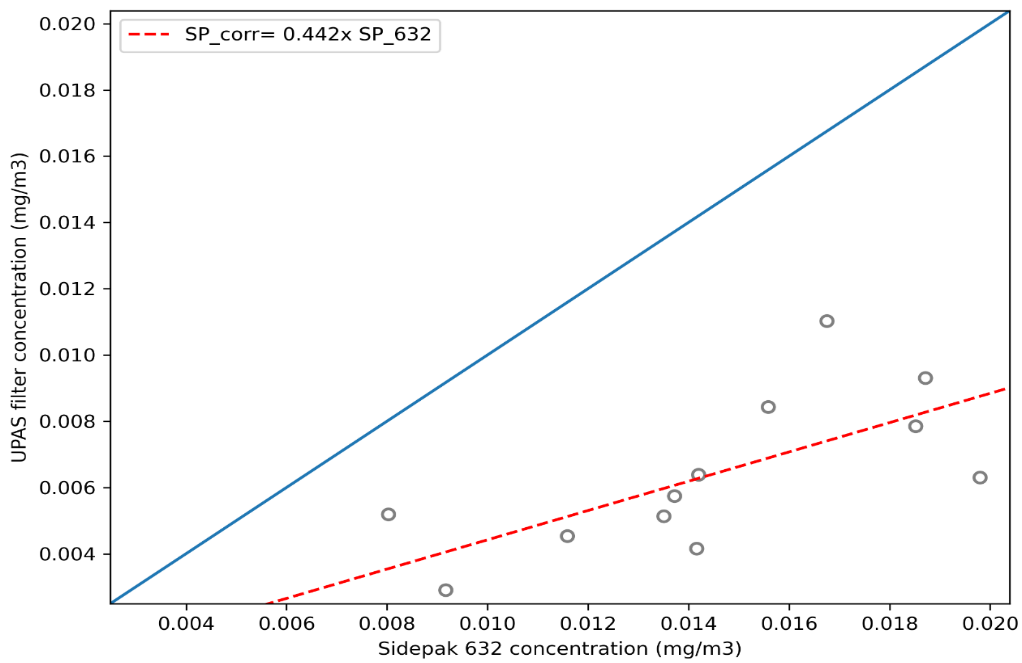


Figure 3.1.1 The correlation between the UPAS and the SidePak SP_632 data for rooftop. The R^2 between UPAS and SidePak SP_632 was observed to be 0.92

Table 3.1.1 The correction factor for all the $PM_{2.5}$ devices used on the bike

Instrument	Reference	ID (X)	Corrected_with (Y)	Slope	Intercept
SidePak	Yes (for all other $PM_{2.5}$ devices)	632	Filter measurement	0.442	NA
SidePak	No	633	SidePak 632 corrected	0.442	NA
SidePak	No	634	SidePak 632 corrected	0.438	NA
PurpleAir	Yes (for all PPA devices)	44_C9_AC	SidePak 632 corrected	0.399	NA
PurpleAir	No	41_8D_13	44_C9_AC corrected	0.382	NA
PurpleAir	No	44_D4_A6	44_C9_AC corrected	0.407	NA
PurpleAir	No	17_65_75	44_C9_AC corrected	0.413	NA
PurpleAir	No	4d_7e_81	44_C9_AC corrected	0.399	NA

3.1.2 The PurpleAir and SidePak Data

The PPA data was filtered for the anomalies between the two channels of the device The channel 'a' and channel 'b' were checked for anomalies after the aggregation of data at one hour resolution and only those data point which showed less than 20% deviation, or 3 micro-gram difference were kept (Malings et al., 2019). Figure 3.1.2.1 shows the correlation Pearson plot and the time series of the PPA sensors. It is to be noted that the data shown in Figure 3.1.2.1 is not adjusted with the filter measurements.

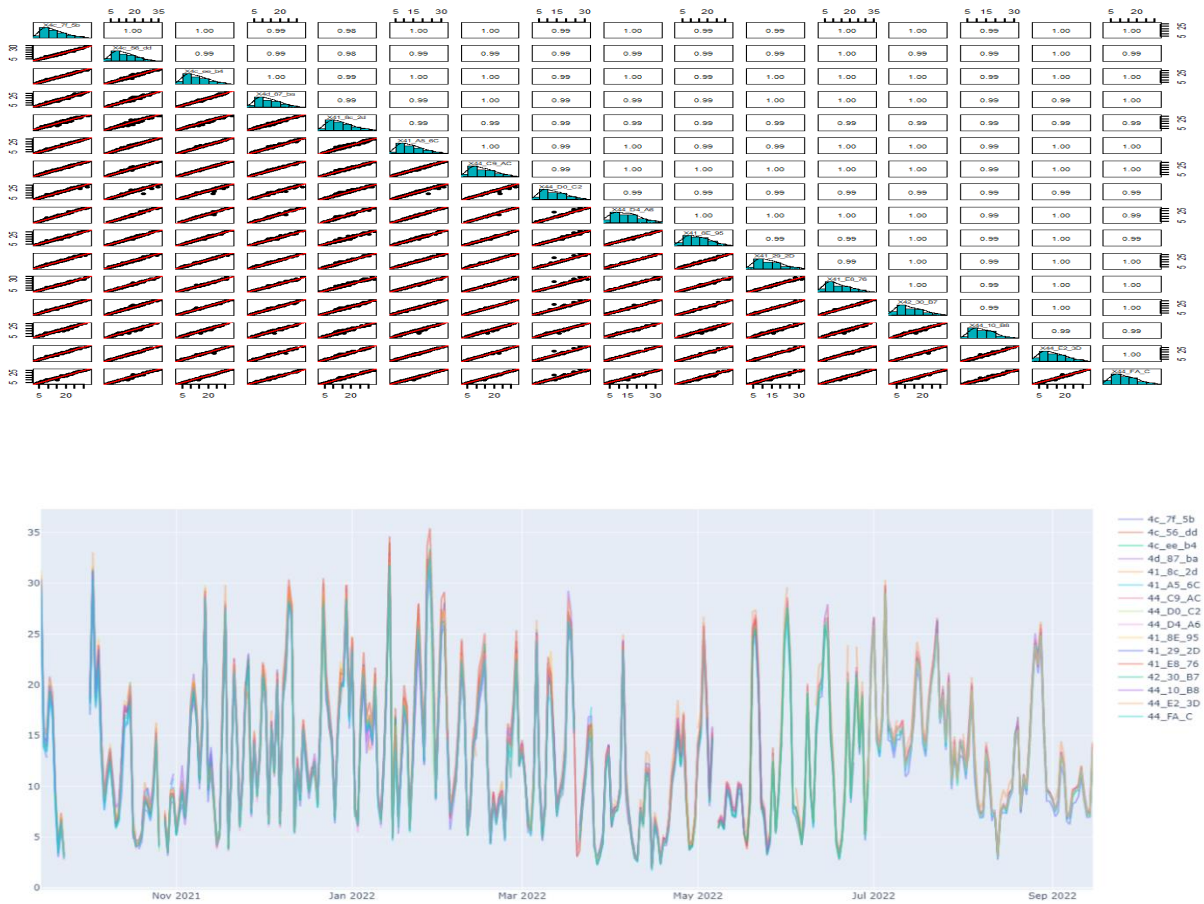


Figure 3.1.2.1 The correlation between the PPA sensors at the central site and the timeseries plot of the uncorrected PPA sensors.

The r-value for every PPA sensor was above 0.98 and there was almost perfect correlation among the PPA devices at the rooftop. A heat plot with all the PPA sensor and the sidepak is given in Figure 3.1.2.2. The r-value for sidepak and PPA sensors was also observed to be above 0.83. Also, it is interesting to note that the correlation between the UPAS and the PPA was always below the correlation between the UPAS and the sidepak. This also shows that the sidepak was more accurate when compared with the sidepak. The scatter plot between sidepak and PPA device '44_C9_AC' is shown in Figure 3.1.1.3.

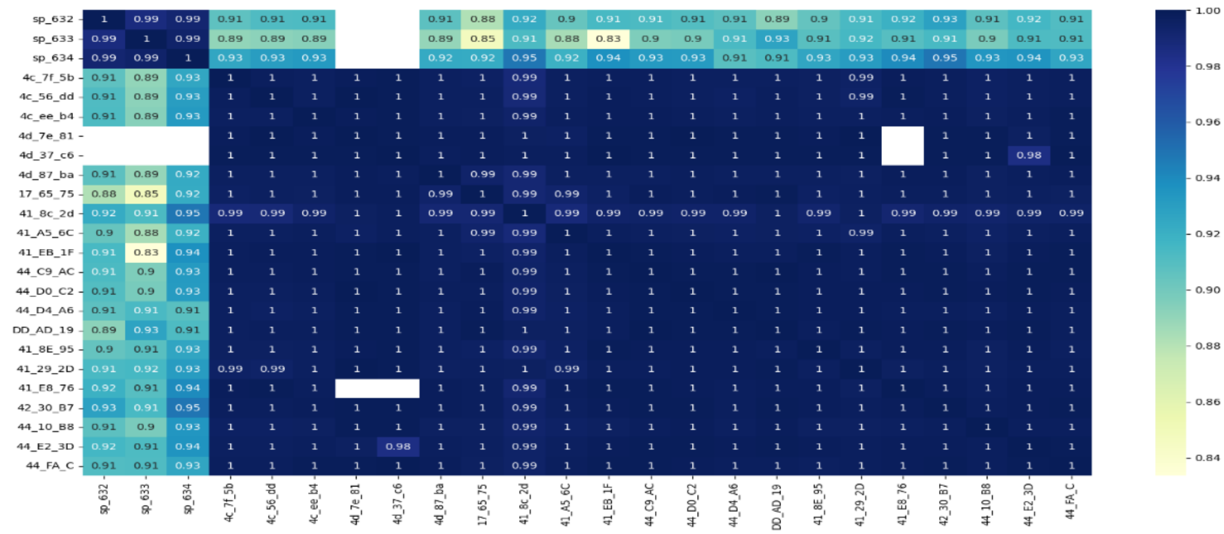


Figure 3.1.2.2 The correlation between the SidePak and the different PPA devices at the central site

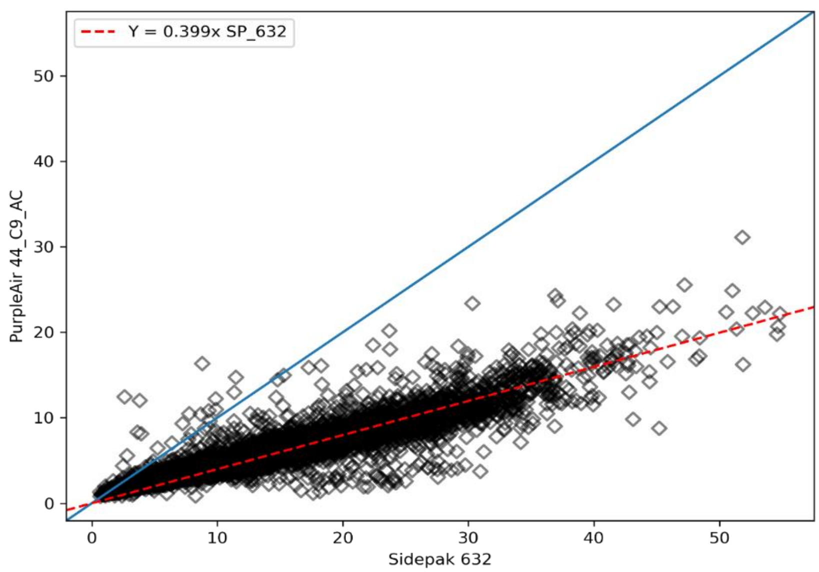


Figure 3.1.2.3 The scatterplot between the SidePak 632 and the PPA 44_C9_AC. The R^2 between the two device was observed to be 0.82

The box plot of all the $PM_{2.5}$ instruments after correction is shown in Figure 3.1.1.4 and as could be observed the median value of the concentration is around 5 μg/m^3.

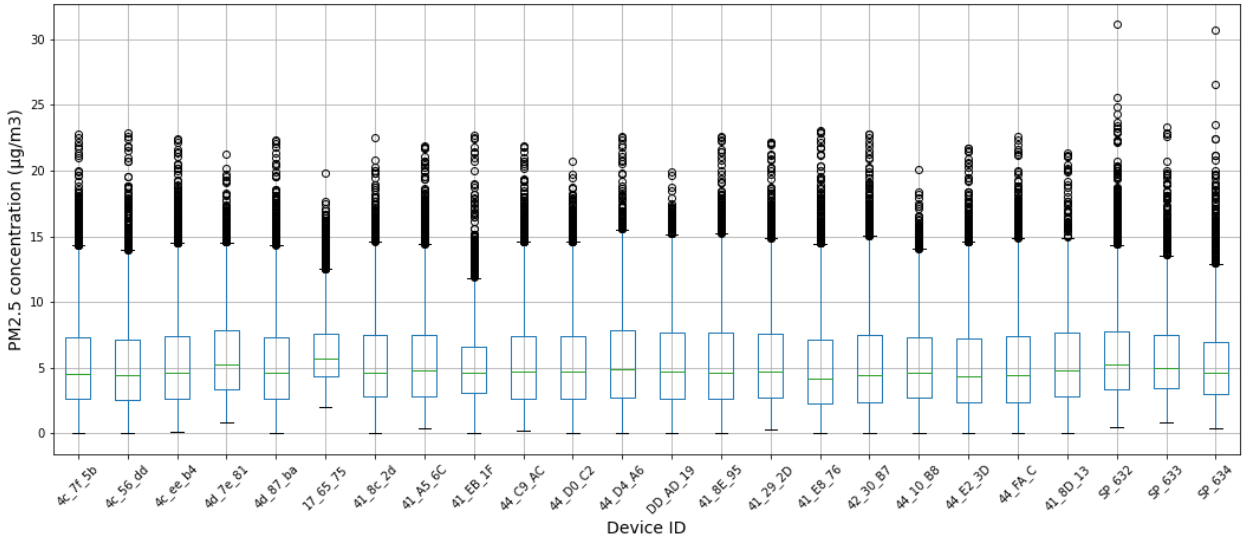


Figure 3.1.2.4 The box plot of the SidePak and PPA sensors after correction at the central site

3.1.3 CPC Data:

The CPC 3783 and the CPC 3007 data scatter plot is shown in Figure 3.1.3.1. We used the CPC 3783 at the central site to be the reference instrument as it had a higher accuracy. The R-squared for both the CPCs was observed to be 0.98. The correction factor for the CPC 3007 is also shown in Figure 3.1.3.1.

Also, for correcting the CPC 3007 data above 100,000 #/cm3 I used the correction factor from from previous study (Westerdahl et al., 2005). Equation 3.1.3 was used to correct the values of CPC 3007 that were above 100,000 #/cm3:

$$PNC_{Corr} = 38456 * \exp(PNCraw * 0.00001) \qquad \text{equation 3.1.3}$$

Where,

PNC_{corr} = The corrected Particle Number Concentration (PNC)

PNC_{raw} = The raw concentration as logged by CPC 3007

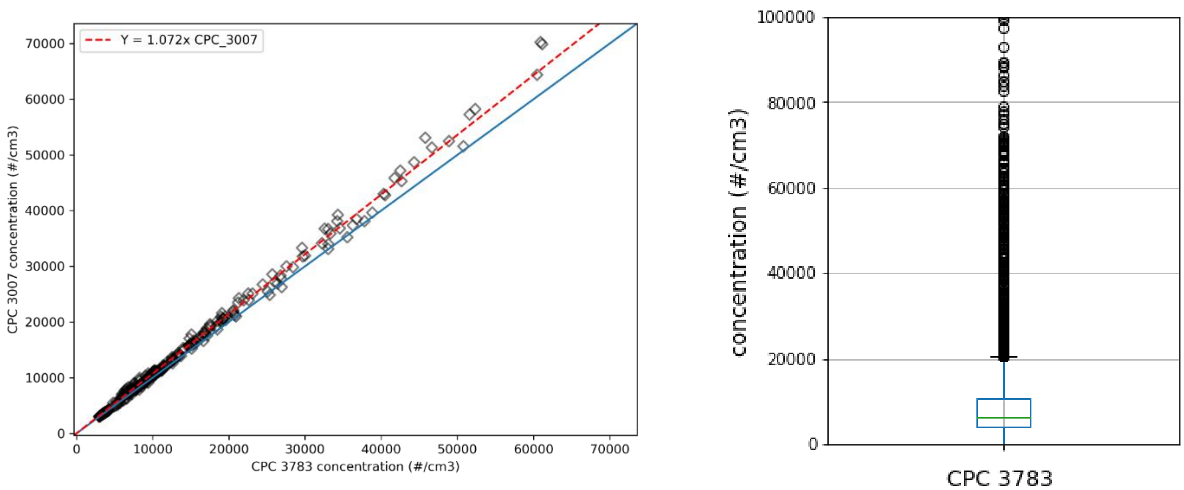
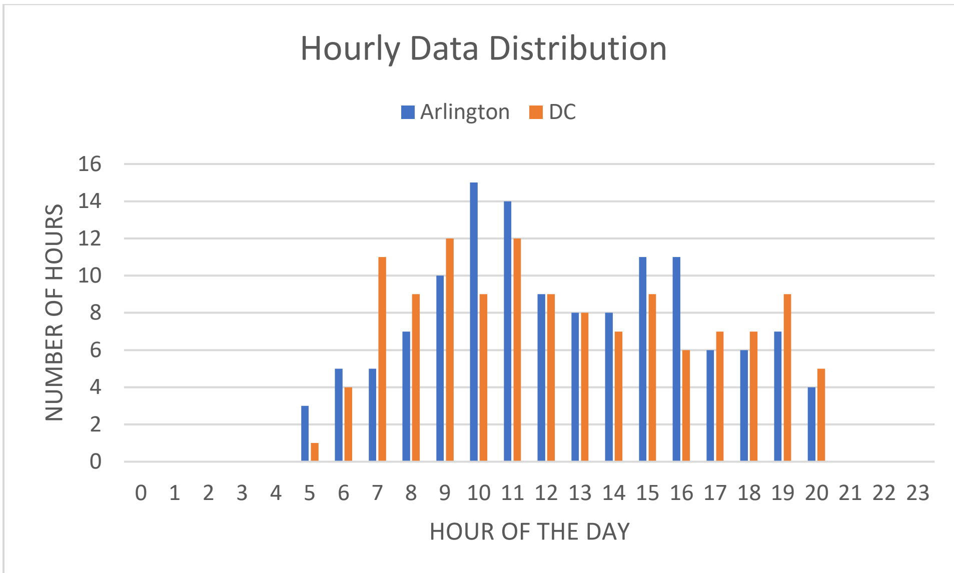


Figure 3.1.3.1 The correlation between the CPC 3007 and the CPC 3783 at the central site and the distribution of one year data collected from the CPC 3783 at the central site.

3.2 The Bike Data Description

The data collected on the bike was for 104 runs in total which included four random sampling runs that could be utilized later on for evaluating the LUR models. Each run was ~2 hours long so we roughly recorded ~208 hours of data from the Arlington and the Washington DC route. The bike data was collected from 4th November 2021 to 17th September 2022. The hourly, weekday, and monthly distribution of the data collected in each unit hour slot is shown in Figure 3.2.1 and Figure

3.2.2. We tried to balance the sample distribution as much as possible. The bike runs were done from 5 am morning to 9 pm. No runs were completed at night because of low activity and safety reasons.



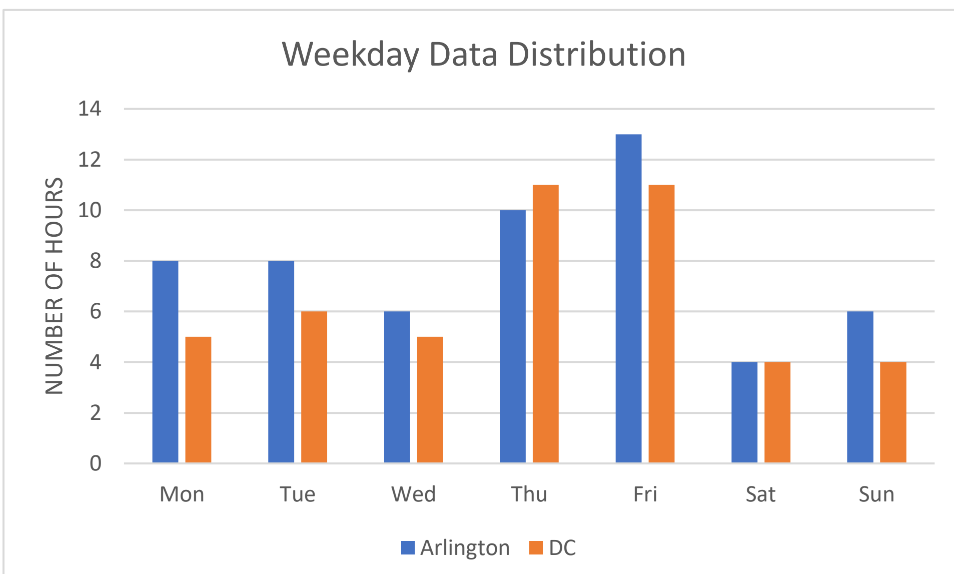


Figure 3.2.1 The hourly and weekday distribution of bike sample runs for Arlington and Washington DC route

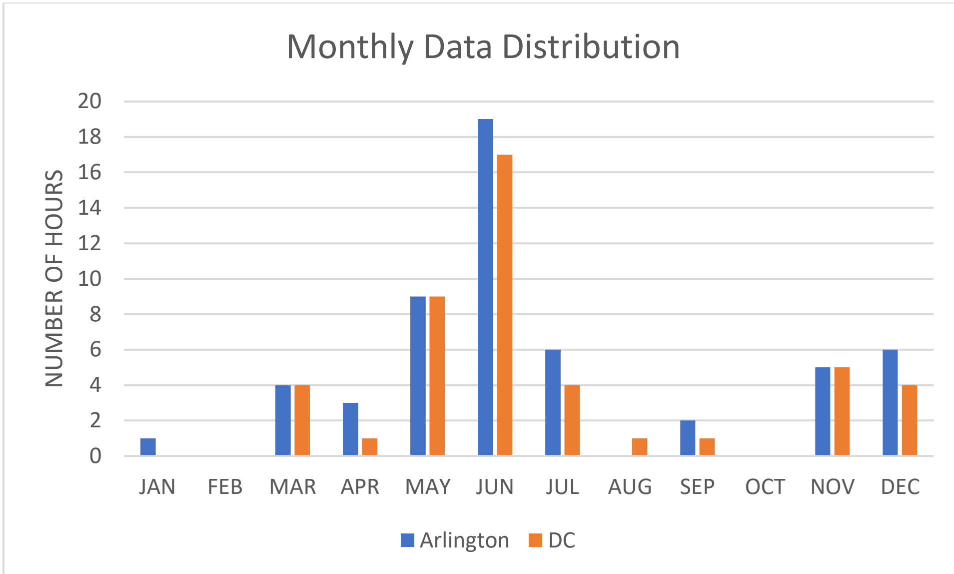


Figure 3.2.2 The monthly distribution of bike sample runs for Arlington and Washington DC route

The descriptive statistics for both routes are given in Table 3.2.1. It is to be noted that Table 3.2.1 shows data only from the main route and does not consider the detours and random routes that were taken during the bike runs. The values shown in Table 3.2.1 were corrected with the reference instruments as given in Table 3.1.1 before the analysis.

Table 3.2.1 The data description as collected from the two routes and the central site

Measurement	Device	Route/location	Count	Mean	Standard Dev	Minimum	25%	50%	75%	Maximum
PNC	CPC 3007	Arlington	194182	88646.9	5092090	2.1	4485.2	6587.4	10869	7.67E+08
PNC	CPC 3007	Wahington DC	180606	32790.8	2453314	5.4	5574.4	8504.2	13991.7	4.21E+08
PNC	CPC 3783	Central Site	-	8867.9	9970	0.2	3891.7	6365.3	10595.2	125692.6
$PM_{2.5}$	SidePak	Arlington	175488	7	5	0.5	4.2	6.2	9.1	1180.8
$PM_{2.5}$	SidePak	Wahington DC	160868	6.4	3.7	1.3	3.7	5.5	8.2	342.6
$PM_{2.5}$	SidePak	Central Site	-	6.0	3.5	0.5	3.3	5.2	7.8	31.1

As can be observed from Table 3.2.1, the average at the central site is less than the average of both routes. Both PNC and $PM_{2.5}$ showed a lower concentration at the central site when compared to the bike runs.

For spatial display, the point data was extracted inside a raster of 100-meter resolution. The mean concentration value was assigned to each grid cell. The spatial raster distribution for PNC and $PM_{2.5}$ is shown in Figure 3.2.3.

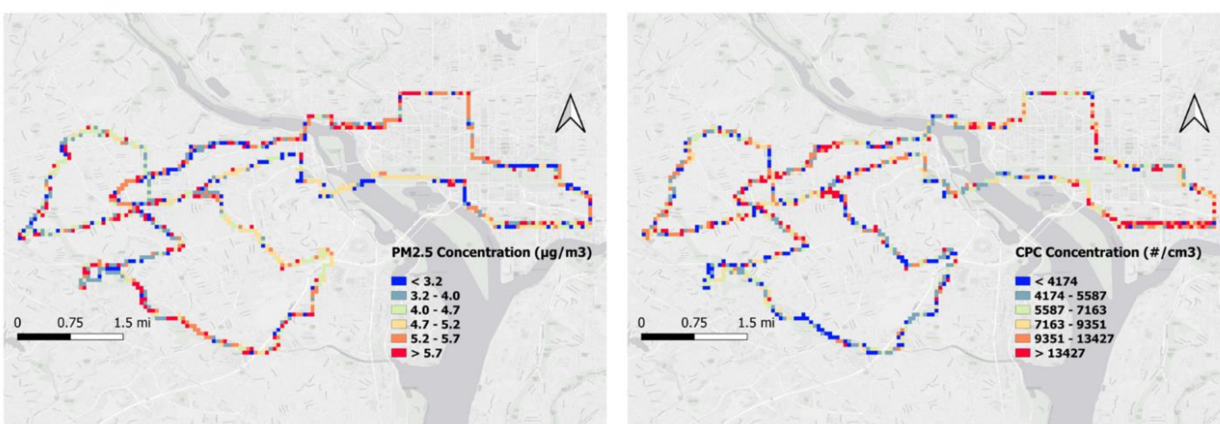


Figure 3.2.3 The spatial distribution of $PM_{2.5}$ and PNC for the two bike routes.

3.2.1 Preliminary Results and Observations

After data cleaning and plotting the spatial data as shown in Figure 3.2.3, it can be observed that the special and unzoned areas in Washington DC and Arlingtonhad low concentrations. This unzoned and special area includes the Washington Monument, the area near Arlington Cemetery, and the open area near the Potomac River. Also, for the Washington DC route, the downtown area seems to have the highest concentration of PNC and $PM_{2.5}$.

The Arlington route had a more similar spread-out concentration when compared to the Washington DC route. The lower concentration for both the PNC and $PM_{2.5}$ is near the "4-mile run" trail and the higher concentration is observed near the downtown area and near the highway.

4.0 Conclusion and Lessons Learned

The methods used in my study are novel and can help one to select the best representative routes for mobile campaigns. The use of GSV images to select a route can help in the identification of potential built environments that are not captured by zoning information, which is traditionally used in the selection of mobile monitoring routes.

I have also shown that the background site registered 0.4 to 1 microgram less $PM_{2.5}$ mass concentration when compared to the DC and the Arlington route. During the campaign, I learned some important lessons that could help other researchers avoid potential pitfalls. These lessons are:

1. After the setup is complete on the bike, one should make a standard operating procedure for initiating the run and terminating the run. During the initial days, I sometimes forgot one or two things which often created problems. It was primarily because there were a lot of moving parts like batteries for cameras, e-bike batteries, two power banks, CPC AA batteries, two SD cards for the camera, and bike lights for safety.
2. One should make sure that enough random runs are completed during the campaign for validation of the models. In our case, we could only collet four samples runs or, ~4% of the total data.
3. The PPA running status was hard to verify as the device had no indicators outside the device so, one should make sure that the cables are connected well before and during the bike runs.

4. The GoPro needs time calibration after each run and if we do not sync the time between two GoPro after changing the battery, it develops a time shift that could be challenging to remove in post-processing of the data.
5. One should get the extra parts before initiating the campaign. This was one point that was especially disruptive as the supply chains were slow during the campaign, and we could not get some devices like the extra UPAS and the extra CPC on time. This risked the bike campaign if any of our devices were to break down during sampling.
6. Check for local factors while selecting a route. We selected the routes primarily based on the algorithmic decision, but it should be noted that some information is not available with Google maps. In our case, Google maps put me on a dangerous road during the random sampling runs that caught me by surprise.

The next steps in this project are to input this data into the LUR models. This data will include the data collected from the mobile monitoring campaign and the data collected at the central site. We will also use the video data that we collect during the mobile measurement. Once the model development is finished, we will validate our models with the data collected from the random routes and the data collected from the PPA network spread in Arlington and Washington DC.

References:

Alas, H. D., Stöcker, A., Umlauf, N., Senaweera, O., Pfeifer, S., Greven, S., & Wiedensohler, A. (2021). Pedestrian exposure to black carbon and PM2.5 emissions in urban hot spots: New findings using mobile measurement techniques and flexible bayesian regression models. *Journal of Exposure Science & Environmental Epidemiology*, *32*(4), 604–614. https://doi.org/10.1038/s41370-021-00379-5

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C. H., & Hamburg, S. P. (2017). High-resolution air pollution mapping with google street view cars: Exploiting big data. *Environmental Science & Technology*, *51*(12), 6999–7008. https://doi.org/10.1021/acs.est.7b00891

Chambliss, S. E., Pinon, C. P. R., Messier, K. P., LaFranchi, B., Upperman, C. R., Lunden, M. M., Robinson, A. L., Marshall, J. D., & Apte, J. S. (2021). Local- and regional-scale racial and ethnic disparities in air pollution determined by long-term mobile monitoring. *Proceedings of the National Academy of Sciences*, *118*(37). https://doi.org/10.1073/pnas.2109249118

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., … Forouzanfar, M. H. (2017). 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*, *389*(10082), 1907–1918. https://doi.org/10.1016/s0140-6736(17)30505-6

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., … Yan, C. (2022). Pollution and health: A progress update. *The Lancet Planetary Health*, *6*(6). https://doi.org/10.1016/s2542-5196(22)00090-0

Hankey, S., & Marshall, J. D. (2015). Land use regression models of on-road particulate air pollution (particle number, black carbon, PM2.5, particle size) using mobile monitoring. *Environmental Science & Technology*, *49*(15), 9194–9202. https://doi.org/10.1021/acs.est.5b01209

Qi, M., & Hankey, S. (2021). Using street view imagery to predict street-level Particulate Air Pollution. *Environmental Science & Technology*, *55*(4), 2695–2704. https://doi.org/10.1021/acs.est.0c05572

Qi, M., Dixit, K., Marshall, J. D., Zhang, W., & Hankey, S. (2022). National land use regression model for no2 using street view imagery and satellite observations. *Environmental Science & Technology*, *56*(18), 13499–13509. https://doi.org/10.1021/acs.est.2c03581

Samad, A., & Vogt, U. (2020). Investigation of urban air quality by performing mobile measurements using a bicycle (MOBAIR). *Urban Climate*, *33*, 100650. https://doi.org/10.1016/j.uclim.2020.100650

Volckens, J., Quinn, C., Leith, D., Mehaffy, J., Henry, C. S., & Miller-Lionberg, D. (2016). Development and evaluation of an ultrasonic personal aerosol sampler. *Indoor Air*, *27*(2), 409–416. https://doi.org/10.1111/ina.12318

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2881-2890).

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ade20k dataset. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 633-641).

Burrowes, V. J., Piedrahita, R., Pillarisetti, A., Underhill, L. J., Fandiño-Del-Rio, M., Johnson, M., Kephart, J. L., Hartinger, S. M., Steenland, K., Naeher, L., Kearns, K., Peel, J. L., Clark, M. L., Checkley, W., Nambajimana, A., Verma, A., Lovvorn, A., Diaz, A., Papageorghiou, A., … Sakas, Z. (2020). Comparison of next-generation portable pollution monitors to measure exposure to pm 2.5 from household air pollution in Puno, Peru. *Indoor Air*, *30*(3), 445–458. https://doi.org/10.1111/ina.12638

Convert your GPS data for use in Google Earth. GPS Visualizer. (n.d.). Retrieved December 17, 2022, from https://www.gpsvisualizer.com/map_input?form=googleearth

Dao, M.-S., Zettsu, K., & Rage, U. K. (2021). Image-2-AQI: Aware of the surrounding air qualification by a few images. *Lecture Notes in Computer Science*, 335–346. https://doi.org/10.1007/978-3-030-79463-7_28

Google. (n.d.). *Google Direction API*. Google. Retrieved December 17, 2022, from https://developers.google.com/maps/documentation/directions/overview

Hankey, S., & Marshall, J. D. (2015). On-bicycle exposure to particulate air pollution: Particle number, black carbon, PM 2.5 , and particle size. *Atmospheric Environment*, *122*, 65–73. https://doi.org/10.1016/j.atmosenv.2015.09.025

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., & Subramanian, R. (2019). Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation. *Aerosol Science and Technology*, *54*(2), 160–174. https://doi.org/10.1080/02786826.2019.1623863

Pillarisetti, A., Carter, E., Rajkumar, S., Young, B. N., Benka-Coker, M. L., Peel, J. L., Johnson, M., & Clark, M. L. (2019). Measuring personal exposure to fine particulate matter (PM2.5) among rural Honduran women: A field evaluation of the ultrasonic personal aerosol sampler (UPAS). *Environment International*, *123*, 50–53. https://doi.org/10.1016/j.envint.2018.11.014

Vernooij, R., Winiger, P., Wooster, M., Strydom, T., Poulain, L., Dusek, U., Grosvenor, M., Roberts, G. J., Schutgens, N., & van der Werf, G. R. (2022). A quadcopter unmanned aerial system (UAS)-based methodology for measuring biomass burning emission factors.

Atmospheric Measurement Techniques, *15*(14), 4271–4294. https://doi.org/10.5194/amt-15-4271-2022

Westerdahl, D., Fruin, S., Sax, T., Fine, P. M., & Sioutas, C. (2005). Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmospheric Environment*, *39*(20), 3597–3610. https://doi.org/10.1016/j.atmosenv.2005.02.034

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., & Künzli, N. (2012). Effect of the number of measurement sites on land use regression models in estimating local Air Pollution. *Atmospheric Environment*, *54*, 634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064

Arku, R. E., Birch, A., Shupler, M., Yusuf, S., Hystad, P., & Brauer, M. (2018). Characterizing exposure to household air pollution within the prospective urban rural epidemiology (PURE) study. *Environment International*, *114*, 307–317. https://doi.org/10.1016/j.envint.2018.02.033

Ganji, A. *et al.* (2020) “Predicting traffic-related air pollution using feature extraction from Built Environment Images,” *Environmental Science & Technology*, 54(17), pp. 10688–10699. Available at: https://doi.org/10.1021/acs.est.0c00412.