Essays on Mapping and Improving Urban Air Quality Monitoring in a Developing Country Setting: The Case of Ethiopia

A Dissertation

Submitted to the National Graduate Institute for Policy Studies (GRIPS)

in Partial Fulfillment of the Requirements for the Degree of

Ph.D. in Development Economics

by

Degenet Shifraw Baue

2023

GRIPS

Abstract:

The growth of urban areas and industrial development has diminished air quality across various cities globally, especially in low-and middle-income nations. This decline severely affects health and overall quality of life, resulting in an estimated annual death toll of three to seven million. Although the concentration of particulate matter (PM2.5) varies over time and space, its large-scale monitoring poses a challenge due to its complexity and cost.

The first paper of this dissertation presents a low-cost exercise in the measurement of air quality to create a map of PM2.5 for Addis Ababa, Ethiopia. The data were collected during the dry and rainy seasons in various land-use areas and analyzed using a land-use regression (LUR) model. The results show that PM2.5 levels are affected by local land use, traffic, and weather conditions. Higher PM2.5 concentrations were found in commercial areas and areas with heavy road traffic, highlighting the importance of land use type in predictions of particulate matter levels. Additionally, meteorological factors significantly impacted concentration levels more than other factors, such as land use type. The methodology used here can be replicated in studies of other urban areas in developing countries where air quality monitoring is limited.

Despite the health hazards related to air pollution, low- and middle-income countries lack air quality monitoring equipment and information. The US currently has fixed air quality monitors in its embassies in developing countries, but the data from single fixed sites provide only limited information and does not accurately reflect exposure levels in areas away from the Embassy. The second paper of this dissertation examines the extent to which a single site

placed arbitrarily at the US Embassy provides a guide to air quality information and its risks across the city. The discrepancy between air quality readings from a single monitor and readings from a low-cost monitor in four sub-cities of Addis Ababa is analyzed. The study used linear difference and multinomial logit estimations to identify the sources of discrepancies in the readings. The results show that the discrepancies are primarily influenced by land use, time, and weather variables, with the time of day playing the most significant role in the discrepancy. The findings suggest that relying on data from one air quality monitor positioned randomly does not provide a complete understanding of the intricate nature of air pollution. To achieve more precise measurements in the future, we propose a technique for positioning fixed air quality monitors to optimize monitor placement through minimizing the total expected cost of the discrepancies. This approach aids in achieving a more thorough decision-making process regarding air quality. Overall, the results of this study shed light on the challenges of air quality monitoring in developing countries and offer solutions for improving the accuracy and accessibility of air quality information.

This dissertation is dedicated to my late Father, Ato Shiferaw Bayu, an excellent

smallholder farmer who has always been a light in times of darkness,

Acknowledgment:

I am immensely grateful to my main advisor, Prof. Alistair Munro, for his exceptional support and guidance throughout the research process and for being there for me during my difficult times with his fatherly approach.

 I am also grateful to my sub-advisors, Prof. Akio Yamazaki and Prof. Yukihiro Kidokoro, for their encouragement, support, insightful comments, and guidance during challenging times. I express my gratitude towards my professors at Addis Ababa University, Dr. Assefa Admassie, Dr Kefyalew Endale, Dr. Yabibal Mulualem, Dr. Wassie Birhanu, and Dr. Tadele Ferede for their motivation and support that helped me pursue a Ph.D. study. Families, friends, and classmates supported me during every stage of my study, notably my sister Mulu Shiferaw, my brothers Getnet Shiferaw, Abey Getahun and Tassew Damtew for their positive spirits when I was in a difficult situation. Abebe Ambachew, Mesfin Sahle, and Alebel Melaku supported me all the way; I thank Aregawi Gebremedhin for his brotherhood support starting from my undergraduate class.

I would like to express my gratitude for the excellent class discussions and group work conducted by my classmates. Special thanks go to Pham Phuong Ngoc, MADRIAGA Cherry Ann Dulay, Somsiriwong Piyakul, Pisitsupakul Kritsanee, Chandan Sapkota, Abinash Dash and Constance Sorkpor for their valuable contributions. I am deeply grateful to my wife, Adanech Berta, for her prayers, patience, and support during my challenging times. I would also like to thank the Ethiopian community at St. Michael church in Tokyo for their prayers and positive spirits. I thank the staff members of GRIPS student office and research support

Table of contents

List of tables

List of Figures

CHAPTER ONE

1. INTRODUCTION
1.1. Overview

Air pollution stands as a significant environmental apprehension in the contemporary world. In 2016 the World Health Organization (WHO) data shows that a staggering 91% of the global population resided in regions where air quality fell below recommended standards. Alarmingly, outdoor air pollution led to the premature deaths of 4.2 million individuals during that year. [1] Notably, cities in low and middle-income countries (LMICs) tend to grapple with more pronounced air pollution compared to their developed counterparts.

Those who live, work, socialize, and commute in densely urbanized LMIC areas face escalated exposure to air pollution due to elevated concentration levels. Despite the evident link between air pollution and human well-being, air quality data and monitoring systems remain deficient in many LMICs. This deficiency translates to an inadequate grasp of exposure levels, concentrations, and pollution sources. The outcome is compromised decision-making and insufficient efforts to mitigate air pollution and enhance public health.

Initiating a response to this quandary necessitates intensified research endeavors and the application of potent air pollution control strategies and technologies. Furthermore, cultivating awareness among the general public and policymakers regarding the severe health

---

[1] https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

ramifications of air pollution can be instrumental in galvanizing change and promoting investments in clean air initiatives.

Addressing this gap, the first paper of this dissertation employs a low-cost monitoring approach. By correlating the monitored pollution levels with spatial land use data, this method predicts pollution concentrations in unmonitored locations within Addis Ababa, Ethiopia's capital. The findings underscore the influence of local land usage, traffic, and weather conditions on PM2.5 levels. Notably, elevated PM2.5 concentrations are discerned in commercial zones and areas characterized by heavy vehicular movement. This accentuates the pivotal role of land usage in predicting particulate matter levels. Moreover, meteorological factors wield a substantial impact on concentration levels, surpassing the influence of other factors like land use patterns.

While low-cost monitors can help predict air pollution in unmonitored locations with limited or no fixed air quality monitoring stations, it is crucial to implement a method for placing representative fixed air quality monitoring stations in developing cities like Addis Ababa that could reflect exposure levels. The WHO Global Ambient Air Quality Database 2018 shows a stark disparity in air quality monitoring between low-income and high-income countries. Low-income countries only have one PM2.5 monitor for every 65 million people, while high-income countries have one monitor for every 370,000 people. For instance, many countries in Sub-Saharan Africa do not have permanent PM2.5 measuring stations, with only 12 African countries having air quality monitoring sites, mostly in capital cities. East Africa, one of the regions with limited air quality monitoring, is facing a growing threat from poor air quality due to rapid urbanization, increasing population, and higher fuel consumption

from the high motorization rate (Avis & Bartington, 2020). Like many cities in developing nations, Addis Ababa relies solely on a single fixed air quality monitoring site for information.[2] While this site provides valuable time-series data, it does not give information for a comprehensive understanding of air quality distribution throughout the city. This data gap also prevents analysis of differences within the city and the impact of air pollution on urban, suburban, and rural areas. The measurement site may also not accurately reflect particulate matter exposure levels for the average resident.

The second paper of the dissertation examines the extent to which the single fixed air quality monitor provides a guide to air quality and its risks throughout the city. The discrepancy of pollution information between the fixed air quality monitor is analyzed by comparing air quality data collected over six months in 2020 and 2021 across four sub-cities over a range of land use sites in Addis Ababa through a low-cost monitor. This study aims to fill the gap in understanding the accuracy of air quality information provided by a single fixed site and how it compares to low-cost monitor data collected throughout the four sub-cities of Addis Ababa. The findings suggest that relying on data from one air quality monitor positioned randomly does not provide a complete understanding of the intricate nature of air pollution. To achieve more precise measurements in the future, we propose a technique for positioning fixed air quality monitors to optimize monitor placement. This research's results will help improve decision-making and action toward reducing air pollution in Addis Ababa and other

---

[2] Currently, the only source of ground-level measurement of PM2.5 in Addis Ababa is from the US Embassy air quality monitor, which records hourly average PM2.5 installed at the parkland of the Embassy, The fixed monitor at Addis Ababa international community school (near Addis Ababa Golf club) is mostly offline and doesn't record air quality data.

cities facing similar challenges. Finally, the findings can inform the deployment of air quality monitoring networks in developing countries, providing more comprehensive and accurate information for effective air pollution control and public health protection. In general, this research brings attention to the difficulties faced in monitoring air quality in developing nations and proposes remedies to enhance the precision and availability of air quality data.

1.2. Organization of the dissertation

The dissertation is organized into four chapters. The first chapter gives an overview and this road map. The second chapter uses a cost-effective approach to predicting air quality using a land use regression model in unmonitored areas. The third chapter explores the specific context of a developing country where a single air quality monitor is the primary ground-level information source of air quality for city residents and air quality decision-making among policymakers. It evaluates the impact of this information on air quality decision-making and proposes a method for determining the optimal location of air quality for additional monitor/s considering the currently available single monitor. Finally, the last chapter gives a conclusion and policy implications.

CHAPTER TWO

2. Low-Cost Mapping of PM2.5 in Addis Ababa
2.1. Introduction

East Africa, a region with limited air quality monitoring infrastructure, faces an ever-growing danger from declining air quality due to rapid urbanization, population growth, and a rise in fuel consumption due to high motorization rates (Avis & Bartington, 2020). The scarcity of air quality monitoring systems in the region exacerbates the situation and makes it challenging to address the problem effectively. In addition, the high expense of air quality monitoring equipment that has undergone proper calibration and certification makes it challenging to measure air quality in countries with low-income countries such as Ethiopia. However, low-cost mobile sensors have made air quality monitoring simple, attainable, and easily accessible, allowing for air quality modeling and evaluations of environmental impacts. The market for portable air pollution sensors, combined with GPS technology, has seen a decline in the cost of collecting air pollution data. As a result, low-cost mobile sensors have become widely used to map air quality at a local level, as reported in various studies (Karppinen et al., 2000; Habermann et al., 2015; Saori et al., 2008; Rivera et al., 2012; Hasenfratz et al., 2015; Lim et al., 2019). Ironically, much of this research has been conducted in developed countries with established networks of high-quality fixed monitoring stations. This disparity affirms the need for increased air quality monitoring in low- and middle-income countries, where air pollution levels are often higher due to a lack of regulations and infrastructure. Using low-cost mobile sensors in these areas can help shed

light on the extent of the air pollution problem and inform efforts to address it. This will improve the quality of life for citizens and contribute to global efforts to mitigate the impacts of air pollution on human health and the environment.

The availability of low-cost mobile air quality sensors presents a unique opportunity to address the air pollution problem in low-income countries and create a more equitable and healthier world.[3] Sensors can now be affixed to rental bikes as they travel around cities or are carried by individuals on foot. This approach was demonstrated in a study by Minet et al. (2017) in Kampala, Uganda, where sensors were mounted on motorbike taxis, known as Bodhas, and powered by the bikes' batteries. The sensors reported the levels of PM2.5 and their location. Despite recent advancements in low-cost data collection methods, a network of monitoring stations alone is insufficient for providing detailed information. The lack of a link between pollutant concentrations and geographic information in these networks requires a model to create a map of concentrations and determine the areas with the poorest air quality so that policies can be implemented to mitigate the effects (Vizcaino & Lavalle, 2018).

PM2.5 is widely recognized as the most important indicator of urban air quality (Cohen et al., 2005) and has been linked to respiratory disease and elevated cardiovascular mortality due to ischemic heart disease, dysrhythmias, heart failure, and cardiac arrest. This is because PM2.5 particles are small enough to penetrate deeply into the lungs (Merbitz et al., 2012). This study aims to gather PM2.5 air quality data through low-cost mobile monitoring device

---

[3] There is no universally agreed definition of a "low-cost" sensor since anything costing less than the instrumentation cost required to demonstrate compliance with the air quality regulations can be termed low-cost. However, the cost should be as low as possible to achieve a sensor-based system for monitoring air pollution, so that widespread deployment is commercially feasible (Kumar et al. 2017).

in four sub-cities of Addis Ababa and construct an air pollution exposure map by applying a land use regression model that associates pollution levels with geographic information. The study will further extrapolate the results obtained from the sample area (the four sub-cities) to produce a comprehensive air quality map for the entire city of Addis Ababa, the capital of Ethiopia. This information will provide valuable insights into the air pollution situation in Addis Ababa and highlight areas that are most affected by air pollution. This, in turn, will help inform decision-making regarding air quality management and mitigation strategies and contribute to improving public health in the city. Using low-cost mobile monitoring devices will also make air quality data collection more accessible and cost-effective, allowing for a more comprehensive understanding of air pollution across the city.

The estimation results and hence the prediction shows that, local land use, traffic, and weather conditions have an impact on PM2.5 levels. The study revealed that commercial areas and areas with high road traffic had increased concentrations of PM2.5, indicating the significance of land use classification in forecasting particulate matter levels. Moreover, the study demonstrated that meteorological factors had a greater impact on PM2.5 concentration levels than other factors like land use classification.

2.2. Survey of Related Literature

LUR models have been widely used in air pollution studies in Europe and North America for more than a decade. However, the technique has recently been adopted in developing countries (Chalermpong et al., 2021; Lim et al., 2019) and used in Sub-Saharan Africa on rare occasions (Coker et al., 2021). A growing number of studies have been conducted in developing countries, where the PM2.5 problem has become more severe as economies and motorization rates have accelerated. Due to a lack of or inaccessibility of fixed monitoring stations in various developing countries, the research relied on purpose-designed measurement sites to collect PM2.5 data for LUR modeling. For example, a Spatiotemporal air quality prediction in Kampala, Uganda's capital, was developed using a network of locally developed low-cost sensors (Coker et al., 2021). The study used monthly average PM2.5 data from twenty-three sites collected with a low-cost sensor throughout 2020. Monthly precipitation, percentage of the population using solid fuel for cooking, distance to Lake Victoria, and green space within 500m of the air quality monitors were the most important spatial and temporal predictors of monthly PM2.5 levels. A survey of related mobile air quality monitoring is summarized in Table 2.1

Table 2.1 Review of Mobile air quality monitoring

Study	Location	Method of Data collection, estimation, and sampling	Major results
Lim et al. (2019)	Seoul, South Korea	The data collection was conducted using low-cost air quality monitors (Airbeams). The sampling campaign was implemented by repeatedly **walking** across five routes during a three-week period (July 23 to August 11) in the summer of 2015. **Pollutant of concern: PM2.5**	The constructed LUR models were used to create prediction maps of street-level PM2.5 concentration levels in Seoul that can be used for personal-level exposure in urban areas where there are more people on the streets than in cars.
Hasenfratz et al. (2015)	Zurich, Switzerland	Mobile sensors were i**nstalled at the top of public transport** vehicles for a two-year period to collect air pollution data. **Pollutant of concern: Ultrafine particles**	LUR was used to create pollution maps with a spatial resolution of 100m × 100m. In this study, pollution maps with different temporal resolutions, such as seasonal resolutions (winter-spring, summer, and fall), were created.

Rivera et al (2012)	Girona province, Spain	Mobile sensor devices were **installed on the sidewalk** of 644 participants in twelve towns of Girona province at 25 sites in Spain. Measurements were conducted during non-rush traffic hours, 9:15-12:45 and 15:15-16:45, between June 15 and July 31, 2009. Measurements were conducted using P-Track counters located perpendicularly to the street above 1.5 meters for 15 minutes in a site at a time. **Pollutant of concern: Ultrafine particles**	The LUR models captured between 36-42% of the spatial variability of UFP in the study area.
Masiol et al. (2018)	Monroe city, New York	Hourly PM2.5 concentrations were measured by placing low-cost monitors **outdoors (backyards) of volunteers' homes** during the heating seasons of 2015/16 (December to March) and 2016/17 (November to March) at 23 and 25 residential locations for those seasons, respectively.	LUR models for each hour of the day and weekdays/weekend days were developed. Model outputs were illustrated by mapping PM2.5 concentrations for midnight and noon on a typical weekday and typical weekend day since these days were found to be times when weather parameters approximated the whole

		Pollutant of concern: PM2.5	season mean values. Maps were computed on a 250×250 m grid resolution.
Jin et al. (2019)	Gansu Province, Lanzhou city, China	Data collection was conducted by **placing palm tubes 2.5 meters above the ground at light poles in Infront of small businesses (such as convenience stores or clothing shops)** for a two-week period in the summer of 2015; the sampling strategy was designed in such a way for security reasons under the surveillance of owners. Stratified random sampling and purposeful selection were used to select sampling sites. **Pollutant of concern: No2**	A predicted map of air pollution for a two-week period is generated and explains 71% of the variance in the measured No2 concentration.
Pope et al., (2018)	Kenya	Three calibrated low-cost optical particle counters were deployed for two months (February-March 2017); two monitors were deployed in the capital city	It demonstrates low-cost sensors could be used to create an affordable network to monitor air quality in cities of developing countries. The results show that, during the research period, PM

| | | Nairobi and one at the outskirts of Nanyunki, upwind of Nairobi.

Urban and background sites were chosen as a criterion for selecting the sites, although no intra-urban stratification was conducted.

Pollutant of concern:PM1, PM2.5, PM10 | mass concentration peaks during morning rush hours and evening in Nairobi rush hours. |

2.2.1. Air pollution and institutional arrangements in Ethiopia

The Ethiopian Environmental Protection Authority (EEPA) and the Ministry of Environment, Forest, and Climate Change guide air quality management in Ethiopia. The EEPA implements and enforces environmental laws and regulations, including air quality. The Ministry of Environment, Forest and Climate Change develops and implements policies and programs for protecting the environment, including air quality (World Bank, 2022). The EEPA has developed National Ambient Air Quality Standards (NAAQS), which set limits for several pollutants, including particulate matter (PM10 and PM2.5). The 2003 ambient environment standards for Ethiopia set a 24-hour PM2.5 concentration not to exceed 65 μm/m^3 and annual concentration not to excceed 15 μm/m^3 (Cao et al., 2016; EEPA, 2003).

The Ethiopian government has implemented several measures to address air pollution in Addis Ababa and other urban areas, including promoting clean technologies and renewable energy sources and improving solid waste management. In addition, Ethiopia has enacted a comprehensive general environmental policy and legal framework, including pollution control. Ethiopia's Constitution (1995), environmental policy, conservation strategy, and various proclamations, regulations, and guidelines all contain provisions that demonstrate the country's legislative commitment to environmental, social, and sustainable development goals. To carry out this commitment, Ethiopia has developed and adopted a comprehensive set of environmental laws and regulations covering environmental and social management and pollution control (World Bank,2022).

Although Ethiopia has put various regulations and policies to improve air quality, enforcement of these regulations is often weak, and the EEPA lacks the resources and capacity to monitor and enforce compliance effectively.

Addis Ababa recently prepared an air quality management including concrete steps to reduce emissions, baseline air quality characterization and projected emission trends, and a detailed air quality management implementation plan (AA EP, 2021).

2.2.2. Air quality monitoring in Addis Ababa

There are two permanent PM2.5 air quality monitors located in Addis Ababa, one BAM-1020 located in the parkland setting of the US Embassy in Gulele sub-city, which is managed by the US State Department's monitoring program, and the other at an international school in Lideta sub-city (which is primarily offline and does not record data). However, despite efforts to monitor air quality in Addis Ababa, as outlined in Table 2.2, none have led to the creation of an air quality map for the city.

Table 2.2: Air quality monitoring in Addis Ababa

Authors	Number of sites	Number of sites	Site selection criteria	Major finding
Kumie et al.(2021)	1 April 2017 – 31 March 2020 (Three years duration)	One fixed site	Availability of sustained power, height from the ground, instrument safety, and lack of any physical barrier that may restrict the free flow of air was considered as site selection criterion in this study.	The study described levels of PM2.5 in Addis Ababa during the study period and examined temporal patterns, also examined the health impacts of PM2.5. The study illustrated two daily extremes of PM2.5, morning (high) and afternoon (low). Sunday has the lowest concentration, while Friday has the highest concentration of PM2.5; Mondays to Thursdays show a continuous increase in PM2.5 concentration.
Watson (2005)	26 January-28 February 2004 (One Month)	Twelve sites	Sites were selected to represent the urban core intended to represent ambient conditions at street level (which were located between 50 and 100m from arterial roads with heavy motor-vehicle and pedestrian traffic),	The study explored the temporal variation of PM10 and CO; the study found that PM10 and CO exhibited daily maxima around 7:00 AM and secondary peaks in late afternoon and evening, suggesting that those pollutants were emitted during periods associated with motor-vehicle traffic.

			other sites in the city were selected to represent variation in space, sites along the eastern periphery of Addis Ababa, were chosen to represent air quality levels outside the immediate urban area.	
Tefera et al. (2020)	November 2015- November 2016 (61 days), samples were taken every six days.	One fixed site	The station is in a typical downtown urban area surrounded by residential houses as well as public institutions and private businesses.	The study characterized annual PM2.5 in both chemical composition and seasonal patterns. In almost 90% of sampled days, PM2.5 exceeded WHO`S guidelines during the study period. Furthermore, the study found that higher PM2.5 concentrations were found during heavy rain season. Compared to WHO`s AQI, 31% and 36% of observed days were unhealthy for everyone and unhealthy for sensitive people.

				The study also pointed out that Meteorological variables, vehicle emissions, biomass fuels, unpaved roads, and construction activity contribute to poor air quality.

2.2.3. Review of literature on calibration and use of low-cost air quality monitors

There is currently no standard calibration protocol for low-cost sensors. Different calibration methods, such as chamber and field testing against various reference instruments, have been used in studies, making the inter-comparison of low-cost sensors impossible (Kumar et al., 2017). Nonetheless, studies provide critical information about the performance of low-cost sensors under a wide range of operating conditions.

This section reviews calibration methods for low-cost air quality monitors (specifically for Particulate Matter sensors) and the method we used to select the air quality monitor. Kumar et al., (2017) examined the performance characteristics of several low-cost particle and gas monitoring sensors and discovered that, in general, sensors performed better (high R^2 values) in the laboratory than in the field. The study discovered that performance deterioration during "on-the-field" conditions is caused by changing particle compositions, sizes, and environmental factors, which impact sensor response. This implies that on-site calibration is critical for reflecting reliability, whereas laboratory calibration is insufficient for assessing the device's performance. A study by Galatoulas et al. (2018) demonstrated the intended application of collecting fine-grained Spatiotemporal PM2.5 profiles in an on-road trial by mounting the developed air pollution monitors on an electric bike as a case study in the city of Mons, Belgium, on 1-minute resolution. In that study, roadside measurements from a mobile laboratory were compared using a calibrated instrument, demonstrating its accuracy.

Several researchers such as Amegah (2018), Galatoulas et al. (2018), Jovaevi-Stojanovi et al. (2015), Manikonda et al. (2016), Santiago et al. (2013), Semple et al. (2013), Sousan et al. (2016), Steinle et al. (2015), and Wang et al. (2015) have evaluated the performance of low-cost PM sensors. However, the lack of a standardized method for assessing the performance of low-cost PM sensors makes comparing the results of different studies difficult. Nonetheless, the performance characteristics of the various sensors are similar, with R^2 values greater than 0.50 compared to reference measurements in different environmental settings (Kumar et al., 2017).

When selecting the low-cost monitor, we used Dylos 1700; we considered the device's data storage capacity, display system, and battery life.[4] Researchers have compared the performance of the Dylos 1700 to that of a reference monitor (Manikonda et al., 2016; Semple et al., 2013; Sousan et al., 2016; Kumar et al., 2017). When compared to a standard reference air quality monitor, Susan et al. (2016) demonstrated that the Dylos 1700 helps estimate aerosol mass concentration at a workplace with $R^2 > 0.97$. Manikonda et al. (2016) compare the performance of four low-cost PM sensors, including the Dylos 1700, to a well-characterized reference instrument. The study discovered that any monitors used in the calibration, including the Dylos 1700, tested performance with adequate precision for monitoring air quality. In addition, factors such as data download method, connectivity, and compatibility with a computer are considered when selecting the device, we used.

---

[4] The Dylos DC1700 is a battery-operated air quality monitor that measures particles in two size ranges: $> 0.5\ \mu m$ (Small particle number counts) and $> 2.5\ \mu m$ (large particle number counts) (http://www.dylosproducts.com/dc1700.html).

2.3. Sampling method and study area
2.3.1. Description of the study area, Addis Ababa

Addis Ababa, with an area of 521.47 km^2, is situated at the base of the "Intoto" mountain in the central region of Ethiopia. The city has a circular shape with a diameter ranging from 30-40 km and is situated between 2,200 to 2,800 meters above sea level (Kume et al., 2010). According to the 2007 census conducted by the Central Statistical Agency (CSA) of Ethiopia, the city had a total population of 2.7 million (Kume et al., 2010). However, projected population estimates show that the city's population has grown to 3.2 million as of 2013 (Central Statistical Agency, 2013). Census is still not conducted, and UN world pop data shows that Addis Ababa has an estimated 5.2 million. For our sampling purposes, we used the actual 2007 census data. [5] The population density in each sub-city is consistent with the projected data. Addis Ababa is surrounded by mountain ranges that stretch from the northeast to the western part of the city, with plains to the east and south. As a result, winds tend to carry emissions toward the western and northwestern areas of the city (Tefera et al., 2020). Inversion of temperatures is common during low-temperature seasons, particularly during early mornings and evenings (Tefera et al., 2020).

Addis Ababa is divided into ten administrative districts with unique land areas and population densities.[6] The sub-cities of Addis Ketema, Lideta, Arada, Kirkos, and Gulele have the highest population densities and rank first through fifth, while the sub-cities of Kolfe, Nefas Silk-Lafto, Yeka, Bole, and Akaki have the lowest population densities and rank last. In terms of land size, the sub-cities of Bole, Akaki, Yeka, Nefas Silk-Lafto, and Kolfe rank first, while the sub-cities of

---

[5] https://vizhub.healthdata.org/gbd-results/

[6] The number of sub-cities in Addis Ababa increased to 11 in October 2020; the new sub-city name is Lemi-Kura.

Gulele, Kirkos, Arada, Lideta, and Addis Ketema rank last (see Table 2.3 for the population density of each sub-city).

The climate of Addis Ababa exhibits three distinct seasons delineated by precipitation trends: the dry season (October to January and May), the period of light to moderate rainfall (February to April), and the primary rainy season (June to September), as documented by Tefera et al. (2020).The National Meteorology Agency (NMA) reports a monthly temperature range of 8 to 25 degrees Celsius within the city. Monthly average rainfall varies from a minimum of 7 mm in November and December to a peak of 280/290 mm in July and August. Correspondingly, the mean monthly relative humidity fluctuates between 45.5% in December and 79.5% in July, according to data from the National Meteorology Agency in 2022 and Tefera et al. (2020).

This interplay of precipitation, temperature, and humidity gives rise to a distinctive weather pattern in Addis Ababa, exerting a substantial influence on air quality. Consequently, comprehending these meteorological patterns and the underlying factors impacting air quality in the city becomes paramount. This comprehension is integral to formulating effective strategies for managing air quality and safeguarding the health and well-being of the city's inhabitants.

Table 2.3: Sub-cities in Addis Ababa with respective population density and area coverage

Sub-city Name	Population in 2007	Projected population in 2014	Land area in square km.	Population density per square km.
Akaki Kality	181,270	211,380	118.08	1,535.14
Nefas Silk-Lafto	316,283	368,883	63.30	4,996.57
Kolfe Keranio	428,895	500,163	61.25	7,002.36
Gulele	267,624	312,096	30.18	8,867.10
Lideta	201,713	235,246	9.18	2,1973.10
Kirko	221,234	258,035	14.62	15,132.28
Arada	211,501	246,680	9.91	21,342.17
Addis Ketema	255,372	297,793	7.41	34,463.15
Yeka	346,664	404,336	85.46	4,056.44
Bole	308,995	360,387	122.08	2,531.08
Total	2,739,551	3,194,999	521.47	5,253.51

Source: Central Statistical Agency of Ethiopia (CSA)

Note: The table describes the population density of Addis Ababa using the 2007 census of Ethiopia and projected data for 2014 from CSA.
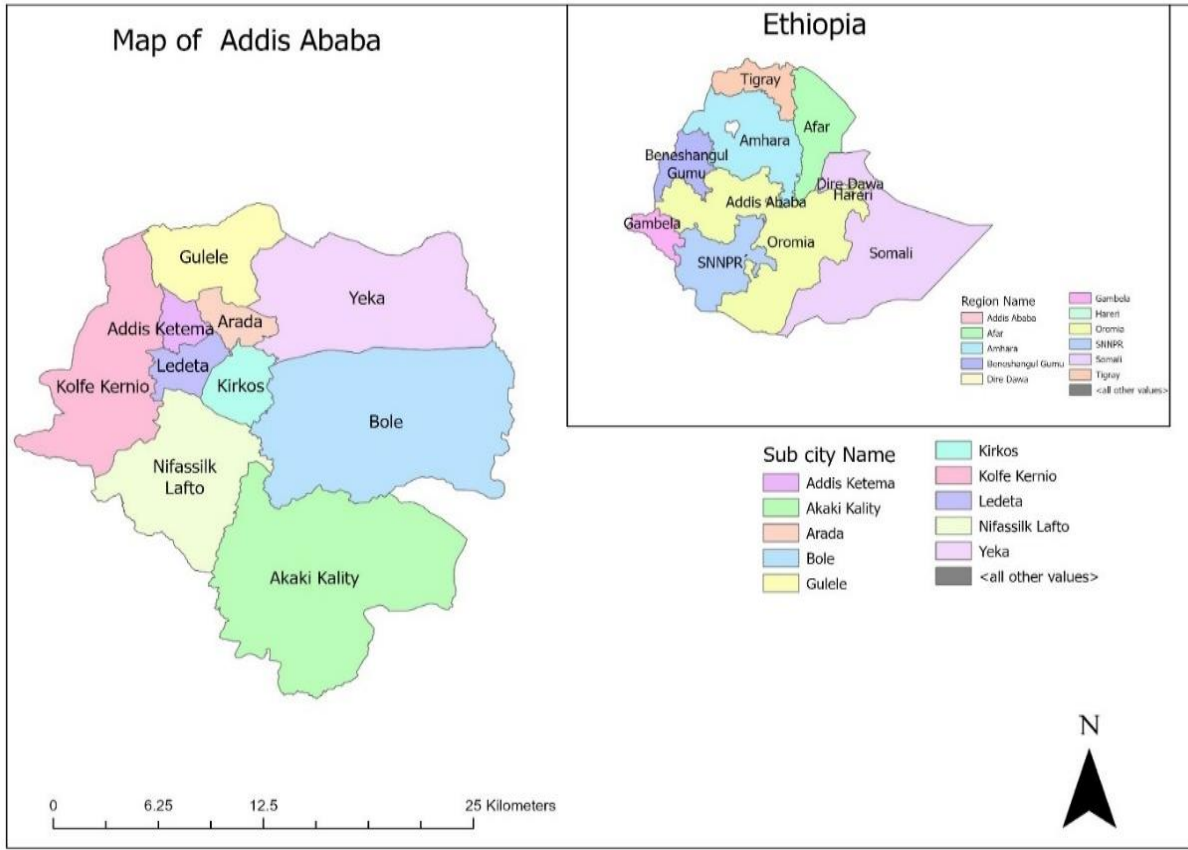
Note: The figure shows the location of Addis Ababa in the country and the ten sub-cities of Addis Ababa

Source: Author's illustration using ArcGIS Pro

Figure 2.1 Map of the study area Addis Ababa

2.3.2. Selecting a sample sub-city from Addis Ababa

The spatial variability of urban air quality is complex due to the variety of land use types and the complexity of the urban environment (Kerckho et al., 2017). To better understand air quality and its connection to the urban environment in developing countries, where air quality monitoring networks often lack careful monitoring network design, it is crucial to create high-resolution pollution maps and identify the local sources of air pollution. However, traditional air quality monitoring instruments can be costly to purchase and maintain, making it challenging to install a dense network of monitors to capture local-scale variability in pollution concentrations (Munir et

al., 2019). There is also a lack of clarity in the literature regarding the definition and regulations of determining the spatial representativeness of an air quality monitoring station, and no established method for determining the appropriate number of monitoring locations needed to meet a specific study goal and setting (Hoek et al., 2008).

To effectively address the negative impacts of pollution on both human health and the environment, it is crucial to have a well-designed air quality monitoring system in place. This will allow for an accurate assessment of air quality and support decision-makers in developing policies to lower air pollution levels (Lozano et al., 2009). In the past, sample stations were typically placed in areas with high pollution levels. While this strategy provides information on the potential exposure to pollutants and the extent of risk, it does not provide insight into the effects of future pollution control measures or the level of exposure in areas outside the immediate vicinity (Nejadkoorki et al., 2011).

It may be beneficial to increase the frequency of pollution monitoring in areas with high population density (Kanaroglou et al., 2005). A study by Pope et al. (2018) conducted in Nairobi, Kenya, found that the dense population and heavy local traffic in urban areas contribute to the city's air pollution. Munir et al. (2019) proposed using Population-weighted Pollution Concentration (considering both population density and pollution levels) and Weighted Spatial Variability (which accounts for the spatial gradients of air pollutant concentration) as crucial factors in determining the placement of air quality monitors. The authors suggest that while the overall number of monitors is determined by budget constraints (Economic indicator), the distribution of monitors should be proportional to the Weighted Spatial Variability (WSV) in each area to account for population density (social indicator) and spatial variability of air pollutant concentrations (environmental indicator).

Human exposure to air pollution is directly tied to population density (residents per square kilometer) (Munir et al., 2019). Among the ten sub-cities of Addis Ababa, the inner core of four sub-cities (Ketema, Lideta, Arada, and Kirkos) have a significantly higher population density than the surrounding semi-rural areas. Given the large size of the city and the contrasting density between the center and outer suburbs, the focus is on measuring air quality in the four central sub-cities with the highest population density: Addis Ketema, Arada, Kirkos, and Lideta (as shown in Figure 2.2). The combined area of these four sub-cities is 41.12 square kilometers, with the highest population density in Addis Ketema (34,463 people per square kilometer) and the lowest in the Kirkos sub-city (15,132 people per square kilometer). The other two sub-cities, Lideta and Arada, have population densities of 21,973 and 21,342 people per square kilometer, respectively.
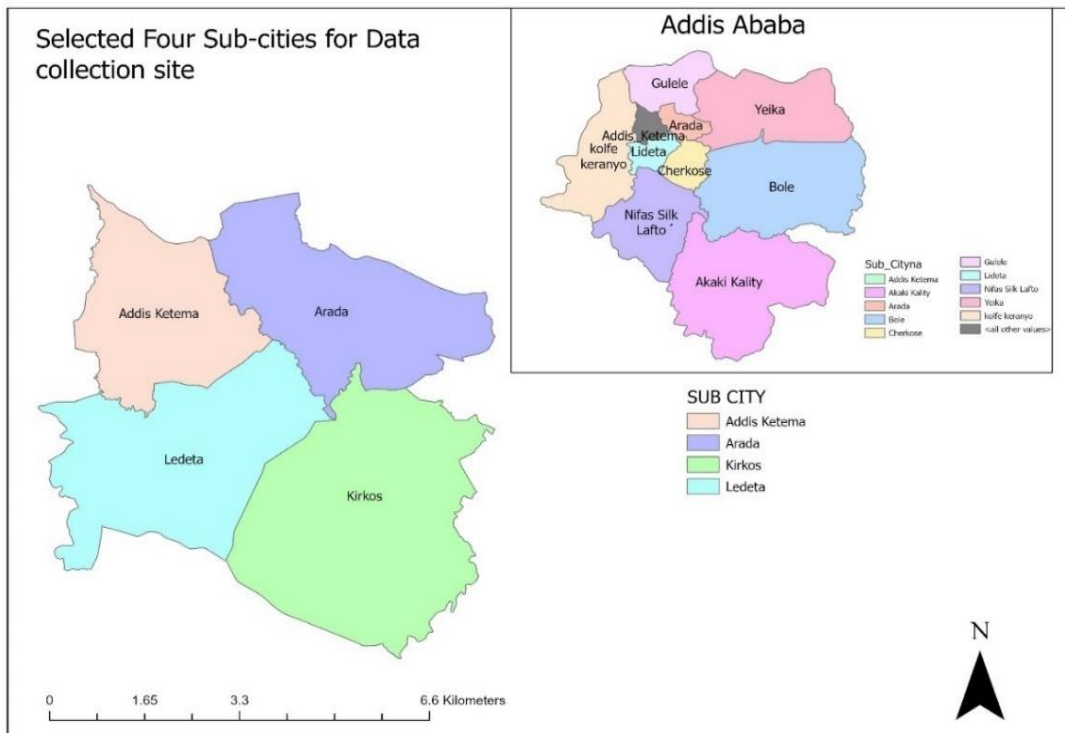
Note: The four sub-cities at the left bottom of the figure are the four central sub-cities of Addis Ababa which are the main study area of this study where air pollution sampling data were taken
Source: Author's illustration using ArcGIS Pro

Figure 2.2: Location of selected sample sub-cities based on population density.

2.3.3. Selecting a sample site from a grid and the rationale for using a grid as a selection mechanism

There is no set method for determining the number of sites required for a land use regression (LUR) analysis. Studies such as Hoek et al. (2008) utilize between 20-100 locations for studies in individual cities, often using a combination of a systematic and random sampling of land use (Munir et al., 2019). Systematic sampling, also known as grid or regular sampling, involves collecting samples at specific locations or over time in a defined pattern, which is used to select a cluster from a sub-city. This method is preferred for its uniform coverage, ease of use, and ability to ensure that the key features of the population are not missed. Additionally, samples taken at regular intervals, such as every node of an area defined by a grid, help estimate spatial or temporal correlations or identify patterns (EPA & Quality Management Division., 2002). To select a cluster from a sub-city sample, the four sub-cities were divided into a grid of 1 km by 1 km. The entire sampling area was divided into 1 km by 1 km squares in order to capture the variability of pollution exposure in the specified exposure area (1 km by 1 km) see Figure 2.3.
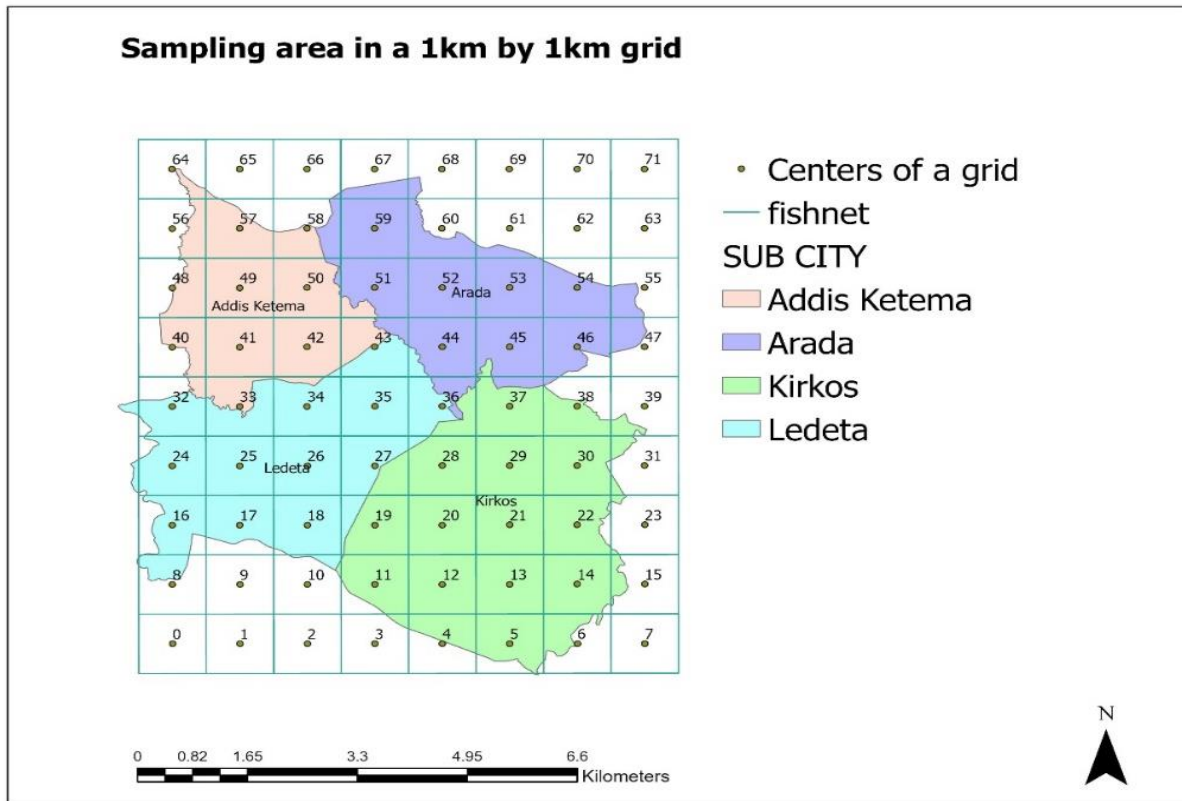
Note: The dots are centers of each grid of 1km square; we used ArcGIS pro software to create a fishnet that covers the four main sub-cities focused for this study

Source: Author's illustration using ArcGIS Pro

Figure 2.3: Sample sub-cities in a grid of 1 km by 1km

The entire study area, which encompasses the four sub-cities and the surrounding area, spans seventy-two km^2. The four sub-cities occupy 41.12 km^2, while the remaining 30.88 km^2 belongs to the surrounding area. The surrounding area was included in the main study area because air pollution in the main study area can be influenced by the surrounding environment since small particles, which are the focus of this study, can travel long distances and therefore impact the concentrations in the main study area. Considering this and the time and money cost of the study to collect the data, we divided the study area into a 1km-by-1km grid for sampling.

We selected thirty-six clusters (1km by 1km square grid areas) from the 72 squares using systematic grid sampling, taking a cluster every 2km. When we look at the distribution of selected

23

grids by sub-city, we have 4, 5,7, 8, and 12 grids of 1 km by 1km for Addis Ketema, Arada, Lideta, Kirkos sub-city and the surrounding area that covers the main study area, respectively (Figure 2.4 for the distribution of selected grids(clusters) between the four sub-cities and the surrounding area).

Studies use different methods to maximize the contrast in variables hypothesized to be potentially significant predictors. Sampling sites were selected to represent different land use areas in the city by considering the distribution of locations to which the model will be applied. In each selected squared grid, we list three different land use types which we can find in common from each squared grid to reflect the spatial distribution of different land use patterns in the study area by using google map and ArcGIS identifier tool (two mixed land use, two asphalt road, two open areas). The definition of these three land use types depends on the area's economic activity and the land use's primary purpose. For example, asphalt roads in this classification are defined as main roads (including one-way and two-way asphalt roads) and bus stations. Mixed area land use includes where people use the area as a commercial and residential area. Open areas include parks, open areas inside churches, stadiums, playgrounds for youths in the city, cemetery areas, and city farms.

The sampling frame contains eight mixed land use sites (4 grids of 2 mixed lands), eight asphalt road sites (4 grids of 2 asphalt roads), eight open area sites (4 grids of 2 open areas) for Addis Ketema sub-city, ten mixed land use sites (5grids of 2 mixed land), ten asphalt road sites (5 grids of 2 asphalt roads), ten open area sites (5 grids of open area) for Arada sub-city; 14 mixed area sites (7 grids of 2 mixed land uses), 14 asphalt road sites (7 grids of 2 asphalt roads), 14 open area sites (7 grids of 2 open areas) for Lideta sub-city; 16 mixed land use sites (8 grids of 2 mixed land uses), 16 asphalt road sites (8 grids of 2 asphalt roads), 16 open area sites (8 grids of 2 open areas) for Kirkos sub-city and 24 mixed land use (12 grids of 2 mixed land use), 24 asphalt road sites (12 grids of 2 asphalt roads), 24 open area sites (12 grids of 2 open areas) sites for the surrounding.

In this way, we end up with 216 potential locations for gathering data (36 grids of 6 sites) (see Table 2.4).

TABLE 2.4: Sampling Frame

	Mixed Land (ML)	Asphalt road (AR)	Green area (GA)	Total	No of Cluster (1kmX1km grid)
Addis Ketema	8	8	8	24	4
Arada	10	10	10	30	5
Lideta	14	14	14	42	7
Kirkos	16	16	16	48	8
Surrounding sub-cities	24	24	24	72	12
Total	72	72	72	216	36

Note: The table illustrates the sampling frame by land use and by grid(cluster)
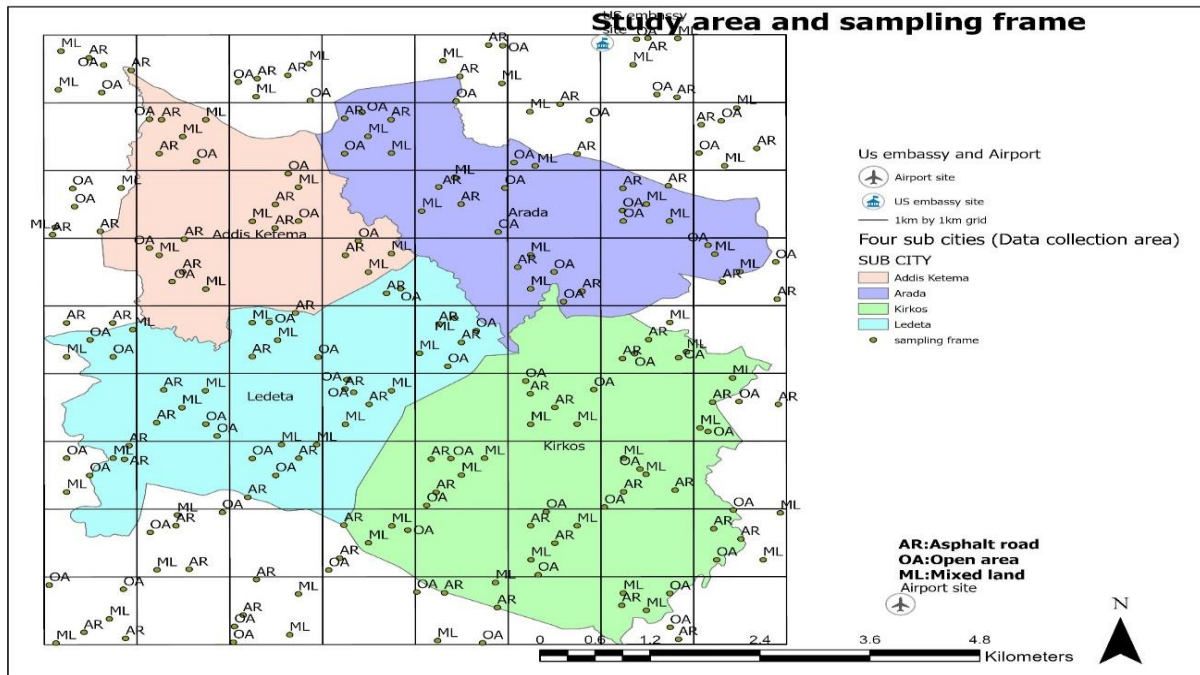
Note: The six sites at each selected grid are sampling frames where we select two sites at random for monitoring air quality in Addis Ababa

Source: Author's illustration using ArcGIS Pro

FIGURE 2.4: Sampling frame

We then use a random number generator to pick two sites out of the six in each grid. Since we do not know the pollution levels for different land use types in the study area, we give each site an equal chance of being selected (i.e., each land use site in a grid has a probability of 1/3 being selected). This gives us seventy-two sites in total.[7] In addition, we selected a one-site adjacent to the US Embassy for calibration purposes. We will demonstrate detailed comparisons and analyses of our Dylos readings and the US Embassy data in Chapter 3, Section 3.2.

The random draw from the total population gives two mixed land use, four asphalt road, and two open area sites for Addis Ketema sub-city; 4 mixed land use, three asphalt road, and three open area sites for Arada sub-city; 5 mixed land use, four asphalt road and five open area sites for Lideta sub-city; 5 mixed land use, six asphalt road and five open area sites for Kirkos sub-city; 4 mixed land use, seven asphalt road and 13 open area sites for the surrounding area which makes the total number of selected sites to be 20 mixed land use (ML), 24 asphalt road (AR) and 28 green areas (GA) sites Table 2.5 and Figure 2.5.

TABLE 2.5: Selected sample sites by a random draw from each cluster

Sub-city	Mixed Land (ML)	Asphalt road (AR)	Green area (GA)	Total
Addis Ketema	2	4	2	8
Arada	4	3	3	10
Lideta	5	4	5	14
Kirkos	5	6	5	16
Surrounding sub-cities	4	7	13	24
Total	20	24	28	72

---

[7] Though it is in the typical range for site selection, seventy-two is an arbitrary number, but it was chosen after we piloted the data collection and found that the maximum number of sites that could be visited in the battery life of the device was 10-12.
Reconnaissance revealed that a small number of sites were not accessible (e.g., embassies or defense force properties), and in these cases, we substituted the proposed location with the nearest feasible site of the same kind.

Note: Sampling sites selected using a random number generator in Excel are shown by land use type
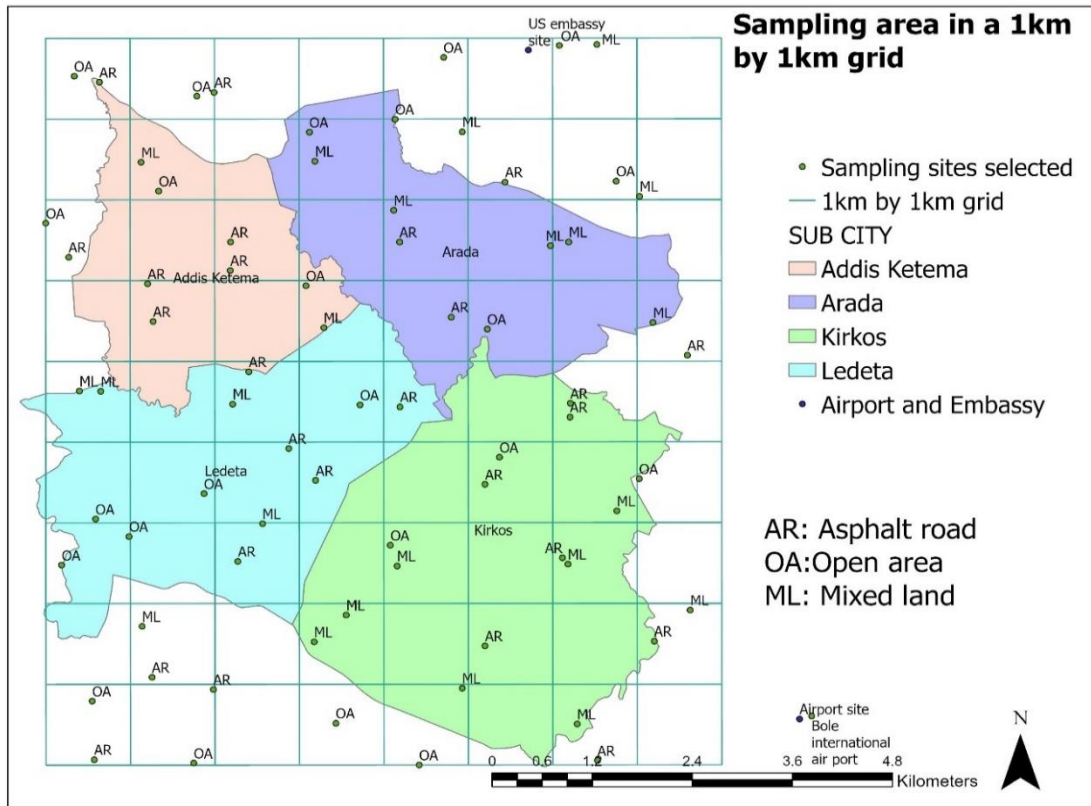
Note: The two sites indicated in the selected grid are randomly selected sites from a total of six sites in each selected grid

Source: Author's illustration using ArcGIS Pro

Figure 2.5: Selected sample sites by land use type

2.3.4. How many sample sites and how long a sample site should be monitored every time?

The selection of air quality monitoring stations is a crucial aspect of any air quality monitoring program, as emphasized by Munir et al. (2019). In this study, the number of sample sites was chosen with the aim of achieving uniform coverage of the study area through systematic grid sampling and by considering the limited data collection time, considering the battery life of the device used. To achieve this, thirty-six clusters were selected, spaced every 2 km. The sample sites

were visited based on the available time for this research and the estimated travel time in Addis Ababa.

A preliminary investigation of travel time in Addis Ababa during August and September 2019 indicated that during peak hours, typically in the morning from 6:30 to 11 am and in the afternoon from 4:30 pm to 9 pm, it took an average of 5 minutes to travel 1 km. During off-peak hours (11 am-4:30 pm and after 9 pm), it took an average of 3 minutes per km. The average of these two gives 4 minutes per km. It must be noted that these travel times are only estimates and may vary significantly in specific areas of Addis Ababa. To minimize the impact of peak hour traffic, data collection during these times was considered to travel against the traffic flow. During the morning, traffic typically moves toward the city center, while in the afternoon and evening, it flows toward residential areas on the city's outskirts (see Table 2.6).

Table 2.6: Travel time by car

Date	Travel time in minutes	Distance traveled in Km	Minutes/km
10-Aug	26	4	6.5
8-Aug	70	16	4.37
8-Sep	26	13	2
14-Sep	41	12	3.41
14-Sep	67	23	2.91
15-Sep	24	11	2.18
16-Sep	184	35	5.25
17-Sep	109	26	4.19
22-Sep	19	5	3.8
27-Sep	25	6	4.16
27-Sep	193	37	5.21
28-Sep	42	6	7
29-Sep	72	15	4.8
Total	898	209	4.29

Note: Pilot travel time in minutes and distance in kilo meters by car in Addis Ababa during a pilot study in August and September 2019.

We also conducted a pilot study in August and September 2019 to determine the feasibility of individual sites and travel plans. With the sensor saving a reading every minute, the pilot demonstrated that after 10 minutes of recording, the sample variance becomes small, implying that waiting for a long time in a site does not add value. Instead, moving to different sites may add value to contrast the air quality due to land use and weather differences. Based on this, the device's battery life, and the average travel times between locations at various times, we determined that 9-10 clusters per day were feasible (See Table 2.6 for the calculated variance trends). As a result, we set the reading time in each sample site to 15 minutes. Thus, we determined the sampling sites as 73 because we chose 36 clusters and one calibration/validation site for a fixed site near the US Embassy.
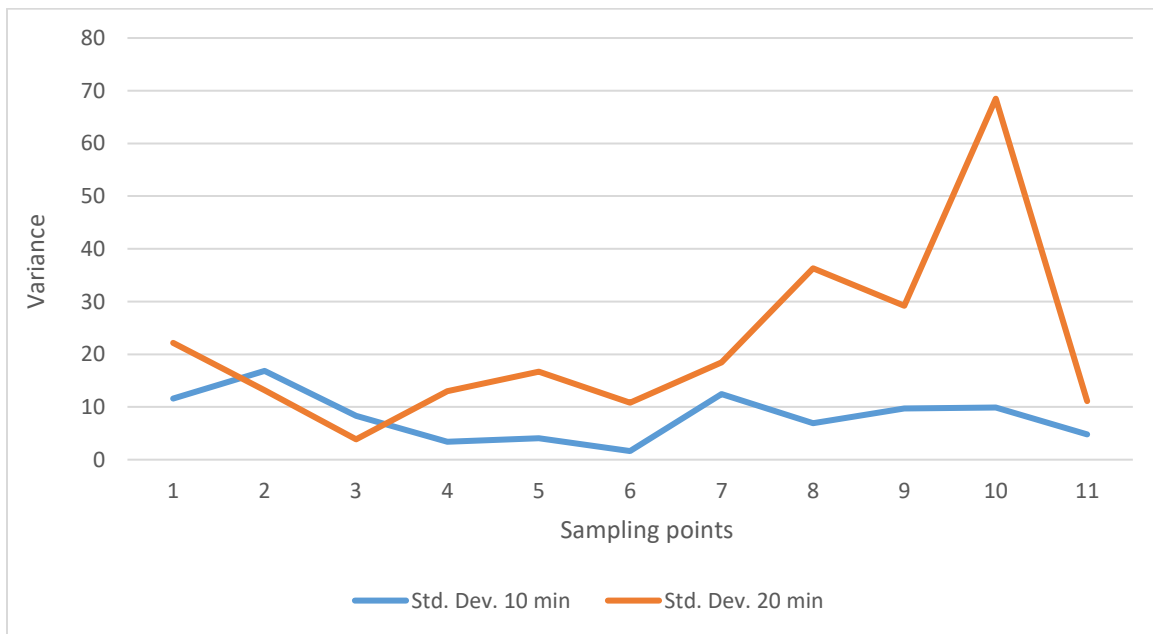
Note: Pilot variance difference, the pilot was conducted in August and September 2019 before the actual data collection

Figure 2.6: Variance calculation of monitoring duration

We divided the number of clusters to be visited over four days based on the portable device's battery life and travel time. Hence, we visited nine clusters and the US Embassy site each day. The average travel time (within and between clusters) for nine clusters is almost 2 hours, excluding the travel time to the first cluster. Recording takes 4 hours and 45 minutes for nine clusters (18 sites) and one calibration site. Moving to the vehicle and packing time and some exceptional delays in a cluster or between clusters took 15 minutes. Hence the total working time each day was 7 hours, excluding travel to the US Embassy site and the first site in the morning. Each day's time breakdown on how many sample sites were visited in the morning, mid-day, and afternoon times are detailed in Table 2.7.

TABLE 2.7: Activities and time breakdown each day.

	Cluster(sites)	Travel time within the cluster	Travel time between cluster	Moving to the vehicle and packing time	Monitoring/reading time	Total working hours in a day
Day 1- Day 4	9(18) Plus a visit to the US Embassy site	45(5 minutes for each cluster)	1 hour and 12 minutes (average 8 minutes between consecutive clusters)	15 minutes (as precautionary time for delay of a traffic jam)	4 hours and 45 minutes	Average 7 hours plus travel to US fixed site since travel time to US site depends on route for each day excluding travel to the first site.

2.3.5. Field visits and adjustments in site location

During the data collection period, a visit to the planned sites was made in the first week (February 10 -February 18, 2020) to ensure the feasibility of collecting air quality data on those sites that were planned to be visited from ArcGIS pro and Google maps. The actual site visit, however, reveals that a few site locations must be adjusted because some locations are inaccessible; thus, the nearest primary land use site is visited. As a result, the number of open areas, asphalt roads, and mixed land use sites has been reduced to 23, 26, and 23, respectively, from the previously planned 28, 24, and 20, implying that the number of sites from each land use did not change significantly, with only five open area sites reduced from the planned sites and changed to three mixed lands and two asphalt road sites. The change was because it was impossible to enter those open areas, some of which were embassies and some of which were prohibited areas, such as defense force areas, as previously explained.

2.3.6. Route identification

Due to our substantial number of sites and the limitation to only one measurement device, it is not feasible to cover all sites within one day during the pilot traveling time. Consequently, we followed different routes to capture the pollution variability when we collected the pollution data at various times. We diversify the time, and the land use sites during measurement days. We followed different routes to travel from one cluster to another (see Figure 2.7). We also reverse the route's order to maximize the variability of visiting a site. The idea behind each route is to travel in adjacent clusters (routes 1 and 2), travel across the study area (route 3), and travel locally (route 4) so that each time when we take a sample of air quality, the time we arrive at that site will differ (e.g., site A is visited in the morning following route 1, and it is visited in the afternoon or during

mid-day following route 2) and the land use we visit will vary. There was a possibility of visiting each site in eight ways (4 routes traveling back and forth) during the first-round data collection period. However, during the actual data collection, we traveled all four routes once, and routes one and two were repeated in the opposite direction; due to time constraints, it was impossible to repeat routes three and four during the first-round data collection.

The number of sites we visited each day was reduced to twelve in the second round of data collection (from 18 in the first round) due to the inability of an experienced driver to collect data in a day for an extended period. The routes were designed to follow a similar pattern to the first-round routes we took. The first route we took during the second round of data collection is like route four in the first round of data collection in that the idea behind both routes is to travel locally, which means we visited nearby sites on each route. The second route we took for the second round of data collection is like route two for the first round of data collection in that it is intended to visit sites from north to south of Addis Ababa, and each site was visited twice during the second round of data collection.

A) First-round data collection routes
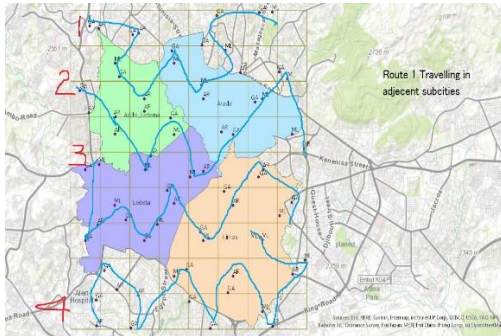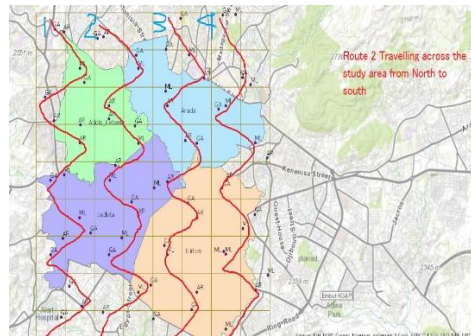
Route 1

Route 2

Route 3
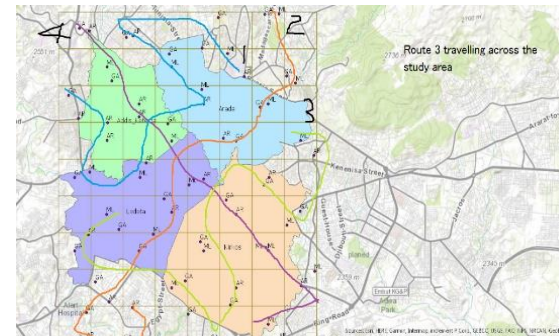
Route 4
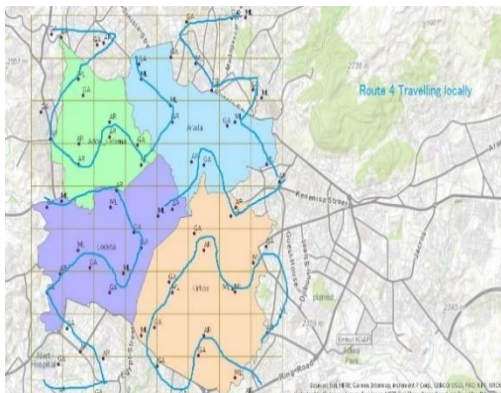
B) Second-round data collection routes

Route 1   Route 2

Note: The first four routes are routes followed in the first round of data collection, and the last two routes are routes of the second round

Figure 2.7: Routes followed for data collection

2.3.7. Method of data collection

A car is used to travel from one site to another, following a predetermined route. A hired driver is responsible for driving, while the researcher is responsible for collecting data. After arriving at each site and turning off the engine, the device is taken away from the vehicle and held in shadow approximately 1.5 meters from the ground. The device automatically starts measuring. In some cases, we put the device on the top of the car since the engine is turned off and no dust is coming out from the car (we checked this from the screen on the device, and there is no interruption on the readings when we are near the car and away from the car). Starting and finishing times of data were recorded in a separate paper so that track records and time-stamped data from the device would be matched later.[8] After recording the data, the device is turned off, and the researcher and the driver proceed to the following location (see Appendix 1).

2.4. Data
2.4.1. Description of collected PM2.5 data in Addis Ababa

Air quality information was gathered in Addis Ababa, encompassing both the dry and rainy seasons. Specifically, data collection took place in February, March, October, and November of 2020 during the dry season, as well as in August and September of 2021 during the rainy season. In the initial data collection phase, a total of 24 days' worth of air quality data were acquired, spanning from February 18, 2020, to March 19, 2020, for all 72 designated sites. Additionally, 12 days' worth of data were collected between October 17 and November 1, 2020, focusing on the

---

[8] Each day in the evening, after downloading the data, the last 15 minutes readings for each site (the initial 1 to 2 and sometimes 5 minutes of data are not taken to allow the device to warm up) were added to the excel sheet of the PM2.5 data file for final analysis.

dry season. For the rainy season, data were gathered from three selected sites out of the 72, with one of these sites located at the airport.[9] As detailed in the sampling method, seventy-two sites were selected to represent the primary land use types in urban environments. The three sites selected during rainy season data collection (asphalt road, open area, and mixed land use) were selected based on convenience sampling as the research assistant collected the data due to the limitation of traveling back to Ethiopia for several reasons (time limitation, and the ongoing security situation in Ethiopia). The rainy season data were collected between August 21, 2021, and September 25, 2021. The first and second-round data was collected on twenty-four weekdays, and the rest twelve days of the data was collected during weekends. Each day eighteen sites were visited during the first-round data collection and twelve sites during the second-round data collection; during the rainy season data collection, all three sites and the airport site were visited. In each site, 15-minute air quality data were collected during the first and second rounds and 20 minutes during the rainy season data collection. The data collection duration during the rainy season is extended by 5 minutes to increase the number of observations and to check the variability of air quality in a site during the rainy season. Details of rainy season data collection and summary are in Appendix 1.

In total, 67.11% of this data is collected during weekdays, and 32.89 % of this data is collected on weekends see Table 2.8 A. Although morning time shows more polluted air than afternoon, a more detailed day classification shows that evening is the most polluted time, followed by early morning, late morning, late afternoon, and early afternoon. September shows the highest mean PM2.5 level, followed by August, March, February, October, and November. Looking at the average PM2.5 by

---

[9] The air pollution data at the airport site (Bole International airport) is collected to check the air quality situation at the airport as airports are special type of land uses and need verification to check the validity of maps predicted based on first and second round data collection.

land use shows that the highest average PM2.5 is observed at the Asphalt Road site (47.54), followed by open area (38.86) and Mixed land (38.02). Saturday shows the highest PM2.5 levels, followed by Monday, Thursday, Friday, Wednesday, and Tuesday, and the lowest pollution is on Sunday. The mean of PM2.5 collected was 41.68, with a minimum of 0 concentration observed on asphalt road evening time during the rainy season and a maximum of 410.98 observed in the open area early in the morning on a weekday during the dry season in February see Table 2.8, Figure 2.8 and Figure 2.9 for details of PM2.5 profile and see Appendix 1:Table 2.A.1 for details of PM2.5 by the site.

The highest variance in PM2.5 is observed at route 4; the highest PM2.5 level is also observed in this route. This route was designed to capture variability in PM2.5 in short distances; the range indicates that higher variance is observed locally within short distances. The lowest PM2.5 and variance are observed in route one, where we travel to adjacent sub-cities, which cover a longer distance than route four. Routes three and four indicate the same variance in PM2.5 (Appendix 1 Table 2.A.2), although routes one and two have the same design (traveling in adjacent sub-cities), while route three is designed to capture traveling across the study area with the longest distance.

TABLE 2.8: Summary statistics of PM2.5 data collected in Addis Ababa

A) By season

Data round	Weekend	weekday	Total
Rainy season	678	1,499	2,177
Round one	2,160	4,314	6,474
Round two	717	1,440	2,157

E) By land use

Land use	Mean	Min	Max	SD
AR	47.54	0	333	36.38
ML	38.02	5.32	262	30.85
OA	38.86	6.61	411	36.66
Total	41.68	0	411	35.07

B) Time of day (2 classifications)

Time	Mean
Afternoon	36.16
Morning	49.62
Total	41.68

F) By the day of a week

Days in a Week	Mean	Percent
Monday	46.14	11.92
Tuesday	37.56	13.57
Wednesday	38.58	16.07
Thursday	42.21	11.09
Friday	41.9	14.45
Saturday	53.12	16.77
Sunday	32.5	16.13

C) By the time of day(5 classifications)

Time	Mean PM2.5	Freq.	Percent
Early morning	77.1	1,562	14.45
Late morning	34.69	2,872	26.57
Early afternoon	25.92	3,367	31.15
Late afternoon	29.34	1,974	18.26
Evening	82.55	1,033	9.56

D) By Month

Month	Mean
February	31.83
March	42.32
August	50.1
September	77.06
October	24.52
November	18.43
Total	41.68

G) By data collection round in minutes

Data round	Rainy season	Round one	Round two	Total
Monday	298	810	180	1,288
Tuesday	299	808	360	1,467
Wednesday	298	1,079	360	1,737
Thursday	299	540	360	1,199
Friday	305	1,077	180	1,562
Saturday	375	1,080	357	1,812
Saturday	303	1,080	360	1,743

Note: Table A and B summarizes PM2.5 by data collection round and land use, Table B and C summarize by the time of day while Table D summarizes by month, Table F and G summarizes by day of a week

A)

B)

C)
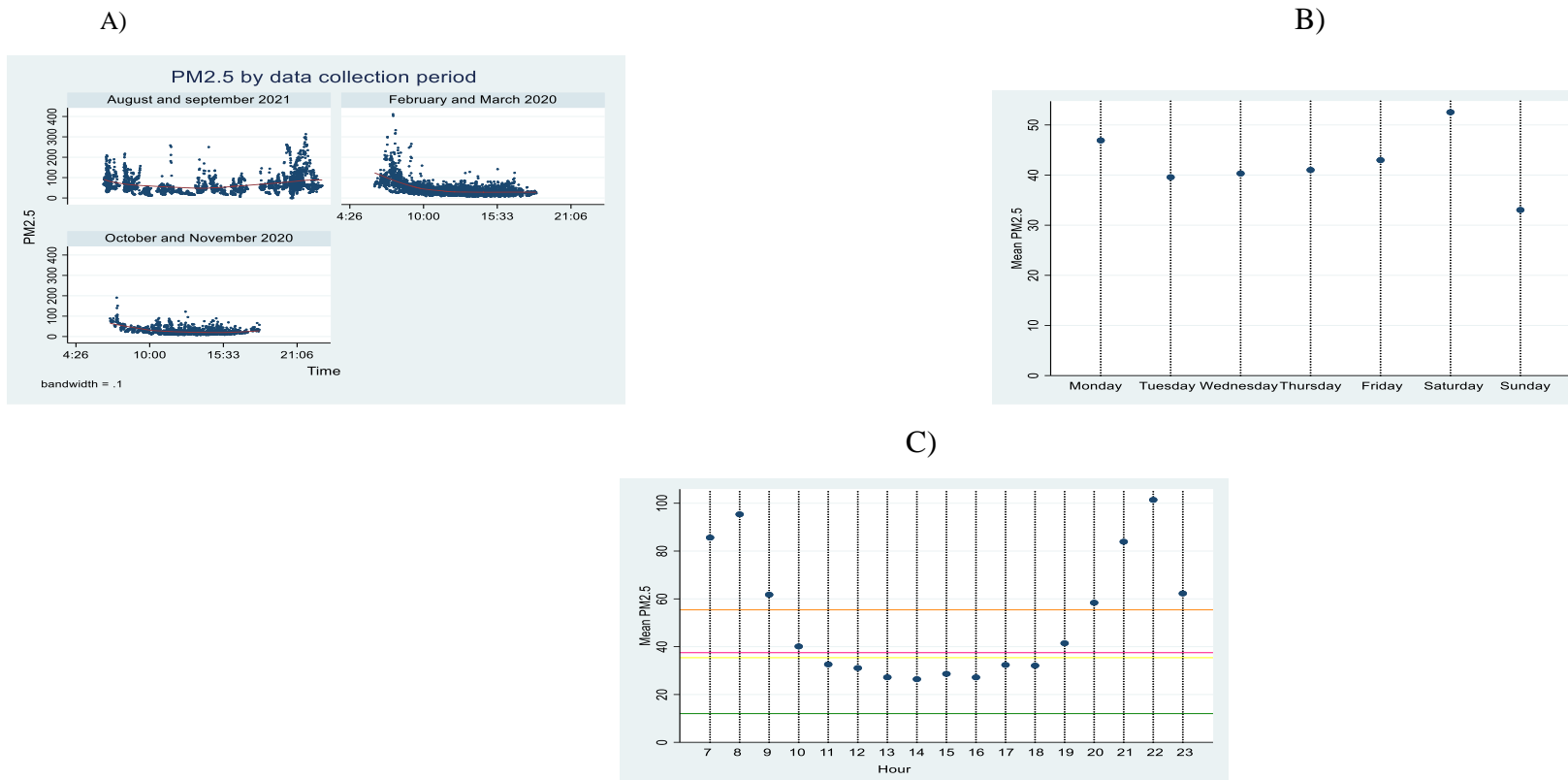
Note: Lowess graph in A shows PM2.5 by data collection round over time of day. Dot graphs B and C illustrate PM2.5 by day of the week and time of the day, respectively. Colored reference lines on the y-axis in Figure C depicts US-EPA air quality index categories and WHO IT-III (green for good air quality level, yellow for the upper limit of moderate air quality, pink for WHO interim target III, orange for the upper limit of unhealthy for sensitive people).

Figure 2.8: PM2.5 profile lowess smoother, and bar graphs
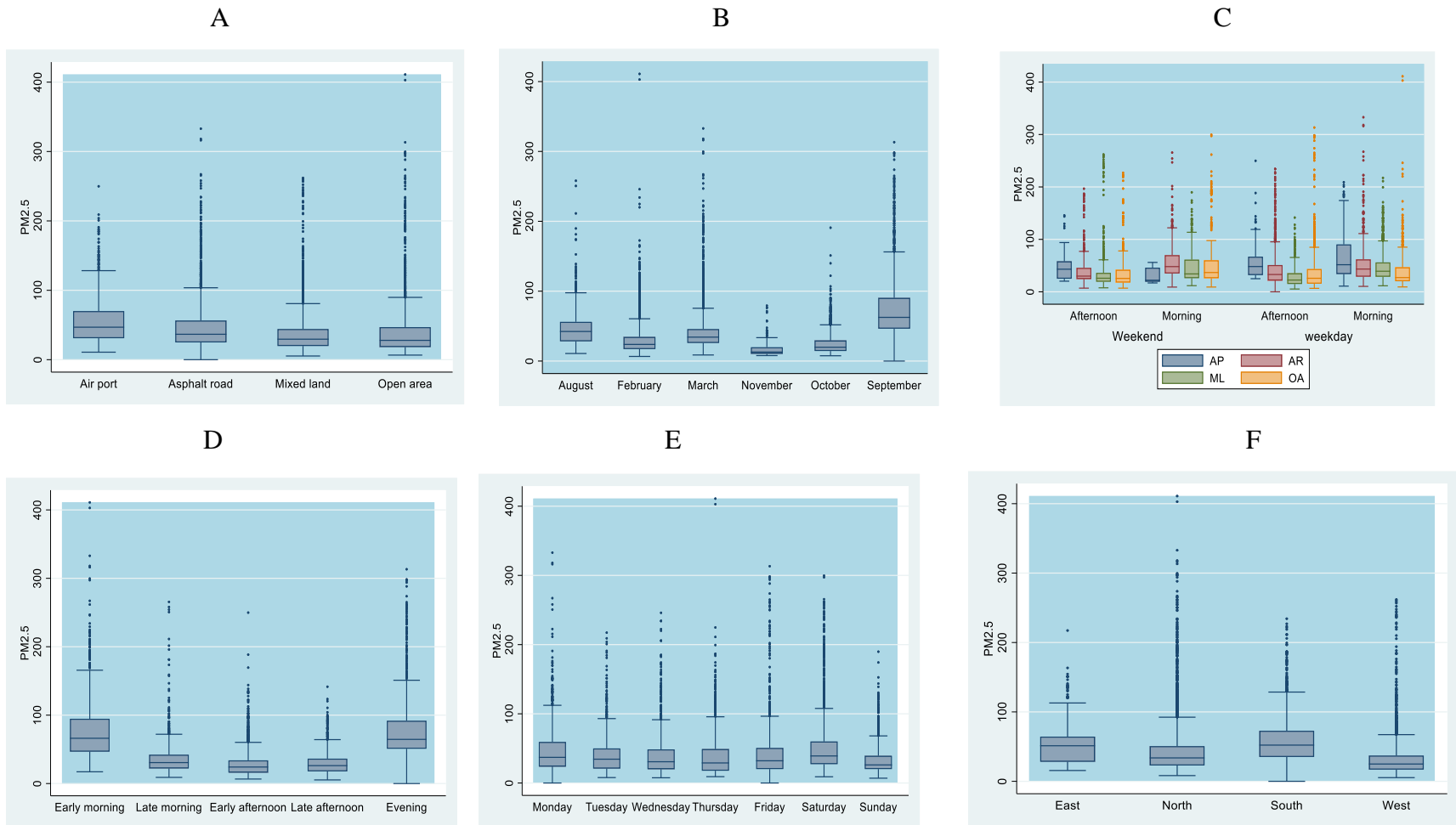
Note: Figure A shows PM2.5 by land use type, B illustrates PM2.5 by month, C shows PM2.5 by land use type over time of day, E shows PM2.5 by the time of day, E shows PM2.5 by day of a week, F shows PM2.5 by wind direction

Figure 2.9: Boxplot

2.4.2. Predictor variables and buffer size selection

Two classes of predictor variables are used in the regression: the first class of variables are specific to each given location in the modeling domain and do not change over time; this class of variables includes land use variables; variables in this category include seven major land use types in addition to two types of roads (unpaved roads and asphalt roads) which were generated from the original data (see the table below). Asphalt roads are generated by adding primary roads, secondary roads, pathways, ruck, and trunk roads) obtained from OpenStreet map, which is freely available. The second class of variables included are general to the area and change over time; this class of variables includes Meteorological variables, weather data such as temperature, humidity, wind speed, and wind gusts were collected from Addis Ababa bole international airport where the data is updated every three hours and whose variability may affect the concentration of particulate matter (PM2.5), weather variables from, National Oceanic and Atmospheric Administration (NOAA) were also used. Using the recent available land use data from Addis Ababa municipality, land use variables are classified into seven broad categories in this study. This includes Existing mixed residences, High-density mixed residences, open areas, public facility land use, commercial, parking land use, and river buffer.[10] Circular buffers were created for 100, 200, 300, 400, and 500m radii using ArcGIS pro to generate predictor variables for various distances from sample sites since there are unclear spatial scale effects of contributing characteristic variables in the literature, which

---

[10] Open area land use variable includes Embassy, urban agriculture, park, church, stadium, cemetery area, playground. public facility land use includes fire and emergency service area land, research centers, schools, various federal and regional bureaus, health centers, historical buildings, culture, and civic centers,
Commercial land use includes hotels, festival sites, market area, plaza, secondary and cattle markets.
Parking land use includes Intercity terminals, bus depot, freight terminal, parking building, bus, air and rail terminal, surface parking.

usually make results study-specific. Studies on air quality monitoring, such as Amini et al. (2014), Chalermpong et al. (2021), Hoek et al. (2008), and Lim et al. (2019) have shown that the impact of a major road on concentrations of traffic-related air pollutants declines exponentially with distance to the road, suggesting that buffer sizes should be selected to take account of known dispersion patterns. Hence following these studies, we restricted the maximum buffer size to 500m since the monitoring sites are at a small distance.

Furthermore, a study by Zhai et al.(2016) developed a LUR model of PM2.5 concentrations at different spatial scales and found that PM2.5 concentrations in Houston were significantly influenced by area ratios of open space urban and medium intensity urban at a 100m scale as well as of high intensity. A study by Hoek et al.(2008) found that the impact of a major road on concentrations of traffic-related air pollutants declines exponentially beyond about 100m from a major urban road or 500 m from a major freeway variability is limited. In this study, buffer size selection for land use and road variables is considered from previous studies (Chen et al., 2010; Eeftens et al., 2012; Jin et al., 2019). The buffer sizes are selected based on results from previous studies showing an exponential decay of pollution away from roads and varying results for land use predictors. Total length and area predictor data were intersected with different buffers, and the sum of the length for roads and total area for land use variables were calculated within each buffer for each site.

We categorized land use types based on primary economic activity that is taking place in that land use area. For instance, churches are included in an open area because church areas in Ethiopia are usually large areas covered by trees, which is also true for embassies and cemetery areas. In the public facility land use, we included mosques and various bureaus because mosques are usually in

small places with a large number of people, like federal and regional bureaus and other public places. A summary of the broad land use type classification is described in Table 2.9.

Table 2.11 A and B show summary statistics of area predictor variables for 100m buffer radius and distance land use variables, respectively. The total area by land use in respective buffers shows that the area is covered by mixed area land use, high-density mixed area land use, road network (in terms of the total area of all types of road networks), open area land use, public facility land use, commercial area land use, river buffer, and parking area land use respectively from the highest to the smallest area except for the smallest buffer (100 m) where the road network and open area land use switch each other. The rest of the land use types followed the same order as the other buffer sizes. Summary statistics of land use predictor variables for various buffer radius is shown in Table 2.12.

The average temperature during the study period in a day is 20.70, with a minimum temperature of 12 0_C and a maximum of 270_C. The average air pressure is 775.34 hpa with a minimum of 772 hPa and a maximum of 779 hPa. The average wind speed during the sampling period in Addis Ababa was 6.2 kts, equivalent to 11.48 km/h. The gust wind speed has an average speed of 7 with a min of 1kts. And a maximum of sixteen kts. The distribution of wind direction during the data collection shows the dominance of wind flow from the northern direction. About 47% of the wind direction is from the northern part of Addis Ababa, followed by the west, south, and east (Table 2.10).

Table 2.9: Summary of Land use variable categories

Variable class	Variable sub classes (components)	Measurement
Open area land use	Embassy, Urban agriculture, Park, Church, Stadium, Cemetery area, Playground	
Public facility land use	Fire and emergency service area land, research centers, schools, various federal and regional bureaus, health centers, historical buildings, culture, and civic centers	
Commercial land use	Hotels, Festival sites, Market area, Plaza, secondary and cattle markets	
Parking land use includes	Inter-city terminals, bus depots, freight terminals, parking buildings, bus, air, and rail terminal, surface parking	
Road network	The sum of the area in square meters of all road types Motorways, primary roads, primary link roads, secondary roads, secondary link roads, service roads, tertiary roads,	
Asphalt road	tertiary link roads, trunk roads, trunk link roads	
Unpaved roads	Tertiary roads, path roads, step roads, secondary roads, track roads, footway roads	Area in meter squares

Table 2.10:Summary statistics of meteorological variables

Variable	Mean	Std. Dev.	Min	Max	Wind direction	Freq.	Percent
Wind speed	6.2	2.68	1	14	East	470	4.35
Wind guests	7	3.06	1	16	North	5,091	47.1
Air pressure	775.34	1.41	772	779	South	1,401	12.96
Temperature	20.7	3.85	12	27	West	3,846	35.58
Rain	0.61	1.46	0	7			

Table 2.11: Predictor variables summary statistics
A). Area of a predictor in a one-hundred-meter radius

Predictor variables	Seventy-two sites (sampling sites)			578 sites (out of sample sites)		
	Mean	SD	Max	Mean	SD	Max
Manufacturing area				550.29	3,725	30,800
Mixed-area land use	9,349	10,916	30,791	8,836.94	11,938	30,802
High-density mixed residence	6,913.35	9,374.16	30,792	1,164.08	4,500	30,791
Total mixed-area land use	16,262.74	10,182.67	30,792	10,001.02	12,319	30,802
River buffer	1,162.94	3,052.59	17,898	2,159.93	6,091	30,800
Road network	3,588.96	3,872.36	15,590	1,784.84	3,106	19,065
Open area	5,221.47	9,696.22	30,790	7,342.71	11,902	30,804
Parking Land use	327.24	1,986.79	15,997	559.14	3,977	30,793
Commercial area	1,620.21	4,929	30,268	506.1	2,723	25,019
Public facility land use	2,411.39	4,451.47	22,677	1,582.07	5,237	30,800
Asphalt road	262.06	291.8	1,118	86.93	185	1,378
Unpaved road	357.97	240.59	879	286.46	289	1,348

B). Distance from the centroid of a prediction point to the nearest land use.

Predictor variables	Seventy-two sites (sampling sites)				Out of sample sites (578)			
Statistics	Mean	Min	SD	Max	Mean	Min	SD	Max
Waste treatment	7,275	907	2,476	11,554	5,407	138	3,396	15,478
Commercial	415	21	285	1,757	2,118	32	1,703	7,218
Bus terminal	1,406	141	648	3,110	4,007	142	2,674	11,880
Manufacturing area	2,774	220	1,021	4,527	2,610	57	1,736	9,046
Forest	2,299	327	936	4,406	1,988	26	1,583	8,259
Festival site	1,758	75	1,097	4,689	4,306	152	2,925	13,665
Embassy	940	61	604	3,078	5,716	49	3,988	15,261
Intercity bus terminal	1,307	-	657	2,898	3,033	7	1,840	8,873
Christian cemetery	1,653	51	809	3,534	2,829	119	2,207	10,499
Urban agriculture	1,151	9	695	2,994	2,243	82	1,745	8,548
Federal bureau	793	58	575	2,919	3,229	26	2,180	11,022
Woreda administration	491	15	304	1,322	2,072	55	1,495	7,267
Primary roads	288	0	228	950	1,492	0	1,533	7,539
Secondary road	522	1	746	3,588	1,964	3	1,697	7,559
Tertiary road	435	1	327	1,297	1,443	1	1,340	6,992
Foot way	560	41	353	2,127	4,957	8	3,914	15,027
High-density mixed residence	415	21	285	1,757	2,118	32	1,703	7,218
River	381	6	411	1,731	514	1	543	3,400

Table 2.12: Summary statistics of Predictor variables for each buffer

VARIABLES	The mean area of predictor variables in m^2 by buffer size				
	500	400	300	200	100
Mixed residential area(a)	293,465	186,050	100,751	42,462	9,349
High-density mixed residential area(b)	166,832	105,538	59,371	26,470	6,913
Total residential area(a+b)	460,296	291,588	160,123	68,931	16,263
River buffer area	26,848	17,529	10,874	5,148	1,163
The total road network in sq meters	85,099	55,324	32,128	13,890	3,589
Open area land use	71,685	48,762	29,290	15,476	5,221
Parking area land use	7,228	5,258	3,225	1,292	327.2
Commercial area land use	46,821	28,716	15,780	7,186	1,620
Public facility land use	65,685	41,580	23,671	10,340	2,411

2.5. Method of Prediction
2.5.1. Land Use Regression (LUR) Model

Land Use Regression (LUR) stands as a methodology employed for predicting air pollutant concentrations at a specific location. This prediction is rooted in the surrounding physical attributes, land usage patterns, and traffic factors. The approach involves amalgamating air pollutant concentration data from multiple sites and formulating stochastic models utilizing predictor variables typically acquired through Geographic Information Systems (GIS)(Chen et al., 2010). Land-Use Regression (LUR) outperforms certain Geographic Information System (GIS) interpolation techniques like kriging and inverse distance weighting. This is primarily because many GIS interpolation methods generate a uniform concentration distribution without incorporating relevant land-use details (Luke, 2017). Moreover, the interpolation of observed concentrations inadequately captures genuine spatial variations due to the typical sparsity of routine monitoring networks, which fails to represent the precise localized fluctuations in pollutant levels (Liu et al., 2015). Researchers use a stepwise forward/backward regression modelling approach, but in this research we used a standard linear regression since there is no reason to put specific criteria to include a land use variable a priori, rather finding the reason why a potential land use variable have such a relationship with the pollutant is a better approach to have a realistic model.

Therefore, the primary model function in this study is,

$$y_{it} = \rho + X_i\beta + W_t\gamma + T_t\mu + \alpha_0 LO_i + \alpha_1 LA_i + \varepsilon_{it} \tag{1}$$

46

Where y_{it} Represents the PM2.5 reading for a site i and time t, X represents a vector of site characteristics-land use, W denotes a vector of weather variables at time t (wind speed, wind direction, temperature, humidity, rain), whereas T denotes a time of day and day vector, LA_i Refers to the latitude of the site i while, LO_i Refers longitude.

2.5.2. Validation of the linear model

LUR is a predictive tool (rather than to establish causation) used to predict air quality out of the sample, either temporally or spatially. For model validation, leave-one-out-cross-validation (LOOCV) is often used to assess the generalizability of the LUR model see Dong et al. (2021) and He et al. (2018)) see also Wu et al. (2017). The dataset is separated into a training set and a testing (validation) set in hold-out validation. The training set is used to develop the model, and the testing set is used to evaluate the model by using the model to predict the output values for the data in the testing set. In k-fold cross-validation, the data set is divided into k subsets, and the holdout method is repeated k times. However, the literature on air quality prediction does not have a standard on how many observations should be used as a training data set and how many data points should be used as a validation data set for an internal validation experiment. For instance, Jin et al. (2019) used different scenarios of training and validation data sets see also Wu et al. (2017) and He et al. (2018).

We checked the generalizability of the model in prediction across buffers through LOOCV R^2 , LOOCV R^2 across buffers doesn't change much, which means that the model can be applied for different buffers as the LOOCV R^2 similar across all buffer's specific regressions (see the LOOCV results in appendix A1: Table 2.A.3). Cook`s distance value was calculated to check for outliers of data points; Cook's distance greater than one is counted as an outlier in LUR studies (Cai et al.,

2020; Li et al., 2020) but the calculated Cook`s D in this study is 0.016 which is lower than one showing that there is no significant outlier observation.

We conducted spatial and temporal Cross-validation, and we can practically apply these validation techniques. The literature uses a random group of observations as a training set, and a testing or validation set, an example of this is by Jin et al. (2019) and Masiol et al. (2018). However, putting a random group of observations as a training set and another as a testing set does not have practical applicability where a group of land use sites can be divided into a group. Our study has three major land use sites: asphalt road sites, open area sites, and mixed land use sites. To perform spatial cross-validation from the three major land-use sites, we used two major land-use sites to develop the model and one group of sites as a testing (validation) set. This process is repeated three times to check that all the groups of sites are separately used as a training (validation) set (see Table 2.13:A).

To conduct a temporal cross-validation exercise, we first considered the five time-of-a-day variables as a selection criterion to use a group of observations as a training set and a testing (validation) set. One group of observations (observed early in the morning) is used as testing (validation set), while the rest are used to develop a model. This process is repeated until the other observations are used as a testing set Table 2.13:B.

A) Spatial Cross-validation (CV) results

Group of land use sites in the training set (# of observations)	Name of testing (validation) set (# of observations)	$CV\ R^2$
Mixed and Open area (6,793)	Asphalt road (3,744)	0.42
Asphalt and open area (7,157)	Mixed land (3,380)	0.40
Asphalt road and Mixed land (7,124)	Open area (3,413)	0.40

B) Temporal Cross-validation results

Number of observations in the training set	Name of testing (validation) set (N)	$CV\ R^2$
9,049	Early morning (1,488)	0.39
7,251	Early afternoon (3,286)	0.36
9,432	Evening (1,105)	0.41
8,681	Late afternoon (1,856)	0.40
7,735	Late morning (2,802)	0.42

2.6. Model output and prediction

Two types of variables are used in regression. First, the buffer area variables (land use variables and area of roads) and length (for roads); are local area-specific variables that are defined over a specific buffer (e.g., 100, 200m). Air quality is affected by local land use. However, it is unclear how to define local. Some significant effects may be lost if we use too big a buffer because the variation in the X variables goes down (if we use a buffer of, say, 20km, then X becomes a constant). On the other hand, there may be measurement errors in land use, and also air quality at a location may genuinely be affected by land use in the larger neighborhood; hence we varied our buffers considering this and following previous works that suggest some useful insight about buffer selection for explanatory variables see Hoek et al.(2008) see also Zhai et al. (2016).

The other group of variables (meteorological variables) is assumed to be uniform across the domain. These variables include rain with its lag, wind speed, temperature, and air pressure (humidity). Non-buffer predictor variables wind speed and lag rain are negatively related to PM2.5 concentration, while air pressure (humidity) is positively related to PM2.5 concentration.

When we look at the relationship between land use variables and PM2.5 concentration, most of the land use variables have expected signs, although some are insignificant. Road network (all types of roads) is positively related to particulate matter concentration (more roads mean more PM2.5 concentration). Public facility land use is negatively related to PM2.5, while commercial area land use variables have a positive relationship with PM2.5. The overall explanatory power of the model is not particularly sensitive to buffer size, as can be seen from the model R^2. However, a smaller 100m buffer picks the most crucial land use variables. Therefore, a smaller buffer is more suitable for prediction on those grounds. Hence in the analysis, we considered the regression output that is based on 100m land use variables see Table 2.14 column 1 see also Appendix 1: Table 2.A.5(a) for regression outputs using distance land use variables and Table 2.5 (b) for the calibrated version LUR results of the Dylos reading.

The dummy for a time of the day indicates that evening contributes the highest concentrations, followed by early in the morning. Conversely, the early afternoon has the lowest concentration effects, followed by the late morning. This indicates that the concentration of particulate matter is associated with household activities and activities related to motor vehicles, as traffic flow is very high in the early morning and evening.

TABLE 2.14: LUR of the full sample

Variables	100m	200m	300m	400m	500m
	PM2.5	PM2.5	PM2.5	PM2.5	PM2.5
Rain	0.264	0.200	0.146	0.105	0.099
	(0.545)	(0.521)	(0.492)	(0.478)	(0.479)
One day lag rain	-2.732***	-2.656***	-2.616***	-2.617***	-2.648***
	(0.825)	(0.839)	(0.844)	(0.849)	(0.852)
Wind speed	-0.818	-0.974	-1.046	-1.097	-1.101
	(0.760)	(0.758)	(0.756)	(0.757)	(0.758)
Humidity	1.821**	1.969**	1.989**	1.976**	1.959**
	(0.740)	(0.745)	(0.753)	(0.754)	(0.750)
Temperature	-5.790*	-6.050*	-6.361*	-6.439*	-6.333*
	(3.223)	(3.301)	(3.326)	(3.369)	(3.388)
Temperature squared	0.136*	0.141*	0.148*	0.150*	0.148*
	(0.073)	(0.075)	(0.076)	(0.077)	(0.078)
High density mixed residence	4.110	3.599	0.657	-1.345	-2.435
	(2.903)	(3.796)	(3.985)	(4.541)	(4.631)
River buffer	6.218	2.544	10.222	21.218	31.767
	(7.226)	(8.761)	(13.010)	(18.835)	(25.086)
Open area	-4.129	-2.873	-2.411	-4.919	-6.743
	(2.721)	(4.074)	(5.957)	(8.192)	(9.484)
Parking land use	-0.792	12.830	49.458**	68.336***	80.840***
	(4.243)	(8.142)	(19.351)	(23.004)	(28.112)
Commercial area	4.778*	4.918	2.264	3.187	4.077
	(2.683)	(3.406)	(5.061)	(6.256)	(6.615)
Public facility land use	-10.882**	-8.804	-10.950	-11.573	-11.224
	(4.535)	(6.033)	(7.582)	(8.394)	(7.974)
Asphalt road	0.011***	0.003**	0.001	0.001	0.000
	(0.003)	(0.001)	(0.001)	(0.001)	(0.000)
Peak hour#Asphalt	-0.002	0.001	0.001	0.001	0.001
	(0.008)	(0.004)	(0.002)	(0.001)	(0.001)
East	-13.620***	-13.480***	-13.180***	-12.904***	-12.743***
	(4.037)	(3.940)	(3.918)	(3.915)	(3.967)
South	-6.814	-7.006	-6.903	-6.741	-6.643
	(4.337)	(4.335)	(4.302)	(4.260)	(4.223)
West	-7.513***	-7.744***	-7.686***	-7.527***	-7.529***
	(2.520)	(2.566)	(2.552)	(2.528)	(2.476)
Late morning	-37.341***	-36.314***	-34.831***	-32.575***	-31.009***
	(3.887)	(4.105)	(4.295)	(4.678)	(5.490)
Early afternoon	-43.551***	-42.758***	-41.381***	-39.137***	-37.614***

	(3.822)	(3.947)	(4.048)	(4.392)	(5.163)
Late afternoon	-39.004***	-37.880***	-36.446***	-34.286***	-32.875***
	(4.136)	(4.212)	(4.318)	(4.669)	(5.525)
Evening	-3.009	-2.302	-2.405	-2.454	-2.076
	(6.498)	(6.136)	(6.008)	(5.983)	(5.964)
Tuesday	-10.892***	-10.660***	-10.506***	-10.369***	-10.232***
	(3.111)	(3.229)	(3.281)	(3.280)	(3.254)
Wednesday	-9.153***	-8.884**	-8.525**	-8.259**	-8.232**
	(3.281)	(3.396)	(3.452)	(3.462)	(3.436)
Thursday	-11.857***	-11.572***	-11.096**	-10.942**	-10.868**
	(4.205)	(4.269)	(4.305)	(4.273)	(4.267)
Friday	-4.104	-3.952	-3.822	-3.750	-3.638
	(3.912)	(4.087)	(4.156)	(4.178)	(4.172)
Saturday	-4.296	-4.210	-4.230	-4.132	-4.109
	(4.278)	(4.302)	(4.311)	(4.311)	(4.284)
Sunday	-10.655***	-10.399***	-10.104***	-9.917***	-9.872***
	(3.197)	(3.316)	(3.373)	(3.398)	(3.391)
Latitude	189.350** *	169.436** *	160.753** *	166.222** *	171.859** *
	(37.885)	(40.901)	(42.458)	(42.792)	(43.399)
Longitude	-9.604	-16.472	-7.650	-15.353	-15.098
	(39.166)	(40.496)	(40.844)	(41.125)	(42.449)
Constant	-1,209.079	-761.470	-1,022.848	-774.594	-837.764
	(1,557.918)	(1,594.348)	(1,598.383)	(1,600.255)	(1,634.062)
Observations	10,537	10,537	10,537	10,537	10,537
R-squared	0.410	0.405	0.404	0.405	0.404
F test(7,71)					
	5.29	6.1	2.65	2.93	2.77
	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Numbers in the first raw shows the buffer radius of a land use variable used in the regression (e.g., 100m refers, in this column all land use variables used are measured in a 100m radius from the site). Land use measurements are expressed in proportions (area of a land use at site i/ total area of a land use at 72 sites)

2.6.1. Out-of-sample prediction

To conduct out-of-sample prediction, we divided the whole city of Addis Ababa into 1km-by-1km grids, with the centroid of each grid being a point where all buffers (100m, 200m, 300m, 400m, 500m) are created to generate land use variables, and all data points (sites) where air quality prediction is taken are counted 578. After obtaining all land use variables (area for land use and road length) for each site, the weather variables available at the city level were merged. The established LUR model predicted the PM2.5 concentrations at the centroids of each grid cell.

2.6.2. Investigating the temporal and spatial structure of measured and predicted PM2.5 concentrations.

Understanding the spatial structure of air pollution is critical for deciding where to place sampling sites, identifying major pollutants and taking corrective action, and developing future urban plans Jin et al., (2019). In this section, we compare measured and predicted PM2.5 concentrations within and outside of sample sites using US-EPA PM2.5 implied risk classifications (See US-EPA implied risk classifications for PM2.5 in Appendix 1: Table 2.A.4. Comparing predicted and raw observations within sampling sites shows that 73% of predicted and actual observations fall into the same US-EPA implied risk category for air quality. When comparing predicted and raw concentrations by land use type, open area sites agree at 91%, mixed land at 73%, and asphalt roads at 56% of the total data. That is, the model has the potential to identify areas/locations with poor air quality and areas/locations with good air quality. Similar temporal patterns of air quality are observed between predicted values and actual air quality readings from Dylos 1700 within sampling sites (poor early in the morning and late evening and lower pollution at other times of the day), indicating that the model captures temporal variability in air quality; see Table 2.16 (a) for in-sample actual and predicted values by the time of day and Table 2.16(b) for out of sample

prediction by the time of day. Table 2.16 shows a comparison between overall actual PM2.5 observations and predicted PM2.5 using US-EPA implied risk categories for PM2.5.

TABLE 2.15: Comparison between raw concentrations and predicted values by land use type.

	Overall matched 73%				AR, Match 56%			
Predicted →	Predicted| Dylos 1700	Good	Moderate	Unhealthy for sensitive people	Unhealthy	Good	Moderate	Unhealthy for sensitive people	Unhealthy
	Good	0	0	0	0	0	0	0	0
	Moderate	0	48	8	0	0	20	16	0
	Unhealthy for sensitive people	0	14	21	4	0	16	32	12
	Unhealthy	0	0	0	4	0	0	0	4

	ML matched 73%				OA, Match 91%			
	Predicted| Dylos 1700 (Mean)	Good	Moderate	Unhealthy for sensitive people	Unhealthy	Good	Moderate	Unhealthy for sensitive people	Unhealthy
	Good	0	0	0	0	0	0	0	0
	Moderate	0	65	9	0	0	61	0	0
	Unhealthy for sensitive people	0	17	4	0	0	9	26	0
	Unhealthy	0	0	0	4	0	0	0	4

TABLE 2.16: Comparison between raw concentrations and Predicted values by the time of day(within sample sites)

a) Actual observations

Actual observations	Early morning	Late morning	Early afternoon	Late afternoon	Evening
Good	0	1	3	1	0
Moderate	2	16	21	12	1
Unhealthy for sensitive people	4	8	6	3	2
Unhealthy	9	2	1	1	6

b) Predicted values

Predicted values	Early morning	Late morning	Early afternoon	Late afternoon	Evening
Good	0	0	0	0	0
Moderate	0	14	28	14	0
Unhealthy for sensitive people	0	13	3	3	0
Unhealthy	14	0	0	0	11

b) Out-of-sample prediction (578 sites)

Out-of-sample prediction	Early morning	Late Morning	Early afternoon	Late afternoon	Evening
Good	0	0	1	0	0
Moderate	0	19	20	10	0
Unhealthy for sensitive people	0	8	3	3	0
Unhealthy	24	0	0	0	12

C) Comparison between predicted and actual Dylos 1700 reading within the model sites (72 sites)

		Actual reading Dylos 1700↓			
	predicted↓ Dylos 1700→	Good	Moderate	Unhealthy for sensitive people	Unhealthy
Predicted→	Good	0	0	0	0
	Moderate	5	40	10	2
	Unhealthy for sensitive people	0	10	6	3
	Unhealthy	0	3	6	15

2.6.3. Mapping Predicted PM2.5

To better understand the spatial and temporal distribution of pollutant concentrations, the obtained values were spatially interpolated using the Inverse Distance Weighted (IDW) method. The concentration values obtained were spatially interpolated to assign a contamination value to each point in the studied area. For spatial interpolation, the Inverse Distance Weighted (IDW) method was used. This method assumes that nearby points will have a greater influence on the interpolating surface than distant points. The interpolating surface is a weighted average of the scatter points, with the weight of each scatter point decreasing with increasing distance between the interpolation point and the scatter point. IDW interpolation does not require any assumptions about the distribution or behavior of the measurements. This method yields exact results at the sampling points and behaves smoothly, with no abrupt changes in measurement points Lozano et al. (2009). Figure 2.11 depicts predicted PM2.5 by the time of day, while Figure 2.12 depicts predicted PM2.5 by wind direction. The maps show that the city center is more prone to elevated pollution and hotspots, whereas less polluted areas are outside, which can be verified from the map for the whole city of Addis Ababa in Figure 2.A.4 and 2.A.5. The less polluted areas have lower population density, more forested areas, and lower road density. On the other hand, the areas with high road density, commercial establishments, and heavy traffic volumes in the city center have elevated pollution levels, emphasizing the influence of land use variables on predicting pollution.

Moreover, the pollution level during peak hours is nearly twice as high as that observed during off-peak hours. This heightened pollution level is concentrated at the city center, where population density is at its peak, and significant commercial activities are bustling. Although the site type is not explicitly known during out-of-sample prediction, one can see that more polluted sites are in

the city center, and less polluted sites are in the outskirts. Furthermore, winds blowing to the south carry more polluted air than winds from other directions.

Examining the pollution variation in the city based on wind direction reveals significant fluctuations, particularly concerning land use patterns when the wind originates from the north or south. This observation can be attributed to the unique topography of Addis Ababa, which features sloping terrain and a mountainous region in the north. These geographical factors influence the airflow within the city, subsequently affecting the dispersion of pollution. When the wind blows from the north or south, it encounters the varying land uses in the city, leading to diverse pollution levels across different areas. The mountain in the north may act as a barrier or channel for the airflow, causing variations in pollution concentration depending on the direction and strength of the wind.

Understanding the impact of wind direction on pollution dispersion is vital for developing effective strategies to address pollution hotspots and improve overall air quality. By identifying the areas most affected by pollution under specific wind conditions, city planners and policymakers can implement targeted measures to mitigate pollution and promote a healthier environment for the city's residents.

However, it is crucial to acknowledge the significant constraint of extending the prediction to additional sub-cities shown in appendix 1 Figure 2.A.3 and 2.A.4. For instance, while the inner sub-cities predominantly exhibit urban characteristics, certain outer sub-cities tend to reflect more rural attributes. Additionally, the lack of sufficient PM2.5 data during rainy and nighttime conditions further prevents us from drawing a conclusive judgment regarding the accuracy of the prediction.

Overall, the predicted PM2.5 concentration maps are consistent with land use distribution and traffic flow at various times of the day. The areas with the poorest air quality tend to be those closest to where the data was collected, primarily commercial activities with dense roads and residential areas. Nonetheless, a few suburbs to the east and south have higher PM2.5 levels. The maps, however, revealed similar spatial pollution patterns across the time of day.

When we look at the predicted pollution level across the city using US-EPA air quality classification and implied risk of PM2.5, 70% of the area in Addis Ababa exhibits unhealthy air in the morning and evening. In contrast, early afternoon (13:00-15:00) has the least polluted air, with approximately 70% of the area having moderate air quality and 28% having good air quality. Late morning (10:00-12:00) and late afternoon (16:00-18:00) show the second least polluted air in the city, with both times of the day having 60% of the area exhibiting moderate air quality. This information suggests that air pollution levels in Addis Ababa vary throughout the day, with the early afternoon having the best air quality and the morning and evening having the worst. It is important to note that prolonged exposure to unhealthy air can harm health, particularly respiratory health. Therefore, residents of Addis Ababa may want to consider taking precautions during times of high pollution, such as limiting outdoor activities, using air purifiers or masks, and staying informed about air quality levels in their area. Additionally, policymakers and city officials can use this information to develop and implement measures to reduce air pollution in the city.

Analyzing prediction maps and their correlation with land use variables provides crucial information for understanding the spatial distribution of air pollution in urban areas. It can also help identify the factors contributing to elevated pollution levels in high-risk areas and inform the development of effective policies and interventions to mitigate pollution levels and promote public health.
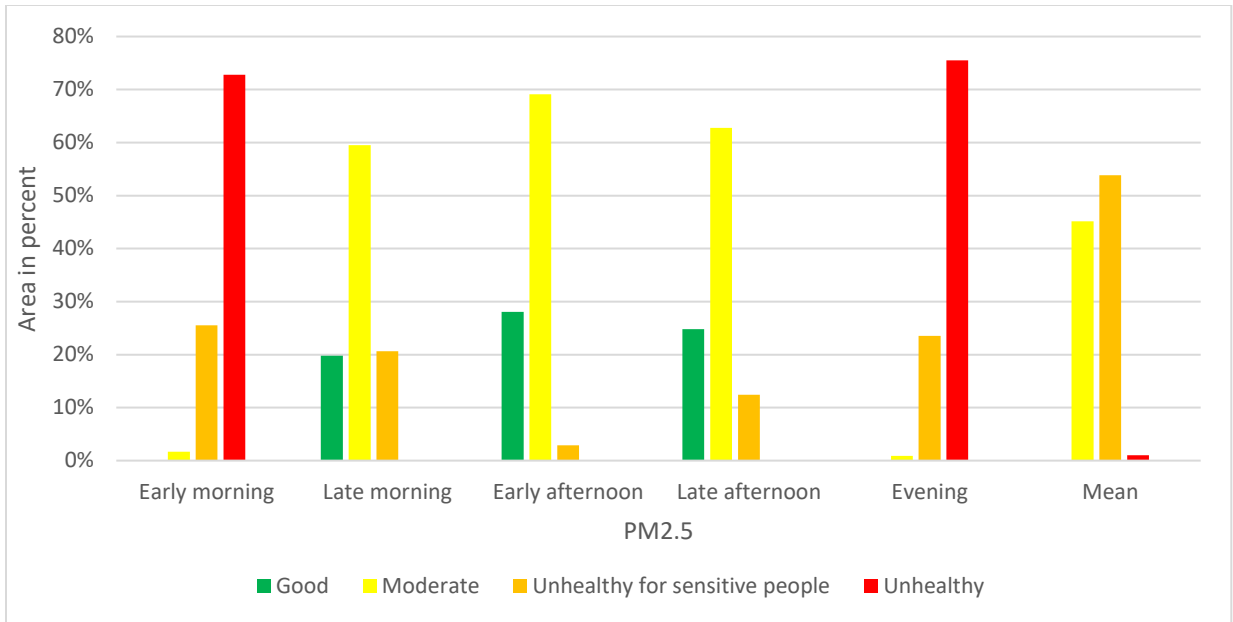
Note: After getting the pollution levels using linear prediction from the land use regression, pollution levels were reclassified using US-EPA air quality classification using ArcGIS pro; areas for each level of air quality were also calculated using the ArcGIS Pro *reclassify* function.

Figure 2.10 Air quality by area

A) PM2.5 in Addis Ababa during Peak hours
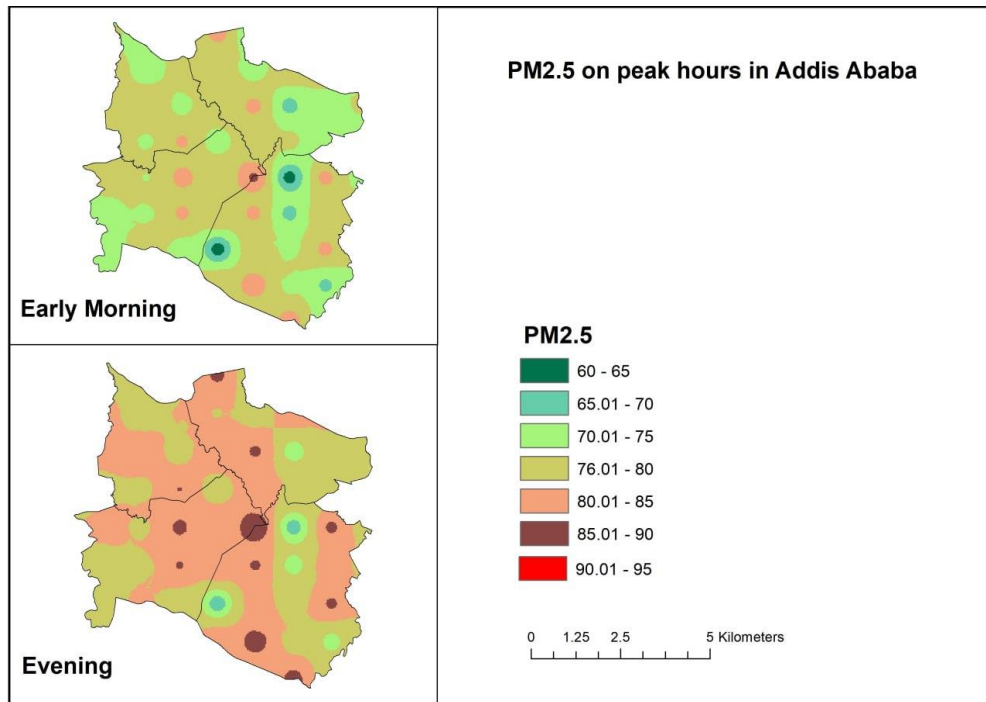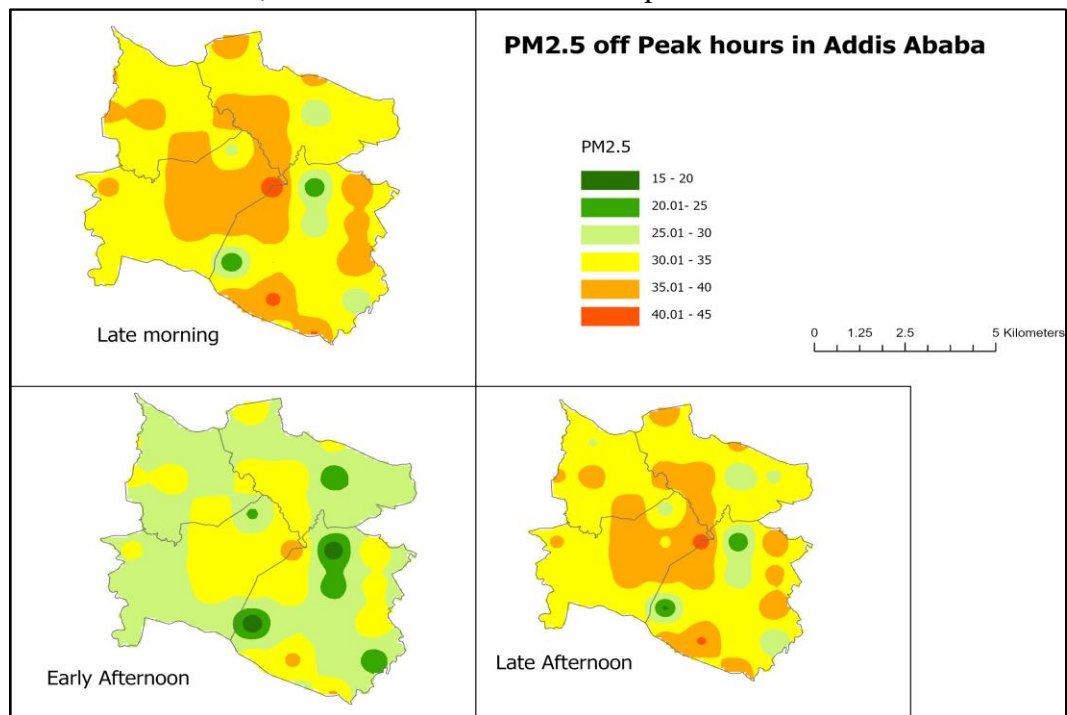
Note: Pollution levels are categorized into six ranges to illustrate spatial and temporal variations.

Source: Author`s illustration using ArcGIS Pro

Figure 2.11: Map of predicted PM2.5 by the time of day for the four sub-cities in Addis Ababa

PM2.5 by wind direction

North

West

PM2.5

- 20 - 30
- 30.01 - 40
- 40.01 - 50
- 50.01 - 60
- 60.01 - 70

0   1.25   2.5   5 Kilometers
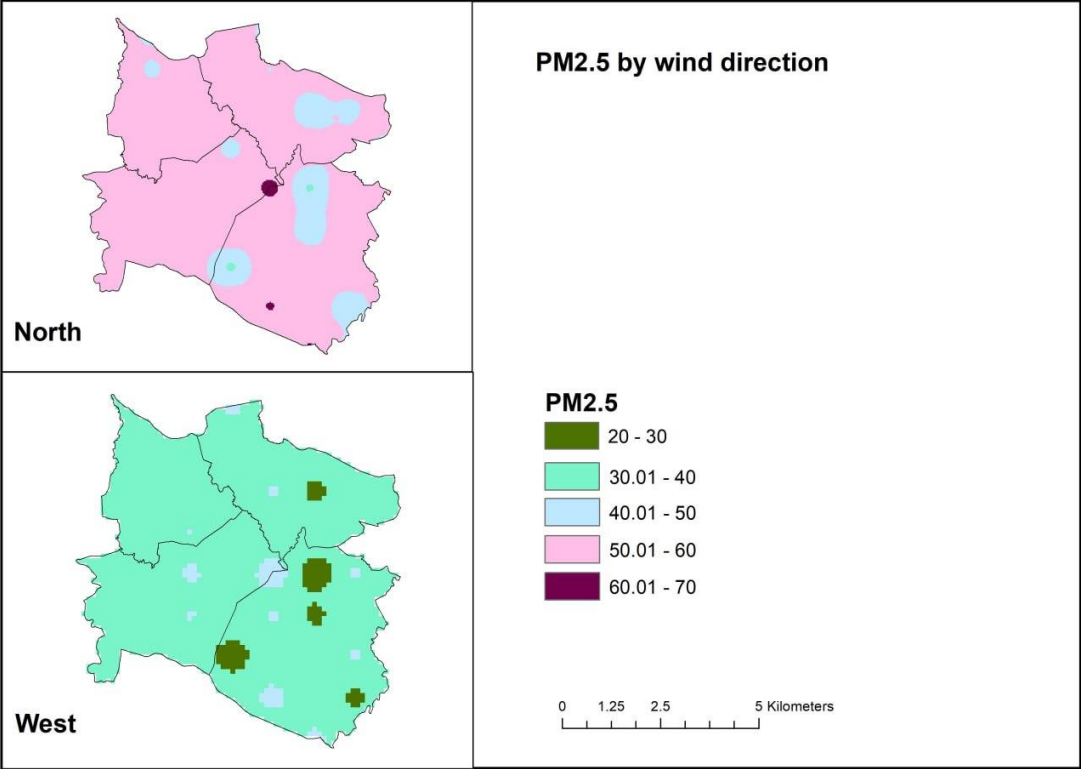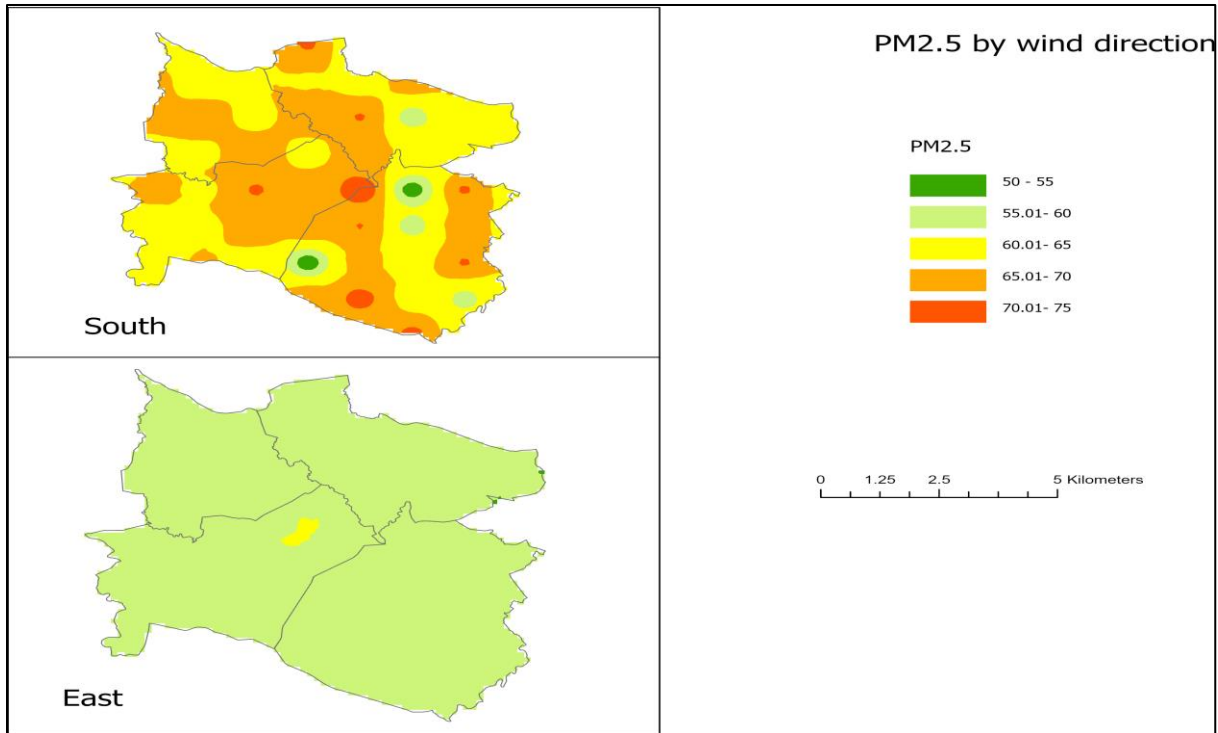
Note: The maps based on wind direction are averages of the data collection period.

Source: Author`s illustration using ArcGIS Pro

Figure 2.12: Map of predicted PM2.5 by wind direction for the four sub-cities in Addis Ababa

2.6.4. Comparison of this study with previous LUR studies

A comparison with previous studies is presented in Table 2.17. The predictive performance of LUR models is mainly measured by R^2 and robustness is measured by LOOCV (Cai et al., 2020; Li et al., 2020). R^2 and LOOCV of this study are 0.398 and 0.394 respectively which is comparably better than that of Chalermpong et al. (2021) in Thailand. This study performance is much lower than the one conducted in Uganda by Coker et al. (2021), where they used locally manufactured mobile air quality monitors (AirQo).

TABLE 2.17: Comparison of this study with LUR studies in Developing countries/Developed countries

Study area	Pollutants of concern	R^2	LOOCV	Reference
Addis Ababa, Ethiopia	PM2.5	0.398	0.394	This work
Taipei–Keelung metropolitan area	PM2.5	0.72	0.53	Li et al. (2020)
	PM10 and O3	0.80	0.72	
	NO2	0.91	0.88	
Aachen, Germany	PM2.5	0.655	-	Merbitz et al. (2012)
	PM10	0.79	-	
Lanzhou, China	NO2	0.71	0.64	Jin et al. (2019)
Bangkok, Thailand	PM2.5	0.321	0.236	Chalermpong et al. (2021)
Uganda, Cities include Kampala and Jinja, Mukono, a Wakiso	PM2.5	0.84	-	Coker et al. (2021)

2.6.5. The Social Benefits of changing land use and Meeting WHO annual interim targets.

Quantifying the social cost of air pollution and the benefit of reducing it is a critical component of public health policy in many developing countries. The benefit of reducing PM2.5 pollution has been discussed by many researchers (Chen et al., 2017; Qu et al., 2020) and the social cost of pollution (Roy, 2016; Yamada et al., 2023; Yin et al., 2017). In this part, I estimate the social benefit of a change in PM2.5 due to a change in a land use type, and I discuss how this policy could be a cost-effective way of reducing PM2.5.

In recent years, Addis Ababa has undergone significant development with the aim of improving its livability and aesthetics. Two projects related to the transformation of land use from private to public have been initiated. The first project, the Addis Ababa Riverside Project, focuses on converting the city's riverside areas into green public spaces for leisure and recreation by developing parks, walkways, and cycle paths along the riverbanks. The second project, the Addis Ababa Beautification Project, aims to enhance the city's aesthetics by planting trees, flowers, and other greenery along streets, constructing roundabouts, and installing public art. These two projects are part of the larger effort by the Addis Ababa City Administration to make the city more livable, sustainable, and attractive to both residents and visitors. The Addis Ababa Riverside Project seeks to reclaim the city's urban rivers and create vibrant public spaces for leisure and recreation. This project aims to transform the banks of rivers, covering a total of 69 kilometers (AfDB, 2021). [11]

The Addis Ababa Beautification Project is focused on improving the city's aesthetics by introducing green spaces and public art. This project includes the planting of trees, flowers, and

---

[11] https://www.afdb.org/en/news-and-events/addis-ababa-beautifying-sheger-river-development-project-56625

other vegetation, constructing roundabouts, and installing public art pieces. The aim of this project is to make Addis Ababa more attractive and inviting while improving air quality and creating more pleasant public spaces for people to enjoy.

Overall, these two projects are aimed at improving the quality of life for residents and visitors to Addis Ababa. The city is undergoing rapid growth and urbanization, and these initiatives aim to ensure that growth is sustainable and people centric. By converting underutilized spaces into vibrant public areas, the city is working to create a more livable and inclusive urban environment. (AfDB, 2021.). [12]

The current low-cost monitor reading average PM2.5 is 42.39 which is unhealthy air for sensitive people. To change the average unhealthy PM2.5 into moderate air (35.4-the is the upper limit of moderate air) we need to change 13,679.06 square meters of commercial land to a public land use or include this amount of public land use in commercial land in each one-kilometer grid. Our intention behind transforming commercial land use into public land use is not centered around the transformation of commercial establishments or buildings into parks. In the context of a rapidly developing city like Addis Ababa in a developing nation, there exist vacant areas even within commercial zones. These spaces have the potential to be repurposed as modest green spaces. Such a conversion would contribute to the reduction of pollution levels. These small green spaces have the potential to mitigate environmental issues that often plague densely populated urban centers. Vegetation and greenery play a crucial role in absorbing pollutants, generating oxygen, and regulating temperature. As a result, converting neglected corners of commercial zones into green

---

[12] https://www.afdb.org/en/news-and-events/addis-ababa-beautifying-sheger-river-development-project

spaces through the currently undergoing beautification projects of Addis Ababa can help combat air pollution and contribute to a healthier urban ecosystem.

$$\Delta PM2.5 = -0.000363 * \Delta \text{ public facility land use} + 0.000151 * (-\Delta \text{commercial land use})$$

$$(42.39 - 35.4) = -0.000363 * X + 0.000151 * (-X)$$

$$6.99 = -0.000511X$$

$$X = 13,679.06 \text{ square meters in each 1 km grid}$$

As part of estimating the health benefit of improved air quality due to a change in land use, I use a method used by (L. Chen et al., 2017; Qu et al., 2020) which was designed by the US Environmental Protection Agency (EPA) to assess the health benefits of improving air quality. Since death is the most significant endpoint of various health effects related to PM2.5 pollution, I choose all-cause death as the end point of health effects.

The health benefits obtained by controlling the PM2.5 concentration can be represented by

$$\Delta Y = Y_0 (1 - e^{-\beta \Delta PM2.5}) * \text{pop}$$

$$\beta = \frac{Ln(RR)}{\Delta PM2.5}$$

In the formula, ΔY represents the change in deaths, Y_0 represents the baseline death rate for all mortality, and β is obtained from the relative risk (RR) linked to a change in exposure. Based on the findings of Hoek et al. (2013) and WHO (2013), the baseline mortality relative risk is taken as 1.06 for a 10 µg/m3 PM2.5 increase. The death rate is obtained from the Global Burden of Disease

(GBD), which is seven per ten thousand people, while the total population of Addis Ababa is derived from the World population database and is 5.06 million.[13]

To calculate the economic benefit of a reduction in air pollution we used the Value of Statistical Life approach. The VSL, reflects the amount individuals are willing to pay to incrementally reduce their risks of death from adverse health conditions that might be caused by environmental pollution (Ho et al., 2023; World bank, 2016). For countries that lack VSL value, VSL is converted from another reference, such as the USEPA or the Organization for Economic Cooperation and Development (OECD).

I used a VSL estimation method from the Organization for Economic Cooperation and Development (OECD) as implemented by the World Bank in 2016. The base VSL estimate represents the mean VSL estimate from a database of WTP studies conducted in high-income member countries of the Organization for Economic Co-operation and Development (OECD). The estimated VSL_{base} by the World Bank in 2016 is \$ 3.83 million. The average gross domestic product (GDP) per capita included for the OECD countries is about \$37,000 (World bank, 2016).

The VSL for Ethiopia is calculated using the following formula.

$$VSL_{Eth} = VSL_{base} * \left(\frac{GDP \; per \; capita_{Eth2021}}{GDP \; per \; capita_{OECD}} \right)^e$$

Where $GDP \; percapita_{Eth2021}$ is Ethiopia's per capita income in 2021, e is the income elasticity of VSL, for low- and middle-income countries, a central value of 1.2 is recommended by WHO, with a range from 1.0 to 1.4 for sensitivity analysis by the World Bank. The calculated VSL for Ethiopia is about \$154,494.74. The total economic benefits of a change in land use in this exercise

---

[13] https://vizhub.healthdata.org/gbd-results/
  https://worldpopulationreview.com/world-cities/addis-ababa-population

amounted to about $1.093 billion per year. That is, if we change the mean PM2.5 by 6.99 due to a change in land use from commercial to public or the inclusion of public land use in the commercial land use (to reach at the upper limit of moderate air), then 7078 deaths can be avoided per year, and the Economic benefit of this is $1.093 billion in one year. In addition, we calculated the benefits of meeting WHO interim targets as interim targets were set by WHO to encourage developing countries in the difficult task of reducing pollution on the way to economic growth. Meeting the lowest annual interim target (Interim target one) of WHO has a monetary benefit of $1.1 billion per year, while meeting the highest interim(interim target I) target gives a benefit of $2.5 billion per year (See table 2.18).

This exercise highlights the benefits of incorporating green spaces in commercial areas as a means to enhance public health and reduce mortality rates and the benefits of gradually meeting WHO-recommended interim targets. It suggests that meeting the lowest interim target and a small increase in greenery can significantly impact the well-being of city residents. The exercise underscores the importance of considering the health benefits of urban green spaces in urban planning and development. Moreover, this exercise is particularly relevant to Addis Ababa, where rapid urbanization and industrialization have led to high pollution levels, posing significant health risks to the population. The results of this exercise provide valuable insights into how the city can mitigate the negative impacts of urbanization and improve the livability of the city by introducing more green spaces in commercial areas. Overall, urban green spaces can play in promoting public health and well-being. By incorporating more greenery in cities, policymakers can create more livable and sustainable environments that support the health and well-being of their citizens.

TABLE **2.18** The economic benefits of meeting WHO annual interim targets and a change in land use

WHO Recommended AQG and Interim annual target	PM2.5	Total annual Economic Benefit in USD	Avoidable deaths per year
IT-4	35	1.1 billion	7,249
IT-3	25	1.7 billion	11,160
IT-2	15	2.2 billion	14,269
IT-1	10	2.5 billion	15,610
AQG	5	2.6 billion	16,837
A change in LU	35.4	1.093 billion	7,078

Note: IT- refers to Interim target; AQG- refers to air quality guideline

2.7. Conclusion

The primary goal of this chapter was to develop a cost-efficient approach for predicting air quality using a land use regression model to predict pollution levels in areas that are not being monitored. The land use regression model used in this study was developed based on a range of variables, including traffic density captured by road types, land use patterns, and other environmental factors that are known to influence air quality.

We created a PM2.5 pollution exposure map for Addis Ababa. The overall explanatory power of the model is not particularly sensitive to the buffer size. This is because essential factors in the model are wind speed, direction, air pressure (humidity), and time of day. Moreover, smaller buffers also pick up the effects of local land use variation. Therefore, a smaller buffer is more suitable for prediction on those grounds. Hence when developing the exposure map for Addis Ababa, we use the smaller buffers as a base for prediction. This way, the exposure map is shown explicitly for the city by the time of the day and wind direction.

In general, in this study, we establish a regression model for predicting PM2.5 and reveal spatial and temporal distribution characteristics. Further, we analyze the relationship between the spatial distribution of PM2.5 with meteorological and land use characteristics and therefore provide a foundation for urban planning, land use regulation, air pollution control, and public health policymaking. This study also provides a basic model for population exposure assessment. Applying land use regression models to heavily polluted areas play a

positive role in achieving a sustainable urban environment and promoting sustainable development in urban environments.

Due to the limited resources and time constraints to facilitate logistics to deploy a large number of low-cost monitors, we could not measure PM2.5 concentrations at multiple sites simultaneously, which is a major limitation of this study. In addition, the low-cost monitoring approach has some other significant limitations. First, the data we used are representative of the period in which we collected, which is typically a constrained time frame. The data collection in this dissertation was carried out in the daytime, between 6 a.m. and 19 p.m., for several days over a 6-month period. Maintaining measurements for longer periods and obtaining more extensive temporal coverage may be preferable. However, low-cost monitor data collection is frequently restricted to rush hours (e.g., commuting routine) or daytime moments for exposure assessments. However, given that this study was conducted within a short period, emission profiles in different locations will not be expected to change dramatically. Therefore, the measured PM2.5 spatial and temporal variation, at a minimum, suggests the heterogeneous pattern of PM2.5 in Addis Ababa.

Care should be taken when using air pollution maps because the mode of transportation used determines the usefulness of the maps (Joris, 2016). In this study, we did not measure pollution levels while traveling from one location to another, but we used a car to get from one location to another. As a result, the exposure maps produced do not represent car users, residents, or pedestrians; rather, the maps depict pollution variability by time and location to investigate pollution exposure further or devise pollution control measures.

A low-cost monitoring strategy will always be spatially and temporally constrained, resulting in a map with insufficient spatial and temporal coverage. A typical low-cost air quality monitoring system is restricted to a few sites and time periods. Depending on the specific configuration, an opportunistic campaign utilizing multiple modes of transportation can result in increased spatial and temporal coverage. Using low-cost and high-performance sensors, data can be collected on a larger scale in an ad hoc, opportunistic fashion. As a result, participatory monitoring, which is based on mobile monitoring, has the potential to deliver large amounts of data needed for producing representative maps with high spatial and temporal coverage.

To address air pollution in Addis Ababa, a comprehensive approach may be necessary. This could involve a combination of measures aimed at reducing emissions from various sources, improving public transportation, promoting alternative modes of transportation such as cycling and walking, and raising awareness about the health impacts of air pollution. For example, the government could incentivize the use of cleaner vehicles by offering tax breaks or subsidies. Additionally, regulations could be put in place to limit emissions from factories and other industrial sources. Improved public transportation options, such as more efficient buses or light rail systems, could encourage people to rely less on personal vehicles.

Moreover, promoting cycling and walking can not only reduce emissions but also promote physical activity and overall health. This could be achieved by building more bike lanes and sidewalks and providing bike-sharing programs.

Finally, raising awareness about the health effects of air pollution could help mobilize public support for action. This could involve public education campaigns, school programs, and community outreach initiatives to inform residents about the risks of exposure to air pollution and how they can protect themselves and their families. By taking a comprehensive approach to reduce air pollution in Addis Ababa, it may be possible to improve air quality and promote public health and well-being in the city.

3. Improving Urban Air Quality Monitoring in a Developing Country Setting: The case of Ethiopia

3.1. Introduction

Despite the extensive links between ambient air pollution and adverse health impacts, there is a paucity of long-term, appropriately calibrated air quality data in low- and middle-income countries. This paucity of air quality information means that residents of these countries do not have sufficient information about the dangers of air pollution nor have the information to use appropriate measures to reduce exposure levels (Avis and Bartington., 2020). The World Health Organization (WHO) global ambient air quality database 2018 version reveals that there was only one PM2.5 ground-level monitor per 65 million people in low-income countries compared to one monitor per 370,000 people in high-income countries (WHO, 2022). Many Sub-Saharan African countries have no permanent measuring stations for PM2.5. In 2022, for example, the WHO ambient air quality database showed that only twelve countries in Africa have air quality monitoring sites primarily located in the capital city. Specifically, one air quality monitor per 28 million people in Sub-Saharan Africa (WHO, 2022). The United States Environmental Protection Agency (US-EPA) has established PM2.5 monitoring at US embassies and consulates in various countries, employing the Federal Equivalent Method (FEM) approved instrument Beta Attenuation Monitor (BAM-Met One 1020) and providing hourly PM2.5 measurements in 27 countries through their website (Singh et al., 2021). Among these countries is Ethiopia, where the capital Addis Ababa provides an example of this paucity of monitoring. Addis Ababa has two air quality

monitoring stations, one located in the parkland setting of the US Embassy installed in 2016, and the other site located at Addis Ababa international community school near Addis Ababa golf club in Lideta sub-city, which is mostly offline and does not record data(Dhammapala, 2019). Despite the installation of low-cost sensors by Addis Air and the United Nations Environment Program (UNEP) leading to a significant increase in the number of air quality monitoring stations, the majority of them remain offline and do not provide air pollution data to the public (WBG, 2021).[14]

While a fixed air quality monitor provides valuable time-series data, a single site cannot give a clear sense of the spatial distribution of air quality in a city. Moreover, depending on the nature of the measurement site, it may also give a biased picture of the exposure to particulate matter for the average resident. In low-income countries where the ability of governments to regulate pollution emissions and work conditions is limited due to weak state capacity, awareness of the public through public broadcasting of air quality will impact pollution avoidance behavior, indirectly reducing the health cost due to pollution.

The US embassies and consulates data has been used to study PM2.5 levels in the urban environment for a variety of purposes, including determining the trend and characteristics of PM2.5 in four mega cities of India by Chen et al. (2020) to examine the variability, trend, and exceedance of PM2.5 levels against WHO and Indian national air quality standards in five Indian megacities by Singh et al. (2021), to assess the association between PM2.5 and

---

[14] This lack of available data poses a significant challenge in addressing air pollution issues, as the public is unable to access real-time information about the quality of the air they breathe. The installation of these monitoring stations was a positive step toward addressing air pollution. Still, more work needs to be done to ensure that the data collected is accessible to the public.

daily outpatient visits for allergic rhinitis in Beijing, China, by Wang et al. (2020), to examine the association between air pollution exposure and gestational diabetes mellitus at trimester and weekly levels in Guangzhou, China a study by Zhang et al. (2020). Studies conducted using US embassies and consulates data so far have been focused on trend analysis of PM2.5 levels and health impact analysis using sparsely located PM2.5 monitors, one monitor for one major city.

Relying on fixed-site air pollution data that is located in a haphazard manner, particularly for polluted cities in developing countries is misleading and is unrepresentative of the actual exposure levels in a particular area for PM2.5. Strand et al. (2006) discovered that as pollution levels increase, the difference between ambient pollutant concentration and average personal exposure tends to rise. Furthermore, Nerriere et al. (2005) found inconsistencies between measurements of ambient air pollution obtained from personal multipollutant samplers and measurements from sparsely located fixed sites. Therefore, caution is necessary when conducting epidemiological studies on the long-term effects of ambient pollution.

This paper investigates the extent to which a single fixed air quality monitor (at the US Embassy) provides a guide to air quality and its risks across the city. To do this, we compared and contrasted data from a high-quality monitoring station based at the US Embassy in Addis Ababa with spatially rich PM2.5 data obtained using a low-cost mobile sensor across the city over a six-month period in 2020 and 2021. Furthermore, this study aims to identify the possible sources of discrepancy in air quality between measurements of the low-cost monitor and fixed air quality monitor.

The results indicate that depending on data from a single air quality monitor placed arbitrarily is inadequate for a comprehensive understanding of the complex characteristics of air pollution. To attain more accurate measurements in the future, we recommend a method for locating fixed air quality monitors that optimizes their placement by minimizing the total anticipated cost of inconsistencies.

This research illustrates the limitations of relying on a sole source of information and suggests using low-cost devices to guide policy interventions for air quality. Furthermore, we provided evidence on how this information could be improved by identifying the source of discrepancy between the two monitors reading and ways to improve air quality monitoring and public awareness in a developing country's major city. Furthermore, we suggest a method for placing fixed air quality monitors that reduces the expected cost discrepancy, considering the current position of existing monitors. By incorporating various factors such as cost (the cost of wrong advice or information), weather patterns and optimizing the placement of fixed monitors, we aim to increase the accuracy and precision of air quality monitoring systems in a developing country setting, thereby providing more comprehensive and actionable information to decision-makers and the public. Additionally, this approach can help to ensure that limited resources are used efficiently.

3.2. Description of the data

We used two main data sets for the PM2.5 profile in Addis Ababa: one from the US Embassy in Addis Ababa's AirNow website, which records hourly PM2.5 for 24 hours a day. [15]

---

[15] AirNow is a partnership of the U.S. Environmental Protection Agency, National Oceanic and Atmospheric Administration (NOAA), National Park Service, NASA, Centers for Disease Control, and tribal, state, and local

Appendix 3 contains details on time series data from the US Embassy in Addis Ababa from 2016 to 2022. The air quality status recorded by the US Embassy over the last six years shows moderate air quality in general, showing more polluted data early in the morning and evening.

The second data set contains PM2.5 data collected over a six-month period in 2020 and 2021 using a portable device-Dylos 1700 (which records PM2.5 every minute) at 72 sites across four inner sub-cities in Addis Ababa (this data set is the same as the data set, we used in the second chapter of this dissertation-refer part 2.4.1 of chapter 2 for details). The land use and weather data from chapter two are used again in this chapter. The land use data comes from the Addis Ababa municipality's land development and management office, which was published in 2015. National Aeronautics and Space Administration-National Oceanic and Atmospheric Administration (NASA-NOAA) and Addis Ababa Bole International Airport provided the weather data through their respective websites. NASA-NOAA weather data is updated hourly, whereas Addis Ababa International Airport data is updated every three hours.

In this section, we compare air quality data readings from the US Embassy in Addis Ababa fixed air quality monitor (located inside the US Embassy's parkland) with data gathered over six months at 72 different land use sites in Addis Ababa using a low-cost hand-held mobile device Dylos 1700. When we compare the readings from the two monitors during the data collection period (2020/21), a difference in the information about air quality and implied risk is noticed, which varies depending on the time of day, land use type, and wind direction. We used the US-EPA air quality classification and health risks described in Appendix 1, Table

---

air quality agencies. AirNow provides air quality data at US embassies and consulates around the world (https://www.airnow.gov/about-airnow/)

2.A.4, and WHO interim targets described in Table 3.A.4 to explain the discrepancy between the two monitors.

To better describe the data, the device was collocated in Addis Ababa in various land use types for calibration exercise, with the results shown in Table 3.3.[16] Details of checking the reliability of the portable device are explained in Appendix 2 (i).

In comparison Table 3.1 panel A, the percentages along the diagonal represent instances where observations from both monitors align. Conversely, disparities between the readings of the two monitors are highlighted off the diagonal. For instance, during the early morning, approximately 14% of the data recorded by the low-cost monitor indicates unhealthy air quality for sensitive individuals, whereas the fixed monitor registers the air quality as moderate. Notably, the cumulative values above the diagonal in the table exceed those below it, underscoring that the fixed air quality monitor documents less polluted levels of air, while the portable device indicates a higher level of air pollution.

Table 3.1 panel B presents the disparities and corresponding matched observations between the two monitors categorized by time of day. For instance, within the data collection period, during the early morning, out of the total monitored data early in the morning, 89% (84% after calibration) indicates air quality above the moderate threshold according to the low-cost monitor. In contrast, the US data records 41% of the data as exceeding the moderate air quality level. The most pronounced discrepancy occurs during the early morning and evening,

---

[16] In the description, from now onwards, we use the term calibrated, which is the result of the low-cost monitor PM2.5 adjustment using a collocation exercise conducted at an open area site

with the smallest overlap in the air quality category readings during these peak hours of the day.

During these peak hours, both monitors exhibit the lowest level of agreement in terms of recording the same air quality category. These findings underscore the significance of temporal patterns in the monitors' readings. This temporal variation in agreement could be attributed to factors such as atmospheric changes, temperature fluctuations, and human activity and land use differences.

Discrepancy and matched observations by land use type are presented in Appendix 2: Table 3.A.1(i) (before the low-cost monitor is calibrated) and Table 2. (ii) (after the low-cost monitor is calibrated). Overall discrepancy and matched observations after calibrating the low-cost monitor are presented in Appendix 2 Table 3.A.2.

Table 3.1: Overall discrepancy and discrepancy by the time of day(in percentage)

	Panel A(%)				
	Dylos↓				
% of total	Good	Moderate	Unhealthy for sensitive people	Unhealthy	Very unhealthy and a hazard
US→ Good	2	5	1	0	0
Moderate	3	40	14	11	1
Unhealthy for sensitive people	0	3	7	6	1
Unhealthy	0	0	0	3	0
Very unhealthy and a hazard	0	1	0	0	0

	Panel B (%)				
Time of day	Dylos raw Above moderate air quality	Dylos Calibrated Above moderate air quality	US above moderate air quality	US and Dylos raw in the same category	US and Dylos are calibrated in the same category.
Early morning	89	84	41	8	9
Late morning	38	32	21	33	33
Early afternoon	23	18	13	37	35
Late afternoon	27	21	10	19	18
Evening	87	85	32	4	4

3.2.1. Air quality warnings and thresholds

The World Health Organization (WHO) has set up three provisional objectives for PM2.5 that require consistent and persistent pollution reduction efforts to be achieved. These objectives are deemed helpful in tracking progress over time in the difficult task of progressively lowering public exposure to PM (WHO, 2022). This section compares the low-cost monitor PM2.5 reading with the measurement obtained from the US Embassy fixed monitor. We utilize the interim WHO benchmarks (Interim targets) for PM2.5 air quality, which are 37.5, 50, and 75 (micrograms per cubic meter), as a guide for air quality. Using these thresholds, we can evaluate the level of air quality and determine whether it falls within the acceptable range. It should be emphasized that PM2.5 concentrations at these levels negatively impact health and surpassing these levels have even more severe effects on public health, such as respiratory issues and cardiovascular diseases.

Panel A of Table 3.2 shows the percentage of times the two monitors exceed WHO interim target III by land use, Panel B shows the percentage of times the two monitors exceed WHO interim target III by the time of day, and Panel C shows the percentage of time the two monitors exceed interim target I, interim target II, and interim target III. According to the low-cost monitor, 41% (36% after calibration) of the data exceeded WHO interim target III, while the fixed monitor records only 17% of the data exceeding interim target III. Comparing the two monitors' readings by the time of day in terms of meeting WHO interim target III shows that more than 80% of the data recorded by the low-cost monitor early in the morning and evening exceeded the interim target, while the fixed monitor shows 65% of the recorded

data meets the target. In general, the US Embassy data understates the pollution levels in the city both in terms of meeting interim targets and in terms of US-EPA implied risk classification of a pollution level. Furthermore, the spatial variation of pollution across the city is not captured, as the US Embassy data shows only the temporal variation of pollution.

Table 3.2: PM2.5 profile in terms of meeting WHO interim targets

Panel A		
Percentage of times exceeding WHO 24-hour interim target III		
Land Use	PM2.5 Dylos raw	PM2.5 Dylos calibrated to the fixed monitor reading	US Embassy
Airport	65	45	15
Asphalt road	48	44	18
Mixed land	34	28	14
Open area	34	30	18
Panel B		
Percentage of times exceeding WHO 24-hour interim target III		
Time of day	PM2.5 Dylos raw	PM2.5 Dylos calibrated to the fixed monitor reading	US Embassy
Early morning	86	81	35
Late morning	34	27	15
Early afternoon	18	14	11
Late afternoon	23	18	4
Evening	87	84	30
Panel C		
Percentage of times exceeding WHO 24-hour interim target		
Monitor	Interim target III (37.5)	Interim target II (50)	Interim target I (75)
PM2.5-Dylos raw	41	26	12
PM2.5 Dylos calibrated	36	22	9
US Embassy	17	6	2
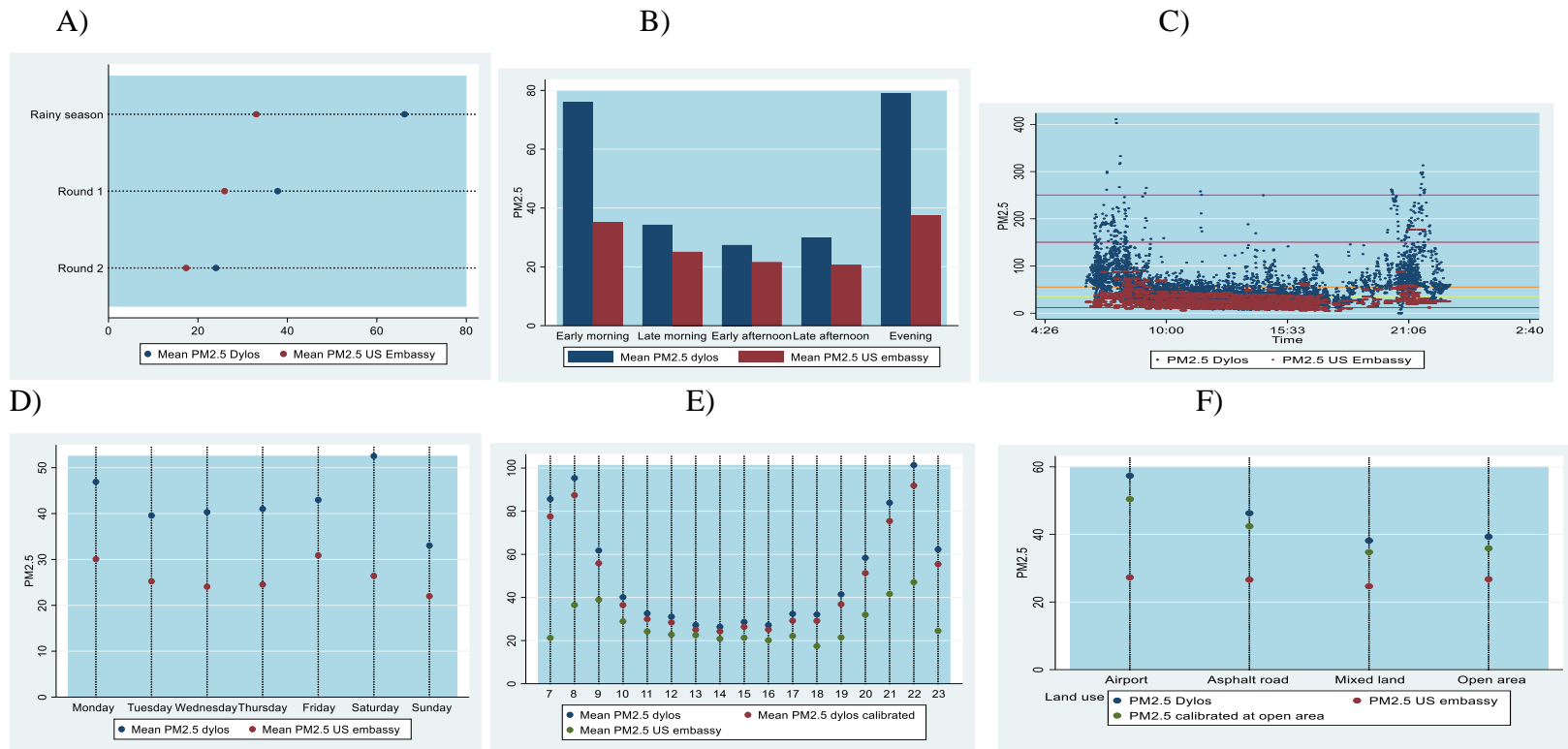
Note: A compares the mean PM2.5 values of the two monitors by data collection round; B, C, and E depicts PM2.5 of the two monitors by the time of day (colored reference lines in C are US-EPA air quality classifications: green for good, yellow for moderate, pink for WHO interim target III, orange for unhealthy for sensitive people); D compares the mean PM2.5 of the two monitors by day of the week; F compares the mean PM2.5 of days after and before calibration by land-use type and compares with US Embassy reading.

Figure 3.1: PM2.5 Profile

Research has been conducted on the importance of low-cost air quality monitors in reducing the cost of monitoring to provide the public with locally relevant pollution information (Maggos et al., 2019; Feenstra et al., 2019; Jova et al., 2015; Manikonda et al., 2016; Miskell et al., 2016; Popoola et al., 2018; Qin et al., 2020; Semple et al., 2013; Sousan et al., 2016; Wang et al., 2015; Williams et al., 2014; Lanphear, 2017). However, the accuracy of low-cost sensors when compared to a reference instrument is in question; as a result, efforts have been exerted to test and verify the accuracy of the low-cost monitors in laboratory and field experiments. In general, the performance of low-cost sensors was found to be in good agreement with high-grade fixed air quality monitors in laboratory experiments (where the environment is controlled) than in a complex field experiment with a range of environmental conditions (Liu et al., 2020).

On-the-field calibration of the portable monitor is performed in two settings to help explain the discrepancy between the two monitors (the US Embassy monitor and the low-cost monitor we used) in this study.[17] Device calibrations were conducted in Addis Ababa, where the research was conducted. The first calibration is carried out by placing the portable monitor near the fixed monitor on an asphalt road (approximately 20 meters away from the fixed monitor), and the second field experiment is carried out in Addis Ababa by placing the portable monitor on similar land use type to where the fixed air quality monitor is placed (at an open area land use-approximately 360 meters away from the fixed monitor).[18] A list of

---

[17] Measurement is conducted by collocating the portable monitor to the fixed monitor at the nearest possible place and approximately 1.5 meters above the ground.
[18] I was unable to enter the office of the US Embassy in Addis Ababa with an electronic device for security reasons of the Embassy; the closest location is the main asphalt road in front of the Embassy, where the portable device could be placed for validation.

previous calibration exercises and calibration exercise results we conducted are summarized

in Table 3.3. Based on model R^2, in correlation to the fixed monitor, the calibration result

conducted in a similar environment to the fixed monitor is chosen from all the field

experiments and literature survey results.

Table 3.3: Calibration results

Review of literature on calibration experiment		
Environment	Function	Source/comment
Indoor	$PM2.5 = 0.65 + 4.16 \times 10^{-5}[PNC] + 1.57 \times 10^{-11} \times PNC^2$	Semple et al. (2013)
Outdoor urban	$PM2.5 = 4.75 + 2.8 \times 10^{-5}[PNC]$	Steinle et al. (2015)
Outdoor rural	$PM2.5 = 1.29 + 1.11 \times 10^{-5}[PNC]$ **PNC** (Particle Number Count) =(Small-large) *0.01	Steinle et al. (2015)
Results from this study's validation exercises		
Open area near Embassy	PM2.5=5.4+0.92Dylos-0.658 Hum $\qquad R^2$=0.661 Correlation coefficient is 0.8	Calibrated values are highly correlated with the US Embassy data
Asphalt road near US Embassy	PM2.5=-0.41+0.86Dylos+0.75Hum-0.5DylosXHum R^2=0.5	Calibrated values are not strongly correlated with the US Embassy data
Use Dylos conversion formula as PM2.5, which means (Small-large) *0.01=Dylos		

Note: A summary of the literature review of Dylos 1700 is shown in the first three rows, and a summary of calibration results of the same low-cost monitor (Dylos 1700) monitor against the fixed monitor reading is shown in the last two rows The correlation between the fixed monitor and Dylos 1700 at the open area site is 0.8.

3.3. Estimating the discrepancy

What is the source of discrepancy between the readings from the fixed air quality monitor and the readings from the portable air quality monitor? We specify two models to identify the sources of a discrepancy more precisely: linear estimation of the difference model and multinomial logit model. First, we estimated the linear difference model with the dependent variable being the difference between the US Embassy reading and the low-cost monitor reading indicated in equation (1).

$$y_{it} = (PM2.5_{\,fixed} - PM2.5\,portable_{it})$$

$$= \rho + X_i\beta + W_t\gamma + T_t\mu + \alpha_0|LO_i - LO_{US\,embassy}| + \alpha_1|LA_i - LA_{US\,embassy}| + \varepsilon_{it} \qquad (1)$$

Where y_{it} represents the difference in readings between the two monitors for site i and time t, $PM2.5_{\,fixed}$ represents the reading from the US Embassy at time t since it is an hourly average, $PM2.5\,portable_{it}$ Denotes the reading from the portable device at site i and time t. X represents vector site characteristics-land use, W denotes a vector of weather variables at time t (wind speed, wind direction, temperature, humidity, rain with its one-day lag), whereas T denotes the time of day and day vector. $|LO_i - LO_{US\,embassy}|$, $|LA_i - LA_{US\,embassy}|$ denotes the absolute value of longitude and latitude difference between each site and the US Embassy to capture the relationship between discrepancy and location difference. As can be seen in the linear regression results (Table 3.4 for regression using area land use variables) and (Appendix 2:Table 3.A.3 for regression using distance land use variables), in both

regression results, geographic features play a systematic role in explaining the discrepancy. Furthermore, meteorological variables have also explained the discrepancy more precisely.

Comparing the results of the difference model using area land use and distance land use variables shows that distance land use variables (distance to the nearest land use from a site) have a more significant effect than area land use variables in explaining the discrepancy. In the second estimation method, we developed a multinomial dependent variable that primarily takes the form of a discrepancy and a no discrepancy in meeting the WHO interim target III. We used a multinomial logit model specified in equation (2) to identify the sources of discrepancy; the response variable (the dependent variable) takes four alternatives showing whether the information the two monitors provide is the same or has a discrepancy in terms of meeting WHO interim target III, indicated in Appendix 2 Table 3.A.4.

$$P(y=j/x) = \frac{e^{x\beta_j}}{1+\sum_{h=1}^{4} e^{x\beta_h}}, \quad j=1,2,3,4 \tag{2}$$

$$\begin{cases} j = 1, if \ US \ embassy \ reading < 37.5 \ and \ Portable \ monitor \ reading \ is < 37.5 \\ \quad j = 2, if \ US \ embassy \ reading < 37.5 \ and \ Portable \ monitor \ reading \geq 37.5 \\ \\ \quad j = 3, if \ US \ embassy \ reading \geq 37.5 \ and \ Portable \ monitor \ reading < 37.5 \\ \quad j = 4, if \ US \ embassy \ reading \geq 37.5 \ and \ Portable \ monitor \ reading \geq 37.5 \end{cases}$$

Table 3.5 shows the result of the estimated MNL using area land use variables, while Table 3.A.5 in Appendix 2 shows the estimated result of the MNL model using distance land use variables. The odds of discrepancy and the match are indicated in columns 1 and 2 (the base is the match, j=1).

The multinomial logit estimation results show that it is more likely for the low-cost monitor to record air quality information that exceeds WHO interim target III and the US Embassy to record air pollution level that doesn`t exceed WHO interim target III when rain increases if the monitoring site is closer to high-density mixed residence, river buffer, parking land use, commercial area, asphalt road, and unpaved roads if the wind blows from the south when the distance between the fixed monitor and the monitoring sites increases (latitude difference). When compared to other times of the day and days of the week, the low-cost monitor is likely to show air pollution levels that exceed WHO interim target III early in the morning and on Friday.

It is less likely for the low-cost monitor to report an air pollution level that exceeds WHO interim target III and the US Embassy to record an air pollution level that doesn`t exceed WHO interim target III when previous rain is higher, when wind speed increases, when the temperature increases, if the monitoring sites are open area sites and near to public facility land uses. For example, a study by Dhammapala (2019) showed that the difference between the US Embassy's air quality record and the other fixed site at a community school is lower when wind speed is higher than 20 km/h.

We can infer from both estimation methods that the difference in readings between the two monitors is primarily caused by land use and weather variables, with weather variables playing a more significant role. When we compare the distance land use variables to the area land use variables (in the linear regression and the MNL regression), the result shows that the distance land use variables play a more significant role in explaining the discrepancy between the two monitors' readings; this means that the distance to the pollution source is

more important in explaining the discrepancy than how much area the pollution source occupied. As a result, we conclude that the US Embassy data does not capture important aspects of air quality that vary throughout the day and by local land use features.

.

TABLE 3.4: Linear difference estimation results

VARIABLES	PM2.5 Difference (No calibration)	PM2.5 US Embassy	PM2.5 Difference (Calibrated)
Rain	-0.159	0.328	-0.138
	(0.568)	(0.394)	(0.537)
Lag rain	1.193	-1.253***	1.012
	(0.765)	(0.332)	(0.714)
Wind speed	-0.265	-1.258***	-0.352
	(0.490)	(0.397)	(0.458)
Humidity	-0.516	1.196***	0.242
	(0.550)	(0.376)	(0.514)
Temperature	0.696***	1.038***	0.708***
	(0.197)	(0.232)	(0.184)
High-density mixed residence	-0.177**		-0.168**
	(0.083)		(0.077)
River buffer	-0.506*		-0.482*
	(0.284)		(0.271)
Road network	-0.366		-0.324
	(0.260)		(0.242)
Open area	0.013		0.013
	(0.111)		(0.105)
Parking land use	0.231**		0.224**
	(0.098)		(0.091)
Commercial area	-0.180		-0.163
	(0.132)		(0.126)
Public facility	0.339**		0.312**
	(0.136)		(0.127)
Asphalt roads	-8.666**		-8.236**
	(3.648)		(3.417)
Unpaved roads	-6.894		-6.491
	(4.155)		(3.916)
North	-9.034**	3.895	-8.172**
	(3.927)	(2.676)	(3.767)
South	-5.490	0.498	-5.114
	(3.678)	(1.855)	(3.421)
West	-6.863*	-2.121	-6.627*
	(3.926)	(2.207)	(3.735)
Late morning	22.571***	-14.382***	19.820***
	(3.240)	(2.505)	(2.988)
Early afternoon	24.921***	-17.145***	21.801***
	(3.193)	(2.757)	(2.946)

Late afternoon	23.091***	-14.321***	20.276***
	(3.489)	(2.897)	(3.229)
Evening	-1.054	5.984	-0.789
	(5.363)	(4.360)	(5.116)
Monday	-3.175	2.106	-2.878
	(4.479)	(2.696)	(4.256)
Tuesday	4.058	-2.070	3.590
	(3.810)	(2.045)	(3.576)
Wednesday	0.310	-4.243	-0.031
	(3.426)	(2.879)	(3.240)
Thursday	4.115	-3.554	3.515
	(3.316)	(3.177)	(3.123)
Saturday	-5.743	-5.300	-5.757
	(3.984)	(3.817)	(3.778)
Sunday	1.238	-4.640	0.784
	(3.397)	(2.839)	(3.235)
Latitude difference	201.688***		187.012***
	(34.590)		(32.628)
Longitude difference	-67.303		-65.451
	(52.813)		(49.306)
Constant	-33.917***	11.204	-35.157***
	(11.380)	(9.780)	(10.561)
Observations	10,462	11,143	10,462
R-squared	0.30	0.26	0.27

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The dependent variable in the first column is the difference between the US Embassy PM2.5 reading and Dylos reading. The dependent variable in the second column is the US Embassy reading. The dependent variable in the third column is the difference between the US Embassy PM2.5 reading and the Calibrated PM2.5 reading. Land use explanatory variables are areas in meters within a one-hundred-square-meter radius of the sampling sites. Coefficients of land use variables and standard errors are multiplied by 1000/1000.

TABLE 3.5: Multinomial logit model estimation results

VARIABLES	MNL (base outcome=1)		
	P(y=2)	P(y=3)	P(y=4)
Rain	0.162***	0.834**	0.355***
	(0.022)	(0.139)	(0.031)
Lag rain	-0.093***	-0.262**	-0.575***
	(0.028)	(0.112)	(0.042)
Wind speed	-0.299***	-0.583***	-0.330***
	(0.021)	(0.076)	(0.030)
Humidity	0.030	-0.259*	0.01
	(0.025)	(0.136)	(0.035)
Temperature	-0.036**	0.318***	0.050**
	(0.014)	(0.061)	(0.021)
High-density mixed residence	0.009**	-0.005	-0.023***
	(0.004)	(0.011)	(0.006)
River buffer	0.052***	-0.315***	-0.027*
	(0.011)	(0.051)	(0.015)
Road network	0.008	-0.019	0.011
	(0.013)	(0.034)	(0.018)
Open area	-0.028***	-0.048***	-0.054***
	(0.005)	(0.013)	(0.007)
Parking land use	0.049***	0.124**	-0.594**
	(0.013)	(0.062)	(0.288)
Commercial area	0.039***	-0.130***	0.054***
	(0.007)	(0.034)	(0.009)
Public facility	-0.025***	-0.011	-0.045***
	(0.008)	(0.022)	(0.011)
Asphalt roads	1.397***	-0.813	0.751***
	(0.189)	(0.507)	(0.269)
Unpaved roads	0.364**	-4.933***	-0.647***
	(0.156)	(0.616)	(0.248)
North	0.022	-0.633**	0.090
	(0.190)	(0.317)	(0.193)
South	0.428**	-0.322	-0.373*
	(0.204)	(0.419)	(0.226)
West	-0.539***	-4.812***	-2.444***
	(0.197)	(0.700)	(0.223)
Late morning	-2.429***	-4.430***	-3.183***
	(0.119)	(0.389)	(0.152)
Early afternoon	-3.320***	-5.176***	-4.320***
	(0.134)	(0.425)	(0.177)
Late afternoon	-2.772***	-9.694***	-5.011***

	(0.127)	(1.125)	(0.223)
Evening	-0.673***	-0.232	-0.438**
	(0.161)	(0.474)	(0.195)
Monday	0.097	-1.845***	0.598***
	(0.128)	(0.305)	(0.146)
Tuesday	-0.132	-1.316***	-0.448***
	(0.122)	(0.324)	(0.159)
Wednesday	-0.151	-6.016***	-1.326***
	(0.117)	(0.790)	(0.160)
Thursday	-0.651***	-2.378***	-1.445***
	(0.137)	(0.383)	(0.183)
Saturday	-0.017	-21.970	-1.994***
	(0.115)	(1,155.107)	(0.170)
Sunday	-0.442***	-5.067***	-1.984***
	(0.116)	(0.602)	(0.182)
Latitude difference	-20.176***	-28.105***	-31.136***
	(1.700)	(5.029)	(2.439)
Longitude difference	23.577***	24.518***	3.841
	(2.485)	(9.309)	(3.837)
Constant	3.184***	3.645	4.471***
	(0.57)	(2.66)	(0.792)
Observations	10,462	10,462	10,462

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The base outcome is j=1(both monitors record WHO interim target III is met, P(y=2) {US Embassy<37.5 and Portable monitor≥37.5}, P(y=3) {US Embassy≥37.5 and Portable monitor >37.5}, P(y=4){US Embassy≥37.5 and Portable monitor ≥37.5}. Area land use variables (in a 100m buffer radius of a sampling site) are used in the regression. Coefficients of land use variables and standard errors are multiplied by 1000/1000.

3.4. Prediction results of the difference

In this part, we present the prediction results of the linear difference model: using a difference map depicted in Figure 3.3 and graphically in Figure 3.2. The difference map is created first by interpolating the predicted results through inverse distance weighting (IDW) using ArcGIS pro, whereby the nearest sites get more weight during interpolation. The map shows that the discrepancy varies by land use and time of day. Looking across maps shows discrepancies by the time of the day; the highest discrepancy is observed early in the morning and evening. Observing the discrepancy by land use type shows (within each map) the highest discrepancy at the center of the city, which means land use types influence the discrepancy. Figure 3.3 a and b show the actual and predicted mean difference PM2.5 by the time of the day, showing a higher mean discrepancy in the evening and early in the morning, which is also the case from the difference map. Figure 3.3 c shows the mean discrepancy by land use types of monitoring sites; the mean difference is highest at an asphalt road, and lower in mixed land and open area sites d shows the actual mean difference and predicted mean difference with and without calibrations. The bar graphs show that the mean predicted difference without calibration is closer to the actual mean difference.
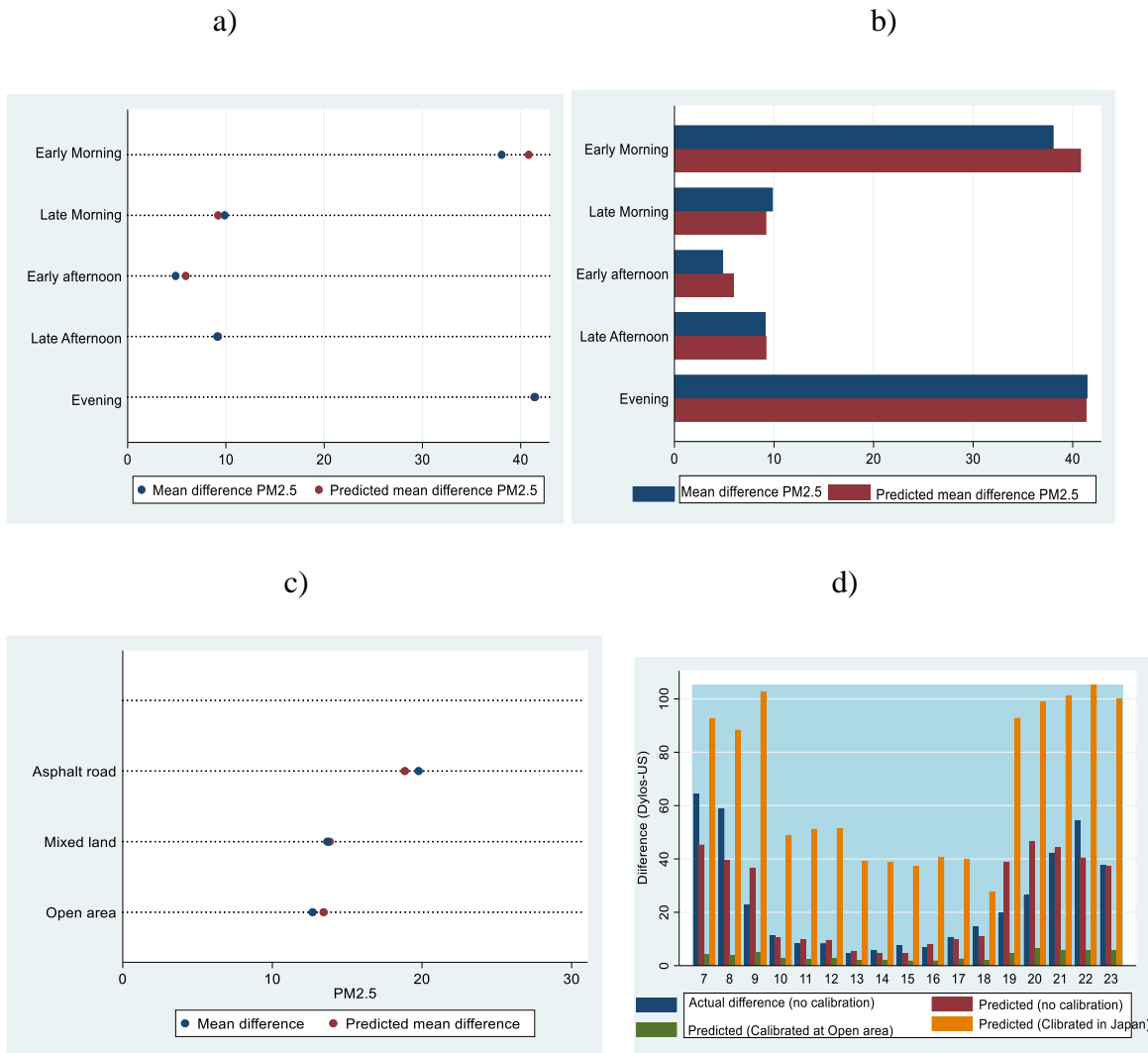
Note: Figure A and B show the mean difference between the low-cost monitor reading and the fixed monitor at the US Embassy by the time of day; Figure C shows the mean difference between the two monitors by land use, while Figure D shows the mean difference by the hour in a day during the sampling period.

Figure 3.2: Predicted difference PM2.5 results
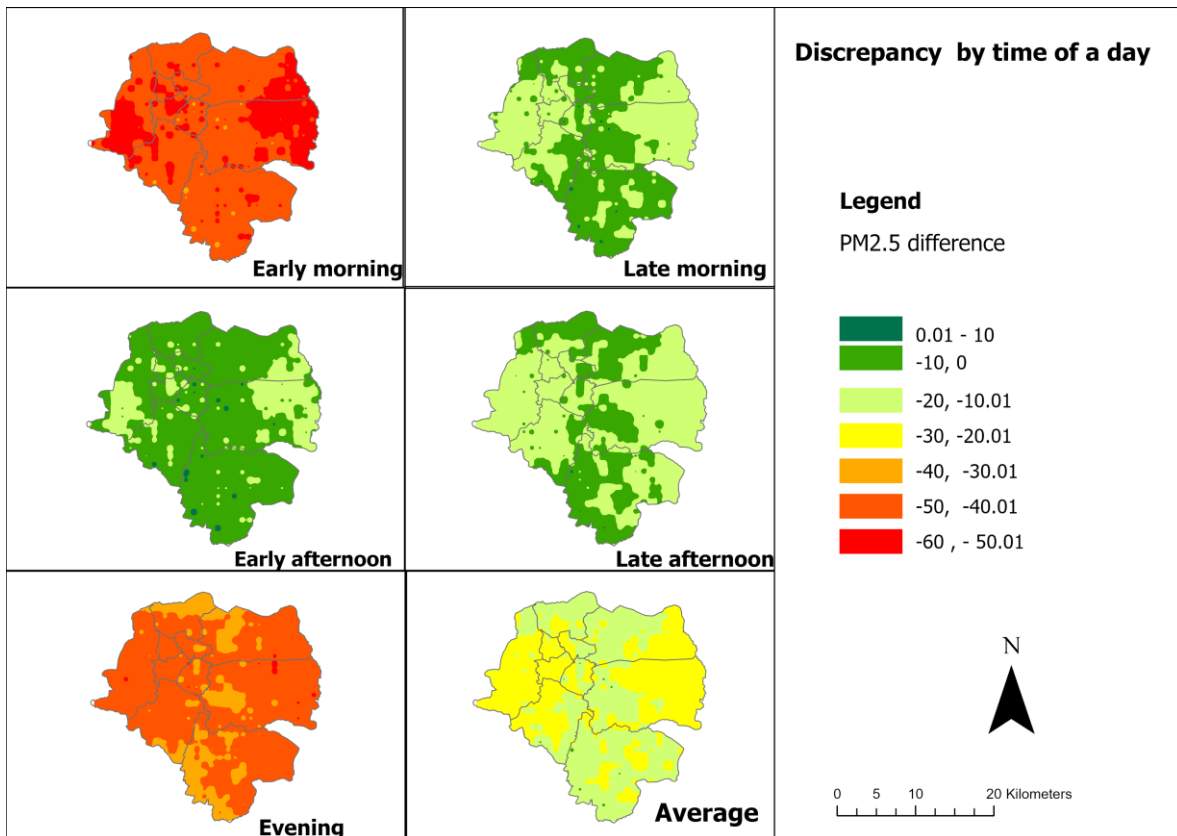
Note: IDW interpolation using the four nearest sites was used to get the map of pollution difference from the fixed monitor and the portable monitor by the time of day; the pollution levels were classified in a range of ten to show spatial and temporal variability in pollution difference

Source: Author`s illustration using ArcGIS Pro

Figure 3.3: Map of predicted air quality difference by the time of day
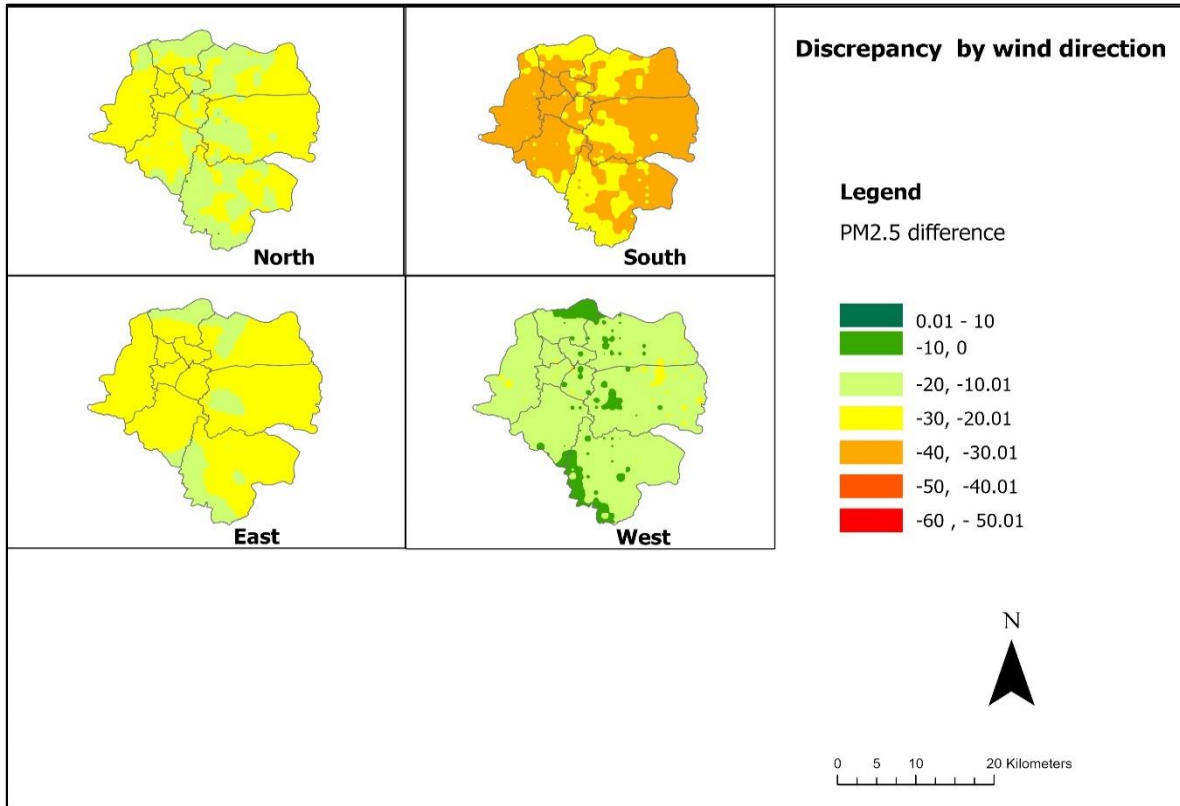
Note: IDW interpolation using the four nearest sites was used to get the map of pollution difference from the fixed monitor, and from the portable monitor by wind direction, the pollution levels were classified in a range of ten to show spatial and temporal variability in pollution difference

Source: Author`s illustration using ArcGIS Pro

Figure 3.4. Map of predicted air quality difference by wind direction

3.5. Suggesting optimal location for air quality monitor

The World Health Organization (WHO) does not have a specific recommendation for the number of air quality monitors used. Instead, they provide guidelines and recommendations for air quality monitoring programs considering the local context and the specific pollutants of concern. WHO recommends that air quality monitoring be conducted at locations representative of the population's exposure to air pollution. In general, this means that monitors should be placed where people spend a significant amount of time, such as in residential areas, schools, and workplaces. The number of monitors needed depends on factors such as the population's size, the area's geography, the sources of pollution, and the availability of resources. In general, it is recommended to have sufficient monitors to capture variations in air pollution levels across the area of interest. Ultimately, air quality monitoring aims to provide accurate and reliable information to inform policies and actions to reduce air pollution and protect public health (WHO, 2021).

Several approaches have been employed to identify the ideal locations for air monitoring stations. In a study by Kanaroglou et al. (2005), a location-allocation algorithm was utilized to identify the optimal sites for a specified number of air pollution monitors to enhance the network's efficiency. Despite its widespread use, this method has received criticism due to the potential for choosing redundant monitoring sites. Alsahli et al. (2018) used a suitability analysis to find a location for a fixed monitor. However, this method does not give the exact location for a fixed monitor. Another study by Munir et al. (2019) used population-weighted pollution concentration and weighted spatial variability that took into account factors such as population density (social indicator), air pollutant levels (environmental indicator), and spatial variability of air pollutant concentrations as the primary elements for determining the placement of air quality monitoring stations. However,

most of the research in this area has focused on cities or locations with a history of pollution data, which is often lacking in developing countries.

Our approach to the optimal location of a fixed air quality monitor is linked to the maximal covering location problem (MCLP) first developed by Church & Revelle (1972) see also Fazel et al. (2011) and ReVelle et al. (2008) where resources for facilities are scarce and the aim is to reach a maximum population coverage. The objective of the problem is to place facilities in eligible positions on a network so that the greatest possible population or other demand metric is met. A population is deemed covered if it is within a predetermined time or distance from one or more facilities. Over time, the problem has been adapted to a variety of scenarios.

To determine the placement of an additional fixed air quality monitor in Addis Ababa, we employed an optimization method to minimize the expected cost of discrepancy. Assume that both sites (the currently available fixed monitor and the proposed monitor) produce information and that everyone bases their decisions on the nearest monitor's information. Then, the further away they are from a monitor, the greatest the discrepancy, and the cost of this discrepancy increases in its square. In other words, the cost of wrong advice is assumed to be proportional to the square of the discrepancy; the farther people are from the monitor, the greater the discrepancy is; this is because air pollution is related to local land use in our discrepancy regression, so we aimed to select a site that would result in the minimum total cost.

In this study, we explored two scenarios for locating "new site":

1. Replicating the US embassy setup: through minimizing the total discrepancy in PM2.5 reading between each monitoring site and the US embassy fixed monitor, and

2. Positioning the proposed monitor in more typical conditions.

If we choose scenario (1), the model provided in appendix 2 (ii) will be applicable. On the other hand, if we opt for scenario (2), the discrepancy equation can be expressed as follows:

$$d_{it} = PM_{it} - PM_n \tag{4}$$

$$(PM_{it} - PM_{fixed}) - (PM_n - PM_{fixed}) = \rho + X_i\beta + W_t\gamma + T_t\mu + \alpha_0(LO_i - LO_{US\ embassy})\delta_{i0} + \alpha_1(LA_i - LA_{US\ embassy})\delta_{i1} - [\rho + X_n\beta + W_t\gamma + T_t\mu + \alpha_0(LO_n - LO_{US\ embassy})\delta_{i0} + \alpha_1(LA_n - LA_{US\ embassy})\delta_{i1}]$$

Where d_{it} refers to the discrepancy in PM2.5 reading between site i and n at time t, X_i and X_n is the site characteristics vector for site i and n (proposed monitor) respectively, W_t is the weather variables vector for time t, T_t is the time of day and week vector for time t. LO refers to longitude, i is the site's subscript, n is the subscript for the nearest fixed monitor, and LA refers to latitude. The deltas are either +1 or -1 and match the signs of the terms in parenthesis. In other words, we are using the absolute value difference of longitude and latitude in the discrepancy equation to pick up the effect of location difference as we used the absolute difference in the regression to show the effect of moving away from the fixed site.

$$d_{it} = \beta(X_i - X_n) + \alpha_0(LO_i - LO_n)\delta_{i0} + \alpha_1(LA_i - LA_n)\delta_{i1} \tag{5}$$

The discrepancy between site i and site n is just the difference in the land use impact plus the longitude and latitude terms difference. All the weather and time of day effects disappear. Hence, we aim to minimize the weighted sum of discrepancy squared with respect to an infinite small change in latitude and longitude of the proposed monitor.

$$\text{Min} \sum_{i \atop LO_n} \sum_{t \atop LA_n} \text{E}(d_{it}^2) = (\beta(X_i - X_n) + \alpha_0(LO_i - LO_n)\delta_{i0} + \alpha_1(LA_i - LA_n)\delta_{i1})^2 \tag{6}$$

As we move the longitude (LO_n) and latitude (LA_n) of the proposed site, we differentiate the above expression with respect to $LO_n\ and\ LA_n$, provided there is no location switches discrepancy changes and the change for a marginal change in latitude and longitude for a given time and location is given by the first order conditions in equations (7) and (8)

$$\frac{\partial \sum_i \sum_t \text{E}(d_{it}^2)}{\partial LO_n} = -2\alpha_0 \sum_i \sum_t \delta_{i0}(\beta(X_i - X_n) + \alpha_0(LO_i - LO_n)\delta_{i0} + \alpha_1(LA_i - LA_n)\delta_{i1}) = 0$$

$$\frac{\partial d_{it}^2}{\partial LO_n} = -2\alpha_0 \sum_i \sum_t \text{E}(\delta_{i0}d_{it}) = 0 \tag{7}$$

$$\frac{\partial \sum_i \sum_t \text{E}(d_{it}^2)}{\partial LA_n} = -2\alpha_0 \sum_i \sum_t \delta_{i0}(\beta(X_i - X_n) + \alpha_0(LO_i - LO_n)\delta_{i0} + \alpha_1(LA_i - LA_n)\delta_{i1}) = 0$$

$$\frac{\partial d_{it}^2}{\partial LA_n} = -2\alpha_1 \sum_i \sum_t \text{E}(\delta_{i1}d_{it}) = 0 \tag{8}$$

Given that the sampling sites were selected randomly, our aim is to identify the optimal location. To achieve this, we calculate the expected total discrepancy across all sites and relevant times. As the derivatives are linear, we proceed to compute the expected values of the derivatives across sites and time, as our objective is to minimize the expected discrepancy. Consequently, we arrive at the expected value of the derivatives.

$$\text{E}\left[\frac{\partial d_{it}^2}{\partial LO_n}\right] = \text{Min}\left[-2\alpha_0 \sum_i \sum_t \text{E}(\delta_{i0}d_{it})\right] \tag{9}$$

$$\text{E}\left[\frac{\partial d_{it}^2}{\partial LA_n}\right] = \text{Min}\left[-2\alpha_1 \sum_i \sum_t \text{E}(\delta_{i1}d_{it})\right] \tag{10}$$

3.5.1. Implementation of the Method

Assume that the new site has the mean land use of 72 sites from the four cities. We use this to find the optimal site out of the 72 squares in the four core sub-cities.

To find the optimal location, we take the expectation of total discrepancy across all sites. Since the effect of weather variables has been ruled out in the discrepancy equation, we are only capturing the effect of land use and location difference to find the optimal location of a representative air quality monitor.

The term $\beta(X_i - X_n)$ is obtained by taking the coefficient β from the discrepancy regression, and $(X_i - X_n)$ is a measure of land use that is obtained by subtracting the mean land use from the respective measure of land use for each site; hence $\beta(X_i - X_n)$ varies by site. This term captures the effect of land use types and location difference on the discrepancy. After obtaining all these, we weight $(\delta_{i0}d_{it})$ and $(\delta_{i1}d_{it})$ by the respective sites' sub-city population density, the weight of sub-city s is calculated as $(W_s = \frac{\text{Population density at sub-city } s}{\text{Population density of the Addis Ababa}} * \text{share of sub-}$ city in total area) and the weight for each site in a sub city s is given by

$W_{is} = \frac{W_s}{n_i}$ where n_i is the number of sites in a given sub city; see the calculation of weights in appendix 2 (ii). The weighted sum of expected discrepancy for a marginal change in longitude and latitude is given in equations 11 and 12.

The weighted sum of expected discrepancy for a marginal change in longitude

$$W_{is} E \left[\frac{\partial d_{it}^2}{\partial LO_n}\right] = W_{is} \{[-2\alpha_0 \sum_i \sum_t E(\delta_{i0}d_{it})]\} = -2\alpha_0 \sum_i \sum_t E(W_{is} \delta_{i0}d_{it}) \tag{11}$$

The weighted sum of expected discrepancy for a marginal change in latitude

$$W_{is} \, E \left[\frac{\partial d_{it}^2}{\partial LA_n}\right] = W_{is} \{[-2\alpha_1 \sum_i \sum_t E(\delta_{i1} d_{it})]\} = -2\alpha_1 \sum_i \sum_t E(W_{is}\delta_{i1} d_{it}) \qquad (12)$$

After obtaining all the calculated values, we adjust the latitude and longitude of the proposed site until the two weighted first-order conditions are close to zero or until it sets the discrepancy to the minimum possible; if the calculated d_{it} is positive, it means that LO_n and LA_n needs to be increased, if the calculated d_{it} is negative then LO_n and LA_n Needs to be reduced. Zero discrepancies are impossible with finite sites as the model does not capture the entire land use types. We adjust longitude and latitude until the first two order conditions are close to zero or the gap is minimum and locate the new site at the point where it sets the discrepancy to a minimum.

We used a trial and error to find the optimal location; we started with a reasonable guess; a location that splits the sampling sites into half makes some sense since we want to reduce the bad (wrong) advice from the monitor (the discrepancy) and the discrepancy is affected by land use; therefore we first took a location which is at the mid-point of the monitoring sites (study area) and started the location adjustment from that point. After doing the algorithm the final location that minimize the total sum of discrepancies found to be a public facility land in one of the four sub-cities of the study area. This sub-city is the fourth most densely populated sub-city See Figure 3.5.

Initially, we made the assumption that the ideal location would exhibit land use characteristics resembling the average of those observed among the 72 sites across the four cities. Now that we have determined the longitude and latitude of this optimal location for a representative monitor, the subsequent task involves developing the land use of the surrounding area in a manner that embodies the average land use characteristics found in the 72 sites, that is why we call it positioning it in a "typical condition".

Nevertheless, our proposition of a representative and stationary air quality monitoring approach stands as a viable choice for acquiring enduring pollution data. The potential for technological advancements in the future could offer cost-effective solutions for establishing multiple automated monitoring stations in urban areas. These stations would be furnished with both mobile/portable PM2.5 monitors and weather monitors, enabling continuous transmission of data through internet or mobile communication systems, particularly beneficial for developing nations. This, in turn, would empower policymakers and local authorities to make informed decisions regarding air quality regulations and interventions to mitigate pollution's adverse effects.

Moreover, the integration of mobile/portable PM2.5 monitors and weather sensors within these stations enhances the monitoring capability, allowing for a more holistic assessment of air quality conditions. This comprehensive data collection can help identify pollution hotspots, track pollution dispersion, and provide insights into the interplay between atmospheric conditions and pollutant levels. While challenges related to data management, maintenance, and calibration of monitoring equipment must be considered, the potential benefits of this approach are substantial. Continuous and accessible air quality data could drive public awareness, advocacy, and community engagement, fostering a collective effort to address air pollution issues on both local and global scales. In summary, the suggestion of establishing a fixed air quality monitoring station, combined with advancing technology, offers a promising avenue to gather enduring pollution data. This strategy has the potential to revolutionize air quality management by providing valuable insights for informed decision-making and driving initiatives to combat the detrimental impacts of air pollution in developing countries.

3.5.2. Sensitivity Check

To check the model sensitivity to changes in the model specification and measurement of land use we conducted three sensitivity checks.

1. The first sensitivity check involves utilizing land use measurements from unmonitored grid points. The outcome of this model suggests that the optimal location is situated in a short distance to the north of the main model suggested location.

2. The second sensitivity assessment gives greater importance to locations neighboring the four primary sub-cities within the designated study area. The suggested monitoring site, as determined by this approach, is positioned in the northwestern quadrant of the study area and exhibits a somewhat relatively pulled out placement when compared to other sensitivity check.

3. The third sensitivity assessment involves incorporating the average area of the two monitored sites within a grid. The result of this model indicates that the suggested monitoring site is positioned to the western side of the main model location (almost the same distance from the main model as from the first sensitivity check).

All sensitivity analysis demonstrates that the suggested monitoring site should be situated in relative proximity to the central area of the four primary core sub-cities, as depicted in Figure 3.6.
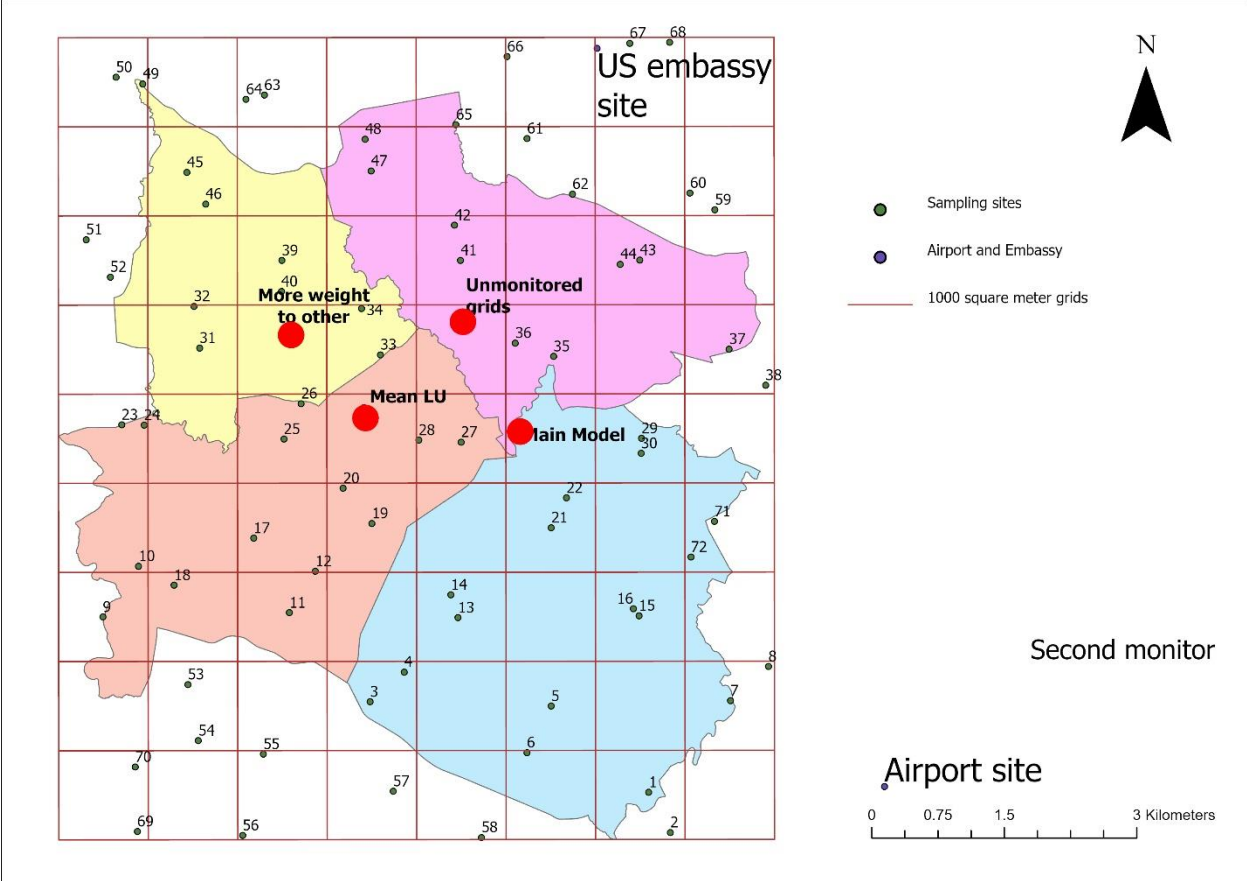
FIGURE 3.5 Optimal location for a fixed monitor

3.5.3. Limitations of the model

Despite its ability to give us the exact location for a fixed monitor, our model, however, has some weaknesses.

1. The impact of information away from sites is unknown in our model, and land use types of measurement may have a measurement error. Furthermore, land use measurement is not updated timely for developing countries, for example the land use data I used is the latest available for Addis Ababa which is the 2015, and land use has changed since then due to rapid change in the city infrastructure development.

2. Our model is quadratic and the solution we get is a Taylors approximation, which is correct locally and may give a result which is not exactly to the global optimum value.

3. Our cost function is quadratic, and we assumed that the cost of discrepancy increases in its square, but we do not have the data for the cost function, and it needs data to verify. Assume that the squared discrepancy is too high or too low than what it is (for example, the air quality outside is good), and someone about to exercise in the morning may decide to stay home than exercise. Is he better off staying home or exercising in polluted air? This needs data on the health benefit of exercising in polluted air or staying home with cleaner air but not exercising.

3.6. Conclusion

The focus of this chapter was to address three primary questions related to air pollution in cities of developing countries. The first objective was to examine whether a single fixed air quality monitor placed haphazardly can accurately reflect the air quality and its associated risks throughout the capital city of Ethiopia by comparing the air pollution data obtained from a low-cost handheld device. The second objective aimed to identify the reasons for the difference in pollution

information obtained from the fixed air quality monitor and the data collected using a low-cost monitor across the four main sub-cities of the Ethiopian capital. The third and final objective was to determine how to locate a fixed air quality monitor that could represent air pollution levels in the four sub-cities of Addis Ababa.

We discovered a correlation between PM2.5 values measured using a low-cost monitor in the local area of Addis Ababa and those recorded at the US Embassy (fixed site). However, significant differences exist between the two data sources, which are systematically related to local environmental features and weather conditions. Therefore, relying solely on a single fixed site placed in an ad-hoc and air quality data from that site may not be adequate for predicting PM2.5 and its risks in Addis Ababa. The air quality warnings generated from currently availed single fixed site data only provide limited guidance regarding local air quality conditions. Moreover, if low-cost devices are deemed accurate for estimating World Health Organization (WHO) interim levels of PM2.5, the guidance produced by the Embassy data is often inaccurate. Given the mortality and morbidity risks of poor air quality in developing countries, spatially rich data is valuable for policymaking. Our findings underscore the need for spatially rich and context-specific data in developing countries to inform policymaking and public health initiatives related to air quality. By adopting a multi-faceted approach incorporating locally collected air quality data from low-cost monitors and data from fixed air quality monitors located at US embassies, policymakers can better understand the complex factors influencing PM2.5 levels and take more targeted actions to protect public health.

Although mobile monitoring data provides a cost-effective means of identifying spatial variations in ambient particulate matter across vast geographical areas, it is essential to acknowledge certain limitations of this study. First, the data collection period was short and did not capture seasonal

fluctuations. Additionally, all data was collected during daytime hours, meaning the model may overestimate ambient concentrations as nighttime levels are typically lower. However, since our aim here is to identify exposure variability, this may not be a big concern.

The method proposed to identify the optimal locations for air quality monitors is (i) based on the expected sum of weighted discrepancies and (ii) by considering the available fixed monitor. This method can enhance the placement of monitors in areas where they are scarce, particularly in developing countries major cities like Addis Ababa. In addition, this approach can improve future air quality readings by providing critical information for public health and policymaking.

This study sheds light on the challenges associated with air quality monitoring in developing countries and offers potential solutions for improving the accessibility of reliable air quality information. The proposed method here represents a promising approach for optimizing the placement of air quality monitors in resource-limited settings, where costly traditional fixed-site monitoring may be impractical on a large scale. By expanding access to reliable air quality information, this approach can potentially improve public health outcomes and inform evidence-based policy decisions.

Furthermore, the study results highlight the importance of considering the current location of fixed monitors when determining the optimal placement of air quality monitors. Considering the weighted sum of expected discrepancies, the proposed method can optimize the efficiency of limited monitoring resources when installing new fixed monitors. Overall, this study provides valuable insights into the challenges of air quality monitoring in developing countries and offers a practical solution for improving the accuracy and accessibility of air quality information. Optimizing the placement of air quality monitors can enable policymakers to make more informed

decisions to reduce exposure to air pollution, protect public health, and promote sustainable development.

CHAPTER FOUR

4. Concluding remarks and policy implications

4.1. Concluding remarks

Air pollution is responsible for causing illness and premature death in millions of individuals worldwide, especially in low- and middle-income countries (LMICs). Many of the world's most populated and polluted cities are in LMICs. Despite this, people living in LMICs are often unaware of the severity of air pollution risks or lack the information necessary to tackle the issue. Several factors contribute to this situation, including inadequate or missing measurement data, limited data accessibility, and ineffective communication strategies. In numerous instances, individuals in LMICs are not cognizant that air pollution is the cause of sickness in themselves and their children, and they do not know how to protect themselves. Air pollution levels in LMIC cities are significantly higher than those in high-income countries, which makes it challenging to alter behaviors and avoid exposure without implementing strict national regulations to curb air pollution. Additionally, addressing air pollution in LMICs is complicated by weak or nonexistent monitoring programs and scientific institutions.

This dissertation has four main objectives related to assessing air pollution and mapping air quality in a developing country setting by taking the case of Ethiopia. The first objective was to develop a low-cost method for assessing air quality using a land use regression model to predict pollution levels in unmonitored areas. The second objective aimed to evaluate the accuracy of a single fixed air quality monitor placed randomly in Ethiopian capital in reflecting air quality and its associated

risks by comparing it with data collected from a low-cost handheld device. The third objective was to identify the reasons for the disparity in pollution information obtained from the single fixed air quality monitor and the data gathered from a low-cost monitor in the four main sub-cities of Addis Ababa. Finally, the fourth objective was to determine the optimal location for a fixed air quality monitor that could represent air pollution levels in the four sub-cities of Addis Ababa.

The first paper of this dissertation involved collecting air quality data in Addis Ababa, the capital of Ethiopia, using a low-cost mobile air quality monitor and creating an air pollution exposure map for unmonitored locations by utilizing a land use regression (LUR) model to relate pollution levels to spatial land use data and weather variables. Field data were collected during dry and rainy seasons, covering a range of land uses at predetermined routes across four sub-cities of Addis Ababa in 2020/21. Each site was visited eight times for 15-20 minutes, with a Dylos 1700 air quality monitor at the four sub-cities in the center of Addis Ababa. The LUR model results showed that PM2.5 concentration levels were sensitive to local land use, traffic, and climatic conditions. Concentrations of PM2.5 were higher in commercial areas and areas with dense networks of roads, indicating the crucial role of land use types in predicting particulate matter concentration. Additionally, meteorological variables had a more significant impact on concentration levels. This modeling approach can be extended to other urban locations in developing countries that lack air quality monitors.

In the second paper of this dissertation, we examined the extent of correlation between PM2.5 levels measured by a low-cost monitor in the four sub-cities of Addis Ababa and those recorded at the US Embassy (fixed site). We discovered significant differences between the two data sources systematically related to local environmental features and weather conditions. Therefore, relying solely on a single fixed site placed in an ad-hoc and air quality data from that site may not be

adequate to inform the PM2.5 risks in Addis Ababa. The air quality warnings generated from the currently availed single fixed site data only provide limited guidance for local air quality conditions. Moreover, if the low-cost device is deemed accurate for World Health Organization (WHO) interim levels of PM2.5, the guidance produced by the Embassy data is often inaccurate. Given the mortality and morbidity risks associated with poor air quality in developing countries, spatially rich data is valuable for policymaking.

The method proposed to determine the optimal locations for air quality monitors based on the weighted sum of expected discrepancies and by considering the available fixed monitor can enhance the placement of monitors in areas where they are scarce, particularly in developing countries like Addis Ababa. By doing so, the accuracy of future air quality readings can be improved, providing critical information for public health and policymaking. The study sheds light on the challenges associated with air quality monitoring in developing countries and offers potential solutions for improving the accessibility of reliable air quality information. The proposed method represents a promising approach for optimizing the placement of air quality monitors in resource-limited settings, where traditional fixed-site monitoring may be impractical on a large scale as they are costly. By expanding access to reliable air quality information, this approach can potentially improve public health outcomes and inform evidence-based policy decisions.

4.2. Policy Implication

The results of this dissertation suggest the following policy implications be considered by policymakers and air quality monitoring program agencies.

4.2.1. Cost-effective spatially rich and context-specific data

Our research underscores the need for spatially rich and context-specific data to inform policymaking and public health initiatives related to air quality in developing countries. By

adopting a multi-faceted approach that incorporates both locally collected air quality data through low-cost monitors and data from fixed air quality monitors usually located at US embassies in many developing countries, policymakers can assess air quality status in the cities and the associated risks.

4.2.2. Opportunistic approach to air pollution exposure assessment

Policy makers could consider developing policies that employ an opportunistic approach to collect data from diverse segments of society and the local environment to monitor exposure levels. For instance, utilizing public transportation to gauge exposure levels during travel and engaging street market vendors to evaluate exposure levels in busy local markets can be effective methods. By gathering data through such opportunistic methods, policymakers can better understand the several factors that contribute to exposure levels in different settings. This information can then be used to develop targeted policies that address specific sources of exposure and mitigate potential health risks. Moreover, involving distinct groups of society and the local context in data collection can promote community engagement and foster a sense of ownership and responsibility toward addressing environmental health issues. This can also help identify and address unique concerns and challenges specific to different communities. Overall, designing policies incorporating opportunistic data collection methods and involving diverse groups of society and the local context can be a valuable approach for policymakers to monitor exposure levels and develop effective strategies to promote environmental health.

4.2.3. Adopting the method of optimal location, which considers the current available fixed monitor.

The proposed method of using a weighted sum of expected discrepancies and considering the available fixed monitor can significantly enhance the placement of air quality monitors, especially in areas where they are scarce, such as Addis Ababa. Policymakers may wish to consider adopting this method when installing additional fixed monitors to find the optimal location.

This method can help improve the accuracy of future air quality readings, providing critical information for public health and policymaking. With accurate data, policymakers can make informed decisions to address air pollution and implement targeted interventions to improve air quality in the identified areas. Furthermore, this method can help save costs associated with monitoring air quality, as it ensures that monitoring is conducted in the most effective locations. This is particularly important in resource-constrained settings, where the cost of monitoring can be a significant challenge. In summary, the proposed method of determining optimal locations for air quality monitors can be an essential tool for policymakers in developing countries like Addis Ababa. By adopting this method, policymakers can improve the accuracy of air quality readings and make informed decisions to address air pollution, promoting public health and environmental sustainability.

4.3. Future research directions

This dissertation highlights several future research directions related to air pollution in developing countries. One potential area for future research is the development and implementation of effective communication strategies to increase awareness and understanding of air pollution risks among the public in low- and middle-income countries and major cities like Addis Ababa. This could involve identifying the most effective channels for communicating information about air pollution.

Another potential research direction is the development of cost-effective and accurate air quality monitoring methods for resource-limited settings. The study presented in this dissertation suggests that land use regression models and low-cost handheld monitors may be effective tools for assessing air quality in areas without fixed monitoring stations. Future research could explore opportunistic methods of data campaigns that involve distinct groups of society to monitor exposure levels; this might involve both indoor and/or outdoor pollution.

Finally, future research could explore the links between air pollution and land use and land cover changes, for example in developing countries, major cities like Addis Ababa, land use change is so frequent from cleaner public land use (e.g., green areas and open spaces into a more polluted and high return commercial and residential areas).

References

AA EP. (2021). Addis Ababa air quality managment plan. *Working Paper*. https://www.epa.gov/system/files/documents/2021-11/final-aqmp-addis-ababa.pdf

AfDB. (2021). Addis Ababa " Beautifying Sheger " River Development Project. *AfDB*. https://www.afdb.org/sites/default/files/2022/07/04/web_sheger_river_report.pdf

Alsahli, M. M., & Al-Harbi, M. (2018). Allocating optimum sites for air quality monitoring stations using GIS suitability analysis. *Urban Climate*, *24*(August), 875–886. https://doi.org/10.1016/j.uclim.2017.11.001

Amini, H., Taghavi-Shahri, S. M., Henderson, S. B., Naddafi, K., Nabizadeh, R., & Yunesian, M. (2014). Land use regression models to estimate the annual and seasonal spatial variability of sulfur dioxide and particulate matter in Tehran, Iran. *Science of the Total Environment*, *488–489*(1), 343–353. https://doi.org/10.1016/j.scitotenv.2014.04.106

Avis, W., & Bartington, S. (2020). *Monitoring Air Quality in Low- Income and Lower Middle-Income Countries*.

Cai, J., Ge, Y., Li, H., Yang, C., Liu, C., Meng, X., & Wang, W. (2020). Application of land use regression to assess exposure and identify potential sources in PM2.5, BC, NO2 concentrations. *Atmospheric Environment*, 117267. https://doi.org/10.1016/j.atmosenv.2020.117267

Cao, J., Cheng, Y., & Yu, C. (2016). Urban air quality management in Ethiopia. *Centre for Science and Environment Material*, *27*(1), 3–6. https://doi.org/10.1177/1420326X17742007

Chalermpong, S., Thaithatkul, P., Anuchitchanchai, O., & Sanghatawatana, P. (2021). Land use regression modeling for fine particulate matters in Bangkok, Thailand, using time-variant predictors: Effects of seasonal factors, open biomass burning, and traffic-related factors. *Atmospheric Environment*, *246*(December 2020), 118128. https://doi.org/10.1016/j.atmosenv.2020.118128

Chen, L., Bai, Z., Kong, S., Han, B., & You, Y. (2010a). A land use regression for predicting NO 2 and PM 10 concentrations in di ff erent seasons in Tianjin region , China. *Journal of Environmental Sciences*, *22*(9), 1364–1373. https://doi.org/10.1016/S1001-0742(09)60263-1

Chen, L., Bai, Z., Kong, S., Han, B., & You, Y. (2010b). A land use regression for predicting NO2 and PM10 concentrations in different seasons in Tianjin region, China. *ScienceDirect*, *22*(2), 1364–1373.

Chen, L., Shi, M., Gao, S., Li, S., Mao, J., Zhang, H., Sun, Y., Bai, Z., & Wang, Z. (2017). Assessment of population exposure to PM2.5 for mortality in China and its public health benefit based on BenMAP. *Environmental Pollution*, *221*, 311–317. https://doi.org/10.1016/j.envpol.2016.11.080

Chen, Y., Wild, O., Conibear, L., Ran, L., He, J., Wang, L., & Wang, Y. (2020). Atmospheric Environment : X Local characteristics of and exposure to fine particulate matter (PM 2 . 5) in four indian megacities. *Atmospheric Environment: X*, *5*(October 2019), 100052. https://doi.org/10.1016/j.aeaoa.2019.100052

Church, R., & Revelle, C. (1972). THE M A X I M A L COVERING LOCATION PROBLEM. *The Johns Hopkins University*, *6*(6).

Coker, E. S., Amegah, A. K., Mwebaze, E., Ssematimba, J., & Bainomugisha, E. (2021). A land use regression model using machine learning and locally developed low cost particulate matter sensors in Uganda. *Environmental Research*, *199*(December 2020), 111352. https://doi.org/10.1016/j.envres.2021.111352

Concentrations, M., Maggos, T., Stamatelopoulou, A., & Loh, M. (2019). *Comparison of Methods for Converting Dylos Particle Number Concentrations Comparison of methods for converting Dylos particle number concentrations to PM2 . 5 mass concentrations. April 2020.* https://doi.org/10.1111/ina.12546

Dacuntoa, , Neil E. Klepeisa, c, Kai-Chung Chenga, Viviana Acevedo-Boltona, R.-. (2015). Determining PM2.5 calibration curves for a low-cost particle monitor: Common indoor residential aerosols. *Environmental Science*, *17*.

Dhammapala, R. (2019). Analysis of fine particle pollution data measured at 29 US diplomatic posts worldwide. *Atmospheric Environment*, *213*, 367–376. https://doi.org/10.1016/j.atmosenv.2019.05.070

Dong, J., Ma, R., Cai, P., Liu, P., Yue, H., Zhang, X., Xu, Q., Li, R., & Song, X. (2021). Effect of sample number and location on accuracy of land use regression model in NO2 prediction. *Atmospheric Environment*, *246*(October), 118057. https://doi.org/10.1016/j.atmosenv.2020.118057

Eeftens, M., Beelen, R., Hoogh, K. De, Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., De, A., Dons, E., Nazelle, A. De, Dimakopoulou, K., Eriksen, K., Fischer, P., Galassi, C., Graz, R., Heinrich, J., Ho, B., Jerrett, M., Keidel, D., … Hoek, G. (2012). *Development of Land Use Regression Models for PM 2 . 5 , PM 2 . 5 Absorbance , PM 10 and PM coarse in 20 European Study Areas ; Results of the ESCAPE Project.*

EEPA. (2003). *GUIDLINE AMBIENT ENVIRONMENT STANDARDS FOR ETHIOPIA*. *August*.

EPA, U., & Quality Management Division, E. (2002). Q u a Q u a l i t y l i t y Guidance for Quality Assurance Project Plans EPA QA/G-5. *USEPA*. https://doi.org/https://www.epa.gov/sites/default/files/2015-06/documents/g5-final.pdf

Fazel Zarandi, M. H., Davari, S., & Haddad Sisakht, S. A. (2011). The large scale maximal covering location problem. *Scientia Iranica*, *18*(6), 1564–1570. https://doi.org/10.1016/j.scient.2011.11.008

Feenstra, B., Papapostolou, V., Hasheminassab, S., & Zhang, H. (2019). Performance evaluation of twelve low-cost PM2.5 sensors at an ambient air monitoring site. *Atmospheric Environment*, 116946. https://doi.org/10.1016/j.atmosenv.2019.116946

Han, I., Symanski, E., & Stock, T. H. (2017). Feasibility of using low-cost portable particle monitors for measurement of fine and coarse particulate matter in urban ambient air. *Journal of the Air & Waste Management Association*, *67*(3), 330–340. https://doi.org/10.1080/10962247.2016.1241195

Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J., & Thiele, L. (2015). Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing*, *16*(PB), 268–285. https://doi.org/10.1016/j.pmcj.2014.11.008

He, B., Heal, M. R., & Reis, S. (2018). Land-use regression modelling of intra-urban air pollution variation in China: Current status and future needs. *Atmosphere*, *9*(4), 1–19. https://doi.org/10.3390/atmos9040134

Ho, T. H., Van Dang, C., Pham, T. T. B., & Wangwongwatana, S. (2023). Assessment of health and economic benefits of reducing fine particulate matter (PM2.5) concentration in Ho Chi Minh City, Vietnam. *Hygiene and Environmental Health Advances*, *6*(August 2022), 100045. https://doi.org/10.1016/j.heha.2023.100045

Hoek, G., Beelen, R., Hoogh, K. De, Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution.

Atmospheric *Environment*, *42*(33), 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057

Jin, L., Berman, J. D., Zhang, Y., Thurston, G., Zhang, Y., & Bell, M. L. (2019). Land use regression study in Lanzhou , China : A pilot sampling and spatial characteristics of pilot sampling sites. *Atmospheric* *Environment*, *210*(October 2018), 253–262. https://doi.org/10.1016/j.atmosenv.2019.02.043

Joris, V. D. B. (2016). *Towards high spatial resolution air quality mapping : a methodology to assess street-level exposure based on mobile monitoring*. Ghent university.

Jova, M., Lazovi, I., & Pokri, B. (2015). *On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter **. *206*, 696–704. https://doi.org/10.1016/j.envpol.2015.08.035

Kanaroglou, P. S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L., & Brook, J. R. (2005). Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmospheric Environment*, *39*(13), 2399–2409. https://doi.org/10.1016/j.atmosenv.2004.06.049

Kumie, A., Worku, A., Tazu, Z., Tefera, W., Asfaw, A., Boja, G., Mekashu, M., Siraw, D., Teferra, S., Zacharias, K., Patz, J., Samet, J., & Berhane, K. (2021). Fine particulate pollution concentration in Addis Ababa exceeds the WHO guideline value: Results of 3 years of continuous monitoring and health impact assessment. *Environmental Epidemiology*, *5*(3), 1–9. https://doi.org/10.1097/EE9.0000000000000155

Li, Z., Ho, K., Hung, S., & Yim, L. (2020). Development and intercity transferability of land-use regression models for predicting ambient PM 10 , PM 2 . 5 , NO 2 and O 3 concentrations in northern Taiwan. *Atmospheric Chemistry and Physics*, *2*.

Lim, C. C., Kim, H., Vilcassim, M. J. R., Thurston, G. D., Gordon, T., Chen, L. C., Lee, K., Heimbinder, M., & Kim, S. Y. (2019). Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environment International*, *131*(July), 105022. https://doi.org/10.1016/j.envint.2019.105022

Liu, W., Li, X., Chen, Z., Zeng, G., Liang, J., Huang, G., Gao, Z., Jiao, S., He, X., & Lai, M. (2015). *Land use regression models coupled with meteorology to model spatial and temporal variability of NO 2 and PM 10 in Changsha , China s Le o n d ,. 116*(2), 272–280. https://doi.org/10.1016/j.atmosenv.2015.06.056

Liu, X., Jayaratne, R., Thai, P., Kuhn, T., Zing, I., Christensen, B., Lamont, R., Dunbabin, M., Zhu, S., Gao, J., Wainwright, D., Neale, D., Kan, R., Kirkwood, J., & Morawska, L. (2020). Low-cost sensors as an alternative for long-term air quality monitoring. *Environmental Research*, *185*(March), 109438. https://doi.org/10.1016/j.envres.2020.109438

Luke, D. (2017). *Development of a land use regression model for daily NO 2 and NO x concentrations in the Brisbane metropolitan area , Australia*. 0–25.

Maggos, T., Stamatelopoulou, A., & Loh, M. (2019). *Comparison of Methods for Converting Dylos Particle Number Concentrations Comparison of methods for converting Dylos particle number concentrations to PM2 . 5 mass concentrations*. *April 2020*. https://doi.org/10.1111/ina.12546

Manikonda, A., Zíková, N., Hopke, P. K., & Ferro, A. R. (2016). Laboratory assessment of low-cost PM monitors. *Journal of Aerosol Science*, *102*, 29–40. https://doi.org/10.1016/j.jaerosci.2016.08.010

Masiol, M., Zíková, N., Chalupa, D. C., Rich, D. Q., Ferro, A. R., & Hopke, P. K. (2018). Hourly land-use regression models based on low-cost PM monitor data. *Environmental Research*,

167(June), 7–14. https://doi.org/10.1016/j.envres.2018.06.052

Merbitz, H., Fritz, S., & Schneider, C. (2012). Science of the Total Environment Mobile measurements and regression modeling of the spatial particulate matter variability in an urban area. *Science of the Total Environment, The*, *438*, 389–403. https://doi.org/10.1016/j.scitotenv.2012.08.049

Miskell, G., Salmond, J., & Williams, D. E. (2016). Science of the Total Environment Low-cost sensors and crowd-sourced data : Observations of siting impacts on a network of air-quality instruments. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2016.09.177

Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M., & Fishbain, B. (2015). Science of the Total Environment On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of the Total Environment, The*, *502*, 537–547. https://doi.org/10.1016/j.scitotenv.2014.09.059

Munir, S., May, M., Coca, D., & Jubb, S. A. (2019). Atmospheric Environment : X Structuring an integrated air quality monitoring network in large urban areas – Discussing the purpose , criteria and deployment strategy. *Elsevier*, *2*(March). https://doi.org/10.1016/j.aeaoa.2019.100027

Pope, F. D., Gatari, M., Ng, D., Poynter, A., & Blake, R. (2018). Airborne particulate matter monitoring in Kenya using calibrated low-cost sensors. *Atmospheric Chemistry and Physics*, 15403–15418.

Popoola, O. A. M., Carruthers, D., Lad, C., Bright, V. B., Mead, M. I., Stettler, M. E. J., Sa, J. R., & Jones, R. L. (2018). *Use of networks of low cost air quality sensors to quantify air quality in urban settings*. *194*(September), 58–70. https://doi.org/10.1016/j.atmosenv.2018.09.030

Qin, X., Hou, L., Gao, J., & Si, S. (2020). Science of the Total Environment The evaluation and optimization of calibration methods for low-cost particulate matter sensors : Inter-comparison between fi xed and mobile methods. *Science of the Total Environment*, *715*, 136791. https://doi.org/10.1016/j.scitotenv.2020.136791

Qu, Z., Wang, X., Li, F., Li, Y., Chen, X., & Chen, M. (2020). Pm2.5-related health economic benefits evaluation based on air improvement action plan in wuhan city, middle china. *International Journal of Environmental Research and Public Health*, *17*(2). https://doi.org/10.3390/ijerph17020620

Rai, A.C., Kumar, P., Pilla, F., Skouloudis, A.N., Di Sabatino, S., Ratti, C., Yasar, A. and Rickerby, D.,(2017). End-user perspective of low-cost sensors for outdoor air pollution monitoring. Science of The Total Environment, 607, pp.691-705

ReVelle, C., Scholssberg, M., & Williams, J. (2008). Solving the maximal covering location problem with heuristic concentration. *Computers and Operations Research*, *35*(2), 427–435. https://doi.org/10.1016/j.cor.2006.03.007

Rivera, M., Basagaña, X., Aguilera, I., Agis, D., Bouso, L., Foraster, M., Medina-ramón, M., Pey, J., Künzli, N., & Hoek, G. (2012). Spatial distribution of ultra fi ne particles in urban settings : A land use regression model. *Atmospheric Environment*, *54*, 657–666. https://doi.org/10.1016/j.atmosenv.2012.01.058

Roy, R. (2016). "The cost of air pollution in Africa", OECD development centre working paper. *OECD Publishing, Paris*, *333*, 56. www.oecd.org/dev/wp.%0A©OECD

Schlenker, W., & Walker, W. R. (2016). Airports, air pollution, and contemporaneous health. *Review of Economic Studies*, *83*(2), 768–809. https://doi.org/10.1093/restud/rdv043

Semple, S., Ibrahim, A. E., Apsley, A., Steiner, M., & Turner, S. (2013). *Using a new , low-cost*

air quality sensor to quantify second-hand smoke (SHS) levels in homes. 1–6. https://doi.org/10.1136/tobaccocontrol-2013-051188

Singh, V., Singh, S., & Biswal, A. (2021). Exceedances and trends of particulate matter (PM2.5) in five Indian megacities. *Science of the Total Environment*, *750*, 141461. https://doi.org/10.1016/j.scitotenv.2020.141461

Sousan, S., Koehler, K., Thomas, G., Park, J. H., Halterman, A., Peters, T. M., Sousan, S., Koehler, K., Thomas, G., Park, J. H., & Hillman, M. (2016). Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols. *Aerosol Science and Technology*, *50*(5), 462–473. https://doi.org/10.1080/02786826.2016.1162901

Steinle, S., Reis, S., Sabel, C. E., Semple, S., Twigg, M. M., Braban, C. F., Leeson, S. R., Heal, M. R., Harrison, D., Lin, C., & Wu, H. (2015). Science of the Total Environment Personal exposure monitoring of PM 2 . 5 in indoor and outdoor microenvironments. *Science of the Total Environment*, *508*, 383–394. https://doi.org/10.1016/j.scitotenv.2014.12.003

Tefera, W., Kumie, A., Berhane, K., Gilliland, F., & Lai, A. (2020). *Chemical Characterization and Seasonality of Ambient Particles (PM 2 . 5) in the City Centre of Addis Ababa*. 1–16.

Wang, M., Wang, S., Wang, X., Tian, Y., Wu, Y., Cao, Y., Song, J., Wu, T., & Hu, Y. (2020). The association between PM2.5 exposure and daily outpatient visits for allergic rhinitis: evidence from a seriously air-polluted environment. *International Journal of Biometeorology*, *64*(1), 139–144. https://doi.org/10.1007/s00484-019-01804-z

Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., Biswas, P., Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., & Biswas, P. (2015). *Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement*. *6826*. https://doi.org/10.1080/02786826.2015.1100710

Watson, J. G. (2005). *Results from a pilot-scale air quality study in Addis Ababa , Ethiopia Results from a pilot-scale air quality study in Addis Ababa , Ethiopia*. *November*. https://doi.org/10.1016/j.atmosenv.2005.08.033

WBG. (2021). Safe to Breathe? Analyses and Recommendations for Improving Ambient Air Quality Management in Ethiopia. *Www.Worldbank.Org.H Street NW, Washington DC*, 72.

WHO. (2021). WHO global air quality guidelines. *Coastal And Estuarine Processes*, 1–360.

WHO. (2022a). The Global Health Cost of PM2.5 Air Pollution: A Case for Action Beyond 2021. In *WHO*. WHO. https://doi.org/10.1596/978-1-4648-1816-5

WHO. (2022b). *WHO ambient air quality database*. 28. https://cdn.who.int/media/docs/default-source/air-pollution-documents/air-quality-and-health/who-air-quality-database-2022---v7.pdf?sfvrsn=c6d52e7b_7&download=true

Williams, R., Kaufman, A., Hanley, T., & Rice, J. (2014). *Evaluation of Field-deployed Low Cost PM Sensors Evaluation of Field-deployed Low Cost PM Sensors*. *December*.

Women, P., & Lanphear, B. (2017). *Modelling Fine Particulate Matter Concentrations inside the Homes of*.

World bank. (2016). The Cost of Air Pollution. *The Cost of Air Pollution*. https://doi.org/10.1596/25013

World bank. (2022). *Bending the pollution curve*. https://www.ptonline.com/articles/how-to-get-better-mfi-results

Wu, H., Reis, S., Lin, C., & Heal, M. R. (2017). Effect of monitoring network design on land use regression models for estimating residential NO2 concentration. *Atmospheric Environment*, *149*, 24–33. https://doi.org/10.1016/j.atmosenv.2016.11.014

Yamada, D., Hiwatari, M., Narita, D., Hangoma, P., Chitah, B., Nakata, H., Nakayama, S. M. M., Yabe, J., Ito, M., Igarashi, T., Ishizuka, M., & Zyambo, G. (2023). Social cost of mining-related lead (Pb) pollution in Kabwe, Zambia, and potential remediation measures. *Science of the Total Environment*, *865*(December 2022), 161281. https://doi.org/10.1016/j.scitotenv.2022.161281

Yin, H., Pizzol, M., & Xu, L. (2017). External costs of PM2.5 pollution in Beijing, China: Uncertainty analysis of multiple health impacts and costs. *Environmental Pollution*, *226*, 356–369. https://doi.org/10.1016/j.envpol.2017.02.029

Zhai, L., Zou, B., Fang, X., Luo, Y., Wan, N., & Li, S. (2016). *Land Use Regression Modeling of PM 2 . 5 Concentrations at Optimized Spatial Scales*. *December*. https://doi.org/10.3390/atmos8010001

Zhang, H., Dong, H., Ren, M., Liang, Q., Shen, X., Wang, Q., Yu, L., Lin, H., Luo, Q., Chen, W., Knibbs, L. D., Jalaludin, B., Wang, Q., & Huang, C. (2020). Ambient air pollution exposure and gestational diabetes mellitus in Guangzhou, China: A prospective cohort study. *Science of the Total Environment*, *699*. https://doi.org/10.1016/j.scitotenv.2019.134390

Appendix 1. Supplementary figures and tables for chapter two

Figure 2.A.1. Sampling Pictures of data collection

Rainy season data summary in Addis Ababa (August - September 2021)

For the rainy season, we collected PM2.5 data from four sites from August 21 to September 25, 2021. Three sites (Asphalt road, Mixed land use, and Open area) were selected based on convenience for sampling as the data were collected by a field data collector, and the airport data were collected to verify the predicted map of air quality with actual data on the ground, previously predicted map of air quality using data from round 1 and round 2 shows the air quality is healthy at the airport compound in Addis Ababa, against the expectation where transportation areas are supposed to be polluted. What we could learn from the data we collected during the rainy season

is that the airport compound has a lower pollution level (PM2.5 is the lowest when compared with the other three land use types visited during the rainy season). This might be because the airport area is well developed, where we could find clear asphalt roads, and most of the other space is an open area with green plants coverage, mostly grasses, or the other reason might be the main pollutant of concern around airports is No_2 or CO, as a study by (Schlenker & Walker, 2016) found that airplanes raise the level of CO (carbon monoxide). PM2.5 is also of concern since particulate matter from jet exhausts is all in a fine fraction which depends on the amount of time planes spend idling on the tarmac. A further explanation might be the case that we collected the air quality data outside where the airplanes land but inside the airport compound, where most travelers use the place as a meeting compound and car parking.

When we look at the air quality level in Addis Ababa by the time of the day for the rainy season (for August 21 to September 25, 2021), the data shows that air pollution levels are highest during morning rush hours (6:00 to 8:59) and evenings (from 18:30 to 23:01) figure 1 in this Appendix. This suggests that the variation in air quality levels in a day-by-time in Addis Ababa comes from transportation.

Emissions from cars or emissions from morning activities at home need to be verified by measuring air pollution indoors. When we look at the air quality in Addis Ababa by the time of the rainy season from the four sites selected for sampling, Friday shows the most polluted air, followed by Saturday, and the lowest pollution level is observed on Thursday.

Figure 2.A.2. Rainy season PM2.5 data
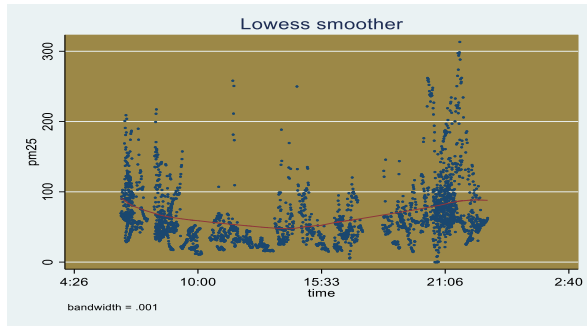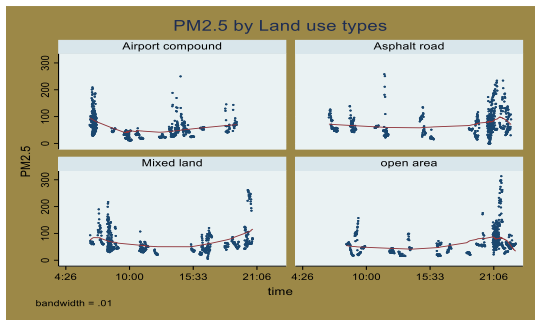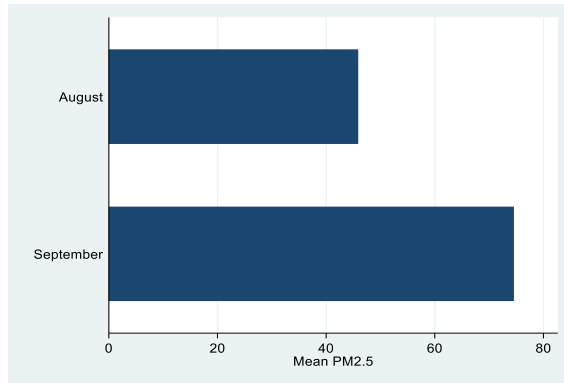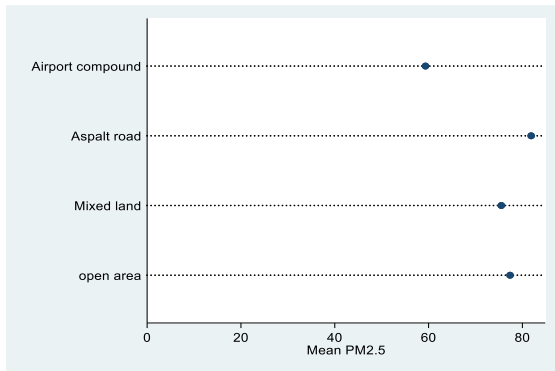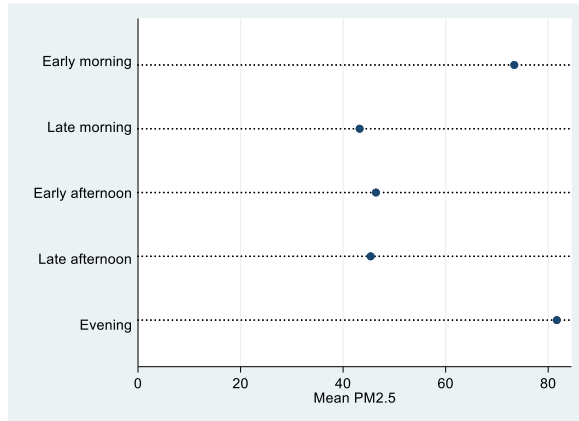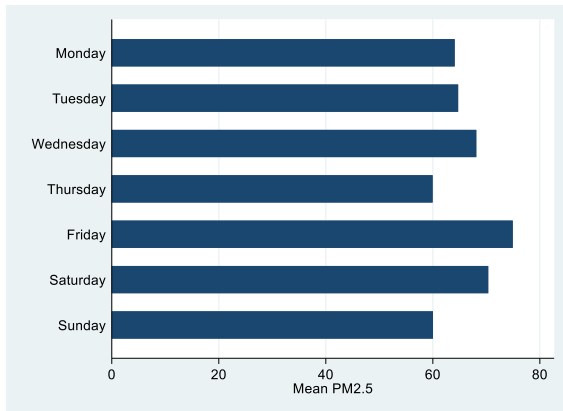
Table 2.A.1. PM2.5 by site
i)   Asphalt road sites (AR)

Site ID	Land use	Mean	Min	Median	Observation in Minutes	SD	Max
2	AR	50.61	13.56	47.84	120	25.11	113.11
5	AR	37.91	11.57	26.64	120	31.52	152.98
7	AR	63.77	12.87	47.85	120	63.28	332.91
11	AR	44.69	16.83	40.77	120	18.66	106.12
16	AR	39.76	12.21	32.19	120	27.90	136.60
19	AR	46.80	10.29	30.62	120	42.62	177.31
20	AR	52.10	10.07	28.07	120	68.61	264.04
21	AR	36.72	8.03	33.98	120	17.00	83.88
26	AR	38.57	12.39	35.81	120	17.79	122.56
27	AR	28.17	6.98	27.50	120	12.73	59.12
29	AR	32.51	7.25	29.94	120	14.92	65.27
30	AR	32.97	8.82	27.61	120	17.66	83.45
31	AR	39.54	15.13	38.00	120	18.05	106.34
32	AR	30.39	13.27	29.12	120	11.80	69.45
36	AR	60.04	8.40	48.25	120	54.51	265.54
38	AR	32.56	12.58	26.91	120	17.42	84.22
39	AR	48.01	18.15	45.12	119	19.17	159.02
40	AR	37.97	12.22	39.34	119	12.17	73.01
41	AR	43.27	13.92	30.91	120	30.76	128.44
49	AR	60.55	9.03	48.63	119	37.53	165.67
52	AR	47.18	17.48	37.24	120	36.01	222.40
54	AR	31.72	10.34	28.54	120	15.33	88.31
55	AR	28.82	7.84	25.33	120	15.99	73.01
62	AR	68.73	0.00	58.67	853	41.91	258.07
63	AR	41.98	11.96	38.33	119	22.14	92.46
69	AR	31.21	11.80	28.36	120	12.12	77.72
Total		47.54	0.00	36.89	3,849	36.38	332.91

ii)   Mixed land-use sites(ML)

Site ID	Land use	Mean	Min	Median	Observation in Minutes	SD	Max
1	ML	33.53	8.47	28.54	120	21.83	81.50
3	ML	23.64	8.12	22.29	120	11.52	56.22
4	ML	24.90	7.84	24.20	120	10.82	68.00
6	ML	31.43	9.90	22.72	120	27.29	106.78
8	ML	39.87	8.26	26.96	120	34.99	123.73
12	ML	31.95	9.26	23.64	120	30.50	143.97
13	ML	29.42	7.40	25.95	120	18.15	81.74
15	ML	33.20	8.33	24.57	117	24.13	105.96
23	ML	32.71	11.84	32.41	120	15.24	80.34
24	ML	27.03	10.54	23.40	120	12.52	75.48
25	ML	28.27	12.40	26.31	120	13.06	122.10
33	ML	27.13	10.96	26.07	120	8.92	52.60
37	ML	35.03	12.57	26.68	120	22.47	127.89
42	ML	43.14	11.64	30.35	120	36.27	170.31
43	ML	27.50	7.61	24.10	120	12.50	64.79
44	ML	22.45	8.70	20.97	120	11.78	53.71
45	ML	27.76	10.85	24.07	120	12.00	76.04
47	ML	33.16	11.31	27.33	120	17.76	96.48
53	ML	24.31	8.65	23.30	120	8.89	43.87
59	ML	34.68	9.27	25.52	120	25.07	140.21
61	ML	61.81	5.32	50.49	833	42.16	261.99
68	ML	26.58	9.28	23.53	120	15.08	69.29
72	ML	33.57	8.25	24.43	120	29.23	126.13
Total		38.02	5.32	29.13	3,470	30.85	261.99

iii)   Open area sites(OA)

Site ID	Land use	Mean	Min	Median	Observation in Minutes	SD	Max
9	OA	30.91	10.62	26.19	120	16.58	79.94
10	OA	26.33	10.75	24.11	120	10.40	52.32
14	OA	26.66	6.82	24.76	120	16.49	87.20
17	OA	23.38	7.63	22.36	120	12.68	89.29
18	OA	20.31	9.17	13.14	120	10.86	41.01
22	OA	21.84	6.61	19.74	120	12.45	50.71
28	OA	26.01	9.23	19.81	120	13.31	81.63
34	OA	26.77	12.83	28.68	119	7.92	50.00
35	OA	35.29	6.97	28.03	120	24.32	89.09
46	OA	30.21	9.76	31.91	120	11.12	54.90
48	OA	31.77	11.80	25.58	120	17.28	83.42
50	OA	47.04	9.64	32.97	120	40.21	173.43
51	OA	34.63	11.28	32.91	120	17.46	93.17
56	OA	30.14	9.34	28.05	120	14.96	84.78
57	OA	20.82	9.64	21.64	120	6.45	37.69
58	OA	27.45	9.41	23.98	120	19.27	78.41
60	OA	61.93	9.16	53.76	851	47.07	313.27
64	OA	41.67	11.63	34.73	120	23.53	86.72
65	OA	49.00	11.50	28.63	120	55.95	221.18
66	OA	52.15	9.30	25.60	120	66.09	299.66
67	OA	22.44	9.86	22.43	119	7.60	38.11
70	OA	25.32	9.11	21.24	120	11.68	50.30
71	OA	40.97	10.02	23.07	120	59.51	410.98
Total		38.86	6.61	27.49	3,489	36.66	410.98

Table 2.A.2. Spatial Statistics of PM2.5 for each route on round one and Round Two data collection period

Round (Route)	Period	N days	Observation in minutes	sites per day	Spatial Mean	Spatial Median	Min	Max	Range	SD
R-1(1)	Feb, 18-Feb24,28	8 (two rounds)	2160	18	31.55	22.95	8.61 site 14	164.72 site 71	156.11	26.018
R-1(2)	Feb 29-March,8	8 (two rounds)	2160	18	45.01	38.39	10.42 site 44	215.12 site 66	204.6	27.58
R-1(3)	March 9,11,13,14	4(one round)	1080	18	40.8	31.64	23.94 site 43	207.15 site 7	183.2	27.8
R-1(4)	March 15,17,18,19	4(one round)	1080	18	36.44	26.15	11.29 site 60	231.04 site 20	219.74	34.55
R-2(1)	Oct 17,19,20,21,22,24	6(One round)	1077	12	29.25	23.68	7.57 site 22	190.84	183.27	18.05
R-2(2)	Oct 25,27,28,29,30, and Nov 1	6(One round)	1080	12	18.79	15.69	7.63 site 17	122.1	114.47	10.1

Table 2.A.3. Leave one out cross-validation results across buffer (LOOCV)

	LOOCV				
	100m	200m	300m	400m	500m
Root Mean Squared Errors	26.57	26.63	26.63	26.63	26.59
Mean Absolute Errors	16.212	16.29	16.26	16.23	16.24
Pseudo-R2	0.39	0.39	0.39	0.39	0.39

Table 2.A.4. US-EPA air quality and implied risk categories for PM2.5

	AQI Category/implied risk	PM2.5	Health recommendation
	Good	0-12	Air quality is satisfactory and possesses little or no risk
	Moderate	12.1-35.4	Sensitive individuals should avoid outdoor activity as they may experience respiratory systems
	Unhealthy for sensitive people	35.5-55.4	The general public and sensitive individuals in particular, are at risk of experiencing irritation and respiratory problems.
	Unhealthy	55.5-150.4	Increased likelihood of adverse effects and aggravation to the heart and lungs among the general public.
	Very Unhealthy	150.5-250.4	The general public will be noticeably affected. Sensitive groups should restrict outdoor activities.

Table 2.A.5 (a). Land use regression results using distance land use variables.

VARIABLES	(1) PM25	(2) PM25calib
Rain	0.022	0.021
	(0.559)	(0.519)
Lag rain	-2.341***	-2.172***
	(0.777)	(0.721)
Wind speed	-1.430**	-1.327**
	(0.615)	(0.570)
Humidity	1.892***	1.097**
	(0.591)	(0.548)
Temperature	0.184	0.170
	(0.241)	(0.224)
River	8.389	7.783
	(29.742)	(27.597)
High-density mixed residence	151.453	140.528
	(145.311)	(134.830)
Waste disposal	-8,370.277	-7,766.529
	(5,584.667)	(5,181.844)
Bus terminal	767.245	711.903
	(1,499.117)	(1,390.986)
Manufacturing	2,397.562***	2,224.626***
	(806.061)	(747.920)
Forest	715.892	664.254
	(1,656.391)	(1,536.916)
Festival	-357.289	-331.517
	(267.102)	(247.836)
Embassy	-601.211*	-557.846*
	(309.889)	(287.537)
Intracity bus terminal	1,349.954**	1,252.582**
	(656.333)	(608.992)
Christian cemetery	480.818***	446.137***
	(125.483)	(116.432)
Urban Agric	-11.926	-11.065
	(47.762)	(44.317)
Federal bureau	79.959	74.192
	(125.552)	(116.496)
Woreda admin	-31.104	-28.861
	(58.813)	(54.571)
Primary roads	2.222***	2.062***
	(0.585)	(0.543)
Secondary roads	2.993	2.777

	(5.148)	(4.777)
Tertiary roads	8.559**	7.942**
	(3.418)	(3.171)
Footway	542.759**	503.610**
	(224.672)	(208.466)
North	10.800***	10.021***
	(3.232)	(2.999)
South	6.450	5.984
	(4.091)	(3.796)
West	3.791	3.517
	(2.881)	(2.673)
Late morning	-39.923***	-37.043***
	(4.565)	(4.236)
Early afternoon	-44.182***	-40.995***
	(4.474)	(4.152)
Late afternoon	-41.106***	-38.141***
	(4.733)	(4.391)
Evening	0.990	0.918
	(5.623)	(5.218)
Monday	3.521	3.267
	(4.037)	(3.746)
Tuesday	-6.054	-5.617
	(4.089)	(3.794)
Wednesday	-3.862	-3.584
	(4.311)	(4.000)
Thursday	-7.288	-6.762
	(4.469)	(4.147)
Saturday	-0.482	-0.448
	(5.097)	(4.729)
Sunday	-7.151*	-6.635*
	(3.765)	(3.493)
Latitude	94.612**	87.788**
	(46.629)	(43.266)
Longitude	62.139	57.657
	(50.468)	(46.828)
Constant	-3,211.017	-2,974.003
	(2,121.891)	(1,968.839)
Observations	11,099	11,099
R-squared	0.399	0.382

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.A.5 (b) LUR full sample

VARIABLES	(100m) PM25calib	(200m) PM25calib	(300m) PM25calib	(400m) PM25calib	(500m) PM25calib
Rain	0.250	0.212	0.184	0.179	0.176
	(0.497)	(0.503)	(0.494)	(0.490)	(0.485)
Lag rain	-2.384***	-2.337***	-2.258***	-2.244***	-2.22*3**
	(0.769)	(0.782)	(0.789)	(0.796)	(0.797)
Wind speed	-1.132**	-1.286**	-1.306**	-1.312**	-1.282**
	(0.564)	(0.581)	(0.588)	(0.594)	(0.595)
Humidity	0.619	0.778	0.727	0.678	0.637
	(0.562)	(0.579)	(0.597)	(0.600)	(0.590)
Temperature	0.194	0.170	0.162	0.168	0.164
	(0.219)	(0.218)	(0.212)	(0.209)	(0.206)
High-density mixed residence	0.130	0.027	0.003	-0.003	-0.005
	(0.082)	(0.028)	(0.014)	(0.010)	(0.007)
River buffer	0.302	0.053	0.056	0.046	0.049
	(0.221)	(0.073)	(0.046)	(0.035)	(0.030)
Road network	0.563**	0.163	0.106	0.076	0.088*
	(0.255)	(0.105)	(0.086)	(0.064)	(0.045)
Open area	0.006	0.019	0.014	0.001	0.001
	(0.103)	(0.042)	(0.026)	(0.019)	(0.015)
Parking land use	-0.064	0.050	0.112*	0.096*	0.063
	(0.107)	(0.052)	(0.056)	(0.051)	(0.038)
Commercial area	0.203**	0.051*	0.014	0.004	0.002
	(0.093)	(0.028)	(0.019)	(0.014)	(0.010)
Public facility	-0.360***	-0.051	-0.016	-0.007	-0.002
	(0.126)	(0.031)	(0.016)	(0.012)	(0.008)
Asphalt roads	5.343	1.594	0.599	0.229	-0.049
	(3.347)	(1.439)	(0.805)	(0.578)	(0.405)
Unpaved roads	5.615	1.418	0.692	0.146	0.070
	(3.54)	(1.520)	(0.822)	(0.616)	(0.439)
North	10.833***	10.496***	10.302***	10.175***	10.221***
	(3.418)	(3.431)	(3.490)	(3.532)	(3.552)
South	4.855	4.419	4.383	4.458	4.651
	(3.644)	(3.665)	(3.693)	(3.724)	(3.770)
West	3.174	2.705	2.556	2.595	2.626
	(2.805)	(2.823)	(2.884)	(2.925)	(2.985)

Late morning	-35.280***	-35.302***	-35.618***	-35.937***	-35.939***
	(4.095)	(4.159)	(4.161)	(4.187)	(4.217)
Early afternoon	-40.316***	-40.385***	-40.608***	-40.925***	-40.884***
	(4.316)	(4.409)	(4.401)	(4.410)	(4.424)
Late afternoon	-36.338***	-36.285***	-36.594***	-36.940***	-37.063***
	(4.350)	(4.417)	(4.438)	(4.462)	(4.502)
Evening	3.183	4.411	4.274	3.918	3.746
	(5.150)	(4.910)	(4.891)	(4.961)	(5.053)
Monday	3.496	3.399	3.059	2.965	2.761
	(3.779)	(3.916)	(3.962)	(3.978)	(3.970)
Tuesday	-6.247	-6.511	-6.773*	-6.857*	-6.86*
	(3.881)	(3.913)	(3.933)	(3.948)	(3.96)
Wednesday	-4.537	-4.705	-4.680	-4.585	-4.681
	(4.149)	(4.198)	(4.243)	(4.265)	(4.266)
Thursday	-7.586*	-7.239	-7.050	-7.003	-7.042
	(4.381)	(4.546)	(4.637)	(4.671)	(4.668)
Saturday	0.084	-0.1206	-0.394	-0.487	-0.542
	(4.958)	(5.036)	(5.042)	(5.045)	(5.039)
Sunday	-5.885	-5.934	-5.918	-5.925	-5.935
	(3.634)	(3.624)	(3.663)	(3.680)	(3.697)
Latitude	185.920***	157.163***	139.192***	138.284***	129.351***
	(35.170)	(39.782)	(44.153)	(46.022)	(46.296)
Longitude	16.069	4.515	2.200	-14.050	-17.514
	(37.207)	(39.681)	(39.350)	(39.288)	(40.601)
Constant	-2248.396	-1541.710	-1290.210	-649.979	-437.078
	(1,501.981)	(1,551.185)	(1,483.131)	(1,483.193)	(1,514.457)
Observations	10,537	10,537	10,537	10,537	10,537
R-squared	0.39	0.38	0.38	0.38	0.38

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: PM2.5 Used here is the calibrated version of Dylos reading

Table 2.A.6 Land use regression (Only land use variables used)

VARIABLES	100	200	300	400	500
	PM2.5	PM2.5	PM2.5	PM2.5	PM2.5
High density mixed residence	-6.849	-8.852	-8.238	-8.326	-11.138*
	(4.461)	(5.360)	(5.379)	(5.932)	(6.307)
River buffer	26.862*	17.903	22.523	30.291	31.567
	(15.480)	(21.173)	(23.405)	(24.083)	(28.167)
Open area	-11.673**	-8.058	-2.196	-0.889	-8.345
	(5.694)	(8.455)	(11.858)	(13.888)	(15.893)
Parking land use	-37.117***	-30.830	-27.546	-7.283	-4.639
	(11.427)	(21.370)	(27.492)	(26.214)	(28.075)
Commercial area	-15.218**	-18.787***	-20.114**	-14.878	-9.216
	(5.901)	(6.346)	(8.166)	(9.686)	(11.532)
Public facility land use	-10.199	2.807	15.740	24.956*	29.343**
	(10.468)	(13.925)	(14.193)	(14.554)	(14.496)
Asphalt road	0.019***	0.004*	0.001	-0.000	-0.000
	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)
Latitude	364.627***	326.817***	287.090***	265.862***	247.939***
	(51.447)	(62.514)	(61.316)	(59.068)	(56.508)
Longitude	227.350***	250.616***	243.940***	235.969***	231.729***
	(67.670)	(85.688)	(87.911)	(88.251)	(85.577)
Constant	-12,060.434***	-12,620.811***	-12,004.292***	-11,503.191***	-11,174.649***
	(2,792.651)	(3,458.645)	(3,527.682)	(3,550.662)	(3,405.806)
Observations	10,537	10,537	10,537	10,537	10,537
R-squared	0.128	0.105	0.104	0.105	0.108

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2.A.7 Correlation between PM2.5 and weather, time, and day of a week

Variables	PM2.5
Rain	0.482
	(0.541)
One day lag rain	-2.094**
	(0.860)
Wind speed	-0.673
	(0.728)
Humidity	2.078**
	(0.887)
Temperature	-6.891**
	(3.335)
Temperature squared	0.161**
	(0.076)
East	-13.244***
	(4.458)
South	-5.686
	(4.328)
West	-7.065***
	(2.559)
Late morning	-39.411***
	(3.858)
Early afternoon	-45.519***
	(3.871)
Late afternoon	-40.221***
	(4.193)
Evening	-4.271
	(6.336)
Tuesday	-8.193**
	(3.391)
Wednesday	-7.200**
	(3.548)
Thursday	-10.321**
	(4.315)
Friday	-4.936
	(3.770)
Saturday	-2.318
	(4.438)
Sunday	-9.848***
	(3.407)
Peak hours#Asphalt road	0.004
	(0.006)
Constant	136.927***

	(29.218)
Observations	10,537
R-squared	0.397

Table 2.A.8. LUR in a 100m buffer and interactions

VARIABLES	PM25	PM25	PM25
Rain	0.264	0.265	0.145
	(0.545)	(0.547)	(0.542)
One day lag rain	-2.732***	-2.730***	-2.777***
	(0.825)	(0.826)	(0.823)
Wind speed	-0.818	-0.815	-0.918
	(0.760)	(0.759)	(0.749)
Humidity	1.821**	1.820**	1.873**
	(0.740)	(0.742)	(0.756)
Temperature	-5.790*	-5.761*	-5.940*
	(3.223)	(3.256)	(3.233)
Temperature squared	0.136*	0.136*	0.140*
	(0.073)	(0.074)	(0.074)
High-density mixed residence	4.110	4.104	2.355
	(2.903)	(2.905)	(2.550)
River buffer	6.218	6.212	6.069
	(7.226)	(7.222)	(7.168)
Open area	-4.129	-4.124	-4.347
	(2.721)	(2.722)	(2.769)
Parking land use	-0.792	-0.772	-1.527
	(4.243)	(4.261)	(4.310)
Commercial area	4.778*	4.982*	4.802*
	(2.683)	(2.768)	(2.808)
Public facility land use	-10.882**	-10.887**	-10.984**
	(4.535)	(4.539)	(4.465)
Asphalt road	0.011***	0.011***	0.011***
	(0.003)	(0.003)	(0.003)
Asphalt#Peak hour	-0.002	-0.002	-0.002
	(0.008)	(0.008)	(0.008)
Commercial#Peak hour		-2.827	-1.618
		(11.072)	(11.843)
High-density #peak hour			12.044
			(12.631)

East	-13.620***	-13.618***	-13.069***
	(4.037)	(4.041)	(4.023)
South	-6.814	-6.817	-6.737
	(4.337)	(4.337)	(4.325)
West	-7.513***	-7.529***	-7.345***
	(2.520)	(2.529)	(2.520)
Late Morning	-37.341***	-37.404***	-35.459***
	(3.887)	(4.030)	(4.209)
Early Afternoon	-43.551***	-43.608***	-41.843***
	(3.822)	(3.951)	(4.219)
Late Afternoon	-39.004***	-39.059***	-37.347***
	(4.136)	(4.268)	(4.435)
Evening	-3.009	-3.000	-2.180
	(6.498)	(6.491)	(6.833)
Tuesday	-10.892***	-10.887***	-10.547***
	(3.111)	(3.116)	(2.954)
Wednesday	-9.153***	-9.152***	-8.639***
	(3.281)	(3.283)	(3.077)
Thursday	-11.857***	-11.790***	-11.454***
	(4.205)	(4.304)	(4.152)
Friday	-4.104	-4.089	-3.911
	(3.912)	(3.920)	(3.818)
Saturday	-4.296	-4.291	-3.971
	(4.278)	(4.283)	(4.094)
Sunday	-10.655***	-10.648***	-10.245***
	(3.197)	(3.206)	(2.949)
Latitude	189.350***	189.366***	197.135***
	(37.885)	(37.890)	(37.904)
Longitude	-9.604	-9.368	-9.996
	(39.166)	(39.379)	(39.298)
Constant	-1,209.079	-1,218.667	-1,264.310
	(1,557.918)	(1,566.035)	(1,562.350)
Observations	10,537	10,537	10,537
R-squared	0.410	0.410	0.411

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.A.9. Pollution by area (using US-EPA air quality level by time of day)

	Health impact	Range of PM2.5	Early morning	Late morning	Early afternoon	Late afternoon	Evening
🟩	Good	0-12	0	102.78	145.49	128.78	0
🟨	Moderate	12.1-35.4	8.8	309.02	358.59	325.92	4.7
🟧	Unhealthy for sensitive people	35.5-55.4	132.38	107.25	14.97	64.34	122.3
🟥	Unhealthy	55.5-150.4	377.877	0	0	0	391.99
🟪	Very Unhealthy	150.5-250.4	0	0	0	0	0

Table 2.A.10 Major land use types in Addis Ababa

Land use type	Area in Square km
Commercial	9.69
Public facility land use	15.77
Parking Land use	10.69
Open area	79.97
Road network	34.59
River buffer	38.68
High-density mixed residence	24.27
Existing manufacturing and storage	10.73
Proposed Manufacturing & Storage	20.68
Proposed Mixed Residence	34.48
Multifunctional forest	60.7
Existing mixed residence	165.38
Other/unclassified	15.04
Total	520.67

Figure 2.A.3 Map of predicted PM2.5 by time of day for Addis Ababa

A. Peak hours
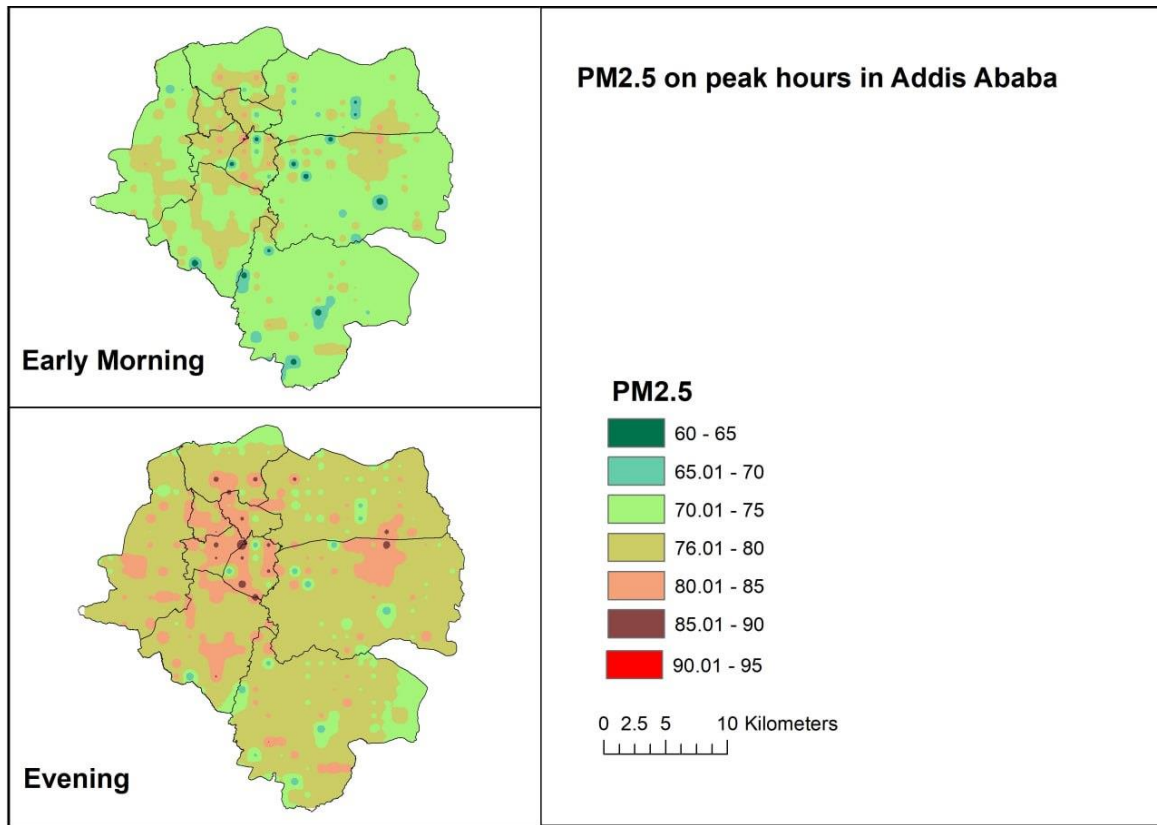
B. Off peak hours
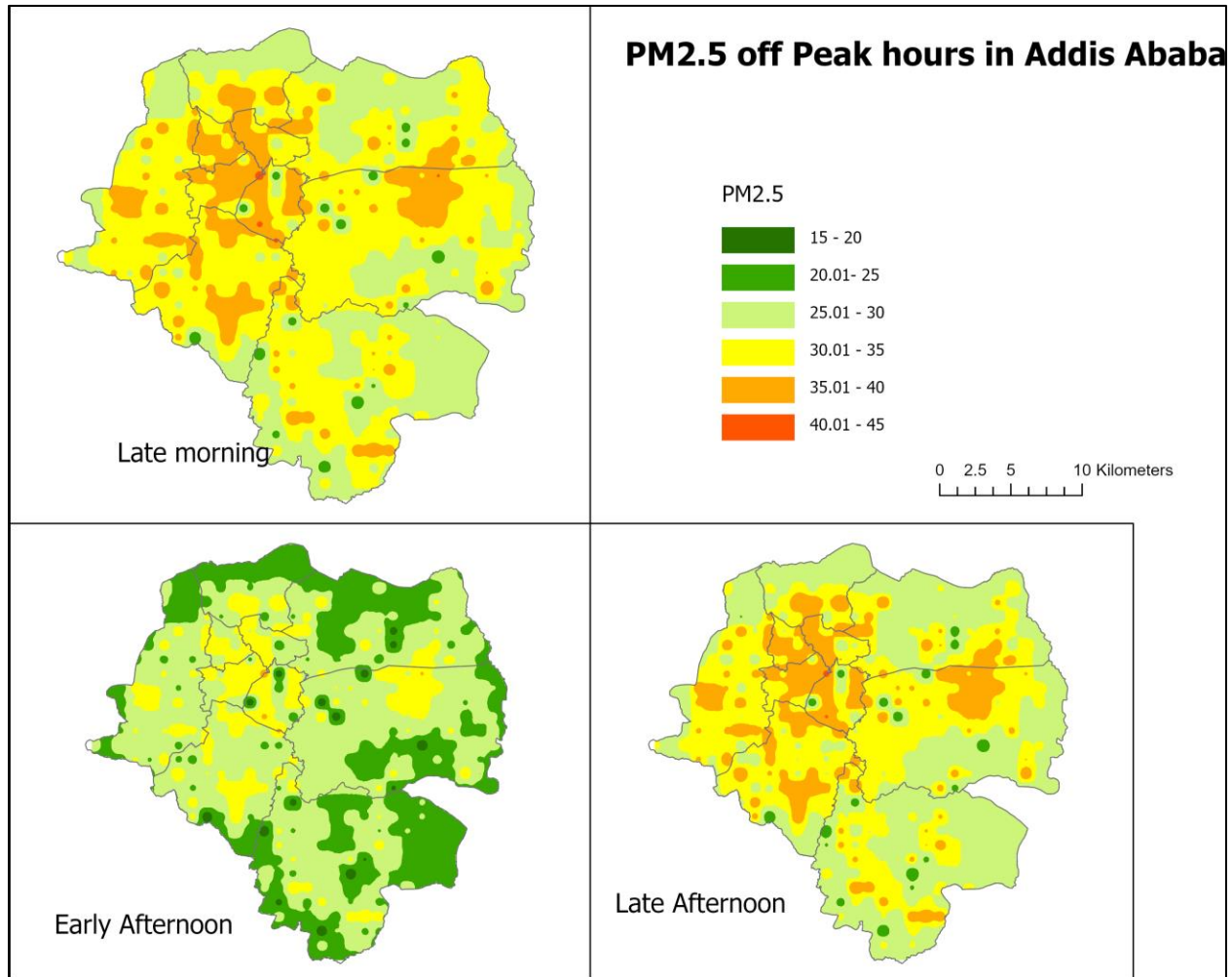
Note: Pollution levels were categorized into six ranges to illustrate spatial and temporal variations. Notably, the maps for peak and off-peak hours showed a significant difference in their range, prompting me to create separate maps for each time period. It's important to observe that while the colors in Figures A and B are identical, they correspond to distinct air quality levels. For instance, the red hue in Figure A signifies a different air quality level compared to its representation in Figure B.

Figure 2.A.4 Map of predicted PM2.5 by wind direction in Addis Ababa

PM2.5 by wind direction

PM2.5

![dark green]	50 - 55
![light green]	55.01- 60
![yellow]	60.01- 65
![orange]	65.01- 70
![red]	70.01- 75

South

East

0   1.25   2.5   5 Kilometers
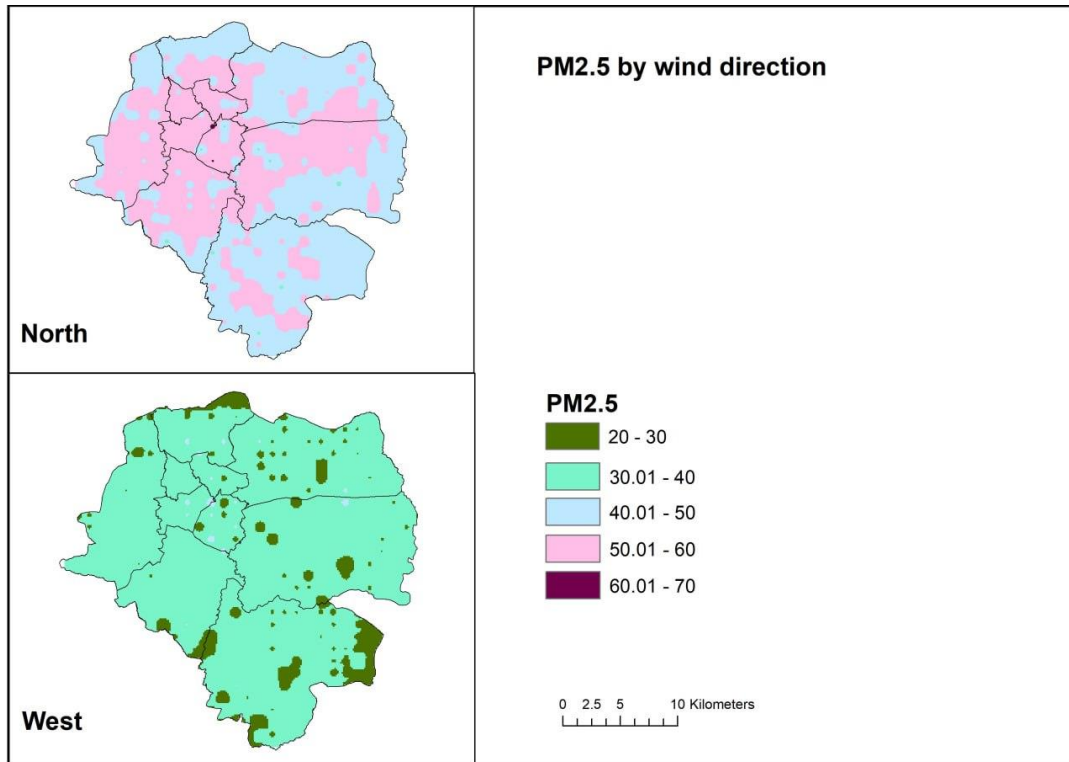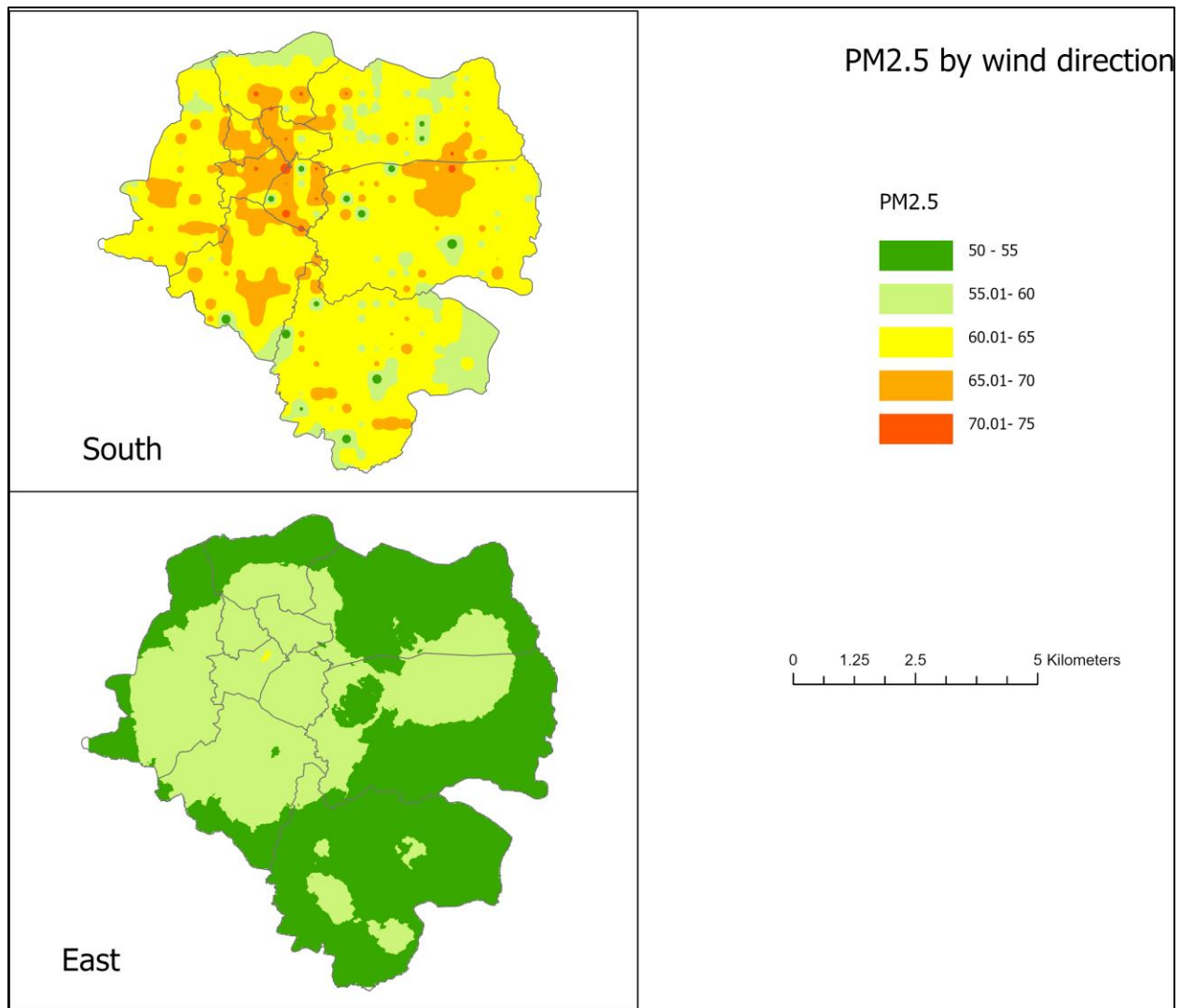
Note: The maps based on wind direction are averages of the data collection period. That is we do not have a map based on wind direction for a specific time period in a day. It's important to observe that the same colors in the two sub-figures are identical, they correspond to distinct air quality levels. For instance, the red hue in the first Figure signifies a different air quality level compared to its representation in the second Figure.

2.A.5 Distribution of land use types in Addis Ababa
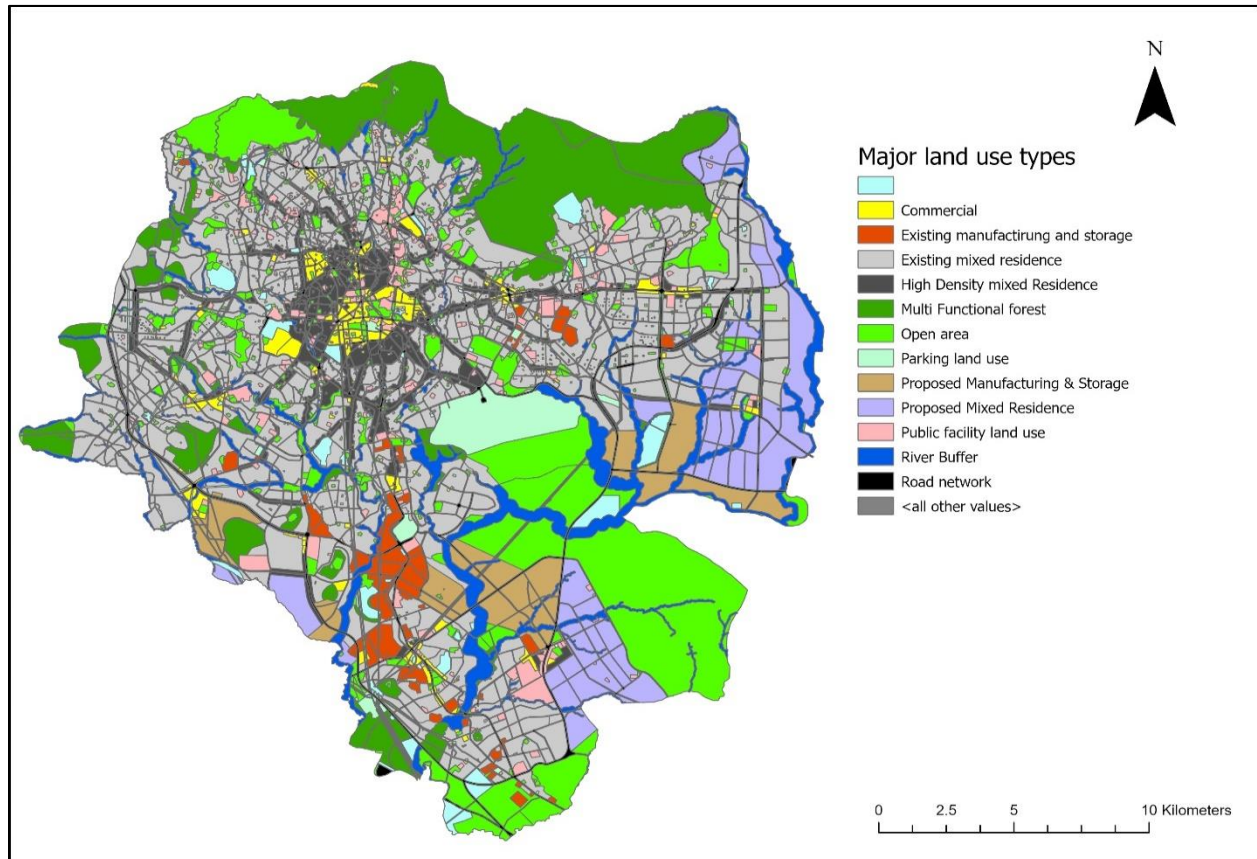
Appendix 2 Supplementary tables and figures for chapter three

i.   Checking the reliability of the Portable device (Dylos 1700)

Validation of low-cost devices against high-quality equipment has been done in both field and laboratory settings. An example of the device we used is a study by Steinle et al. (2015), where the performance of Dylos 1700 was validated in comparison with equivalent instruments (TEOM-FDMS) at two national monitoring network sites in the UK (R^2 = 0.9 at a rural background site, R^2 = 0.7 at an urban background site).

For this device, the recorded data has two main components: date/time and particle number counts. The monitor counts particles in the form of two components (particles >0.5μm per cubic fit/100, which includes both small particles and large) and (particle size > 2.5 μm for large/coarse particles). By subtracting the coarse/large particles from the overall number counted, we can get an estimate for the total number of particles less than 2.5 μm (Han et al., 2017, Dacuntoa et al., 2015); this calculation, however, does not include particles with diameters less than 0.5μm which contributes less to mass concentration Dacuntoa et al (2015).

In the validation experiment, the portable device (Dylos 1700) is placed as close as possible to the fixed air quality monitor location in Addis Ababa in two land use types, one at an asphalt road in front of the US Embassy which is nearly 50 meters away from the fixed monitor in Addis Ababa and one at an open area site which is nearly 360 meters away from the fixed monitor. The fixed air quality monitor uses a virtual impactor to separate particles by size and uses plastic scintillation as a detector, while the Portable device (Dylos 1700) uses light scattering technology to detect the particles.

Placing the two devices close to each other will help to correct the slopes and offset the intercept values of the portable device to improve the accuracy of results Munir et al.,(2019). It is

recommended that the devices be collocated for several days or weeks in academic research, although there is no exact number of validity experiment days specified. Kumar et al. (2017) pointed out that before performing any PM monitoring task, the sensors should be properly calibrated since the long-term (more than a week) performance of the low-cost sensors remains unknown since environmental factors such as temperature and relative humidity might influence PM sensors. Furthermore, sensors' response might change due to "day of use" and dust accumulation Kumar et al., (2017). In this pilot study, we placed 326 minutes at an asphalt road site and 104 minutes at an open area site in 2020 and 2021. We also included humidity in the validation experiment to check if these variables influence the performance of the portable device. In checking the validity of the device, the Portable (Dylos1700) device measurements are regressed against fixed device measurements, where the data from the fixed air quality monitor is taken as a dependent variable while the readings from the portable are taken as an independent(X value) following Munir et al., (2019). As has been shown in 3.3 of the main text, the calibration results from the open area site give us a better estimate of the fixed monitor data, and we used those calibration results to adjust the portable reading to the fixed monitor reading.

Table 3.A.1: Discrepancy and matches by land use and wind direction.

	Dylos Raw (i)						
	Panel A				Panel B		
Land use	Dylos Above moderate air quality	US above moderate air quality	Matched	Wind direction	Dylos Above moderate air quality	US above moderate air quality	Matched
Airport	67	19	4	East	64	52	3
Asphalt road	52	21	30	North	47	29	51
Mixed land	38	20	32	South	74	28	9
Open area	37	22	34	West	27	4	37

	Dylos calibrated(ii)						
	Panel A				Panel B		
Land use	Dylos Above moderate air quality	US Above moderate air quality	Matched	Wind direction	Dylos Above moderate air quality	US Above moderate air quality	matched
Airport	60	19	4	East	61	52	3
Asphalt road	47	21	31	North	42	29	50
Mixed land	32	20	32	South	67	28	10
Open area	33	22	33	West	23	4	37

Table 3.A.2: Overall Discrepancy and matches after calibrating the portable device (in %)

			Dylos calibrated↓			
	Panel C					
	US| Dylos 1700	Good	Moderate	Unhealthy for sensitive people	Unhealthy	Very unhealthy and a hazard
US→	Good	2	5	1	0	0
	Moderate	4	44	14	9	1
	Unhealthy for sensitive people	0	4	7	5	1
	Unhealthy	0	0	1	0	0

Very unhealthy and a hazard	0	1	0	0	0

Table 3.A.3. Linear difference estimation results using distance land use variables.

VARIABLES	(1) Difference PM2.5 (No calibration	(2) PM2.5 US Embassy	(3) Difference PM2.5 (Calibrated)
Rain	0.356	0.328	0.353
	(0.652)	(0.395)	(0.617)
Lag rain	0.837	-1.253***	0.679
	(0.722)	(0.332)	(0.672)
Wind speed	0.154	-1.259***	0.053
	(0.468)	(0.397)	(0.435)
Humidity	-0.754	1.197***	0.050
	(0.494)	(0.376)	(0.461)
Temperature	0.999***	1.039***	1.006***
	(0.193)	(0.232)	(0.184)
River	-14.096		-12.575
	(38.797)		(36.590)
High-density mixed residence	-45.373		-35.266
	(165.821)		(154.783)
Waste disposal	14,726.369**		13,201.634*
	(7,215.208)		(6,738.198)
Bus terminal	-2,853.102*		-2,730.890*
	(1,493.755)		(1,399.719)
Manufacturing	23.618		57.768
	(1,043.834)		(973.266)
Forest	94.786		162.825
	(1,758.352)		(1,665.629)
Festival	406.250		388.967
	(350.177)		(340.333)
Embassy	354.930		317.537
	(384.928)		(362.977)
Intracity bus terminal	-121.291		-85.906
	(670.171)		(623.877)
Christian cemetery	-412.654*		-369.147*
	(227.474)		(221.195)
Urban Agric	-13.199		-13.260
	(67.599)		(63.492)
Federal bureau	36.385		46.322
	(170.657)		(163.225)
Woreda admin	36.904		32.881
	(45.339)		(42.205)
Primary roads	-2.353***		-2.204***

	(0.555)		(0.517)
Secondary roads	-0.867		-0.820
	(4.312)		(4.008)
Tertiary roads	-7.327		-6.574
	(5.470)		(5.137)
Footway	-420.459***		-383.602***
	(150.810)		(139.689)
North	-7.581*	3.896	-6.772*
	(3.840)	(2.676)	(3.687)
South	-5.387	0.498	-4.971
	(3.434)	(1.855)	(3.180)
West	-5.113	-2.121	-4.897
	(3.744)	(2.207)	(3.582)
Late morning	26.075***	-14.382***	23.188***
	(4.187)	(2.506)	(3.919)
Early afternoon	26.990***	-17.145***	23.828***
	(3.396)	(2.758)	(3.155)
Late afternoon	25.776***	-14.322***	22.879***
	(4.049)	(2.898)	(3.791)
Evening	2.411	5.985	2.597
	(5.502)	(4.361)	(5.257)
Monday	-2.502	2.106	-2.198
	(4.268)	(2.696)	(4.056)
Tuesday	2.971	-2.070	2.573
	(3.535)	(2.046)	(3.298)
Wednesday	-0.923	-4.243	-1.185
	(3.120)	(2.879)	(2.927)
Thursday	4.004	-3.555	3.472
	(2.989)	(3.178)	(2.805)
Saturday	-4.113	-5.301	-4.192
	(3.933)	(3.818)	(3.732)
Sunday	2.161	-4.641	1.658
	(3.367)	(2.839)	(3.206)
Latitude difference	86.255**		81.555**
	(42.670)		(40.248)
Longitude difference	-180.765**		-170.941**
	(78.260)		(73.516)
Constant	-43.698***	11.204	-45.261***
	(9.195)	(9.781)	(8.746)
Observations	11,024	11,143	11,024
R-squared	0.299	0.266	0.266

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The dependent variable in the first column is the difference between the US Embassy PM2.5 reading and Dylos reading. The dependent variable in the second column is the US Embassy reading. The dependent variable

in the third column is the difference between the US Embassy PM2.5 reading and the Calibrated PM2.5 reading. Land use explanatory variables are / distance in meters for ease of description.

Table 3.A.4. WHO interim target levels and 24 hours mean concentration.

Interim target level	24 hours mean pm2.5 (μm^3	Annual mean pm2.5 (μm^3	The basis for the selected level
Interim target I (IT-1)	75	35	These levels are associated with about a 15% higher long-term mortality risk relative to the AQG level.
Interim target-II (IT-2)	50	25	In addition to other health benefits, these levels lower the risk of premature mortality by approximately 6% [2–11%] relative to the IT-1 level.
Interim target-III (IT-3)	37.5	15	In addition to other health benefits, these levels reduce the mortality risk by approximately 6% [2-11%] relative to the -IT-2 level.
Air quality Guideline (AQG)	25	10	These are the lowest levels at which total cardiopulmonary and lung cancer mortality have been shown to increase with more than 95% confidence in response to long-term exposure to PM2.5.

Table 3.A.5 (a) Multinomial logit model estimation results using distance land use variables.

VARIABLES	MNL (Base outcome j=1)		
	P(y=2)	P(y=3)	P(y=3)
Rain	0.130***	1.159***	0.303***
	(0.021)	(0.105)	(0.031)
Lag rain	-0.083***	0.289***	-0.507***
	(0.027)	(0.070)	(0.039)
Wind speed	-0.274***	-0.554***	-0.316***
	(0.021)	(0.065)	(0.029)
Humidity	0.103***	-0.415***	0.125***
	(0.024)	(0.103)	(0.034)
Temperature	-0.019	0.197***	0.113***
	(0.015)	(0.046)	(0.021)
River	4.287***	-22.574***	7.812***
	(1.629)	(7.342)	(1.694)
High-density residence	-32.199***	-1.860	14.098*
	(6.313)	(19.069)	(8.334)
Waste treatment	348.220	-15,046.978***	-6,342.145***
	(396.918)	(2,376.657)	(983.763)
Bus terminal	757.891***	164.981	230.573**
	(68.583)	(295.594)	(109.261)
Manufacturing	62.049	476.624***	741.330***
	(70.062)	(154.433)	(83.507)
Forest	202.826*	1,002.917***	886.898***
	(106.056)	(265.858)	(142.687)
Festival site	-197.090***	-125.731	57.897**
	(43.246)	(90.531)	(23.724)
Embassy	-92.726***	-230.964***	-94.601***
	(16.658)	(60.434)	(22.131)
Bus terminal	-110.238***	-162.278	-91.079
	(38.588)	(199.134)	(59.165)
Christian cemetery	12.450	91.098***	-11.868
	(15.497)	(21.751)	(17.874)
Urban agriculture	-8.392***	9.683*	-9.524***
	(2.840)	(5.768)	(3.477)
Federal bureau	25.229***	165.689***	61.256***
	(8.412)	(21.995)	(10.131)
Woreda admin	-19.821***	4.024	-80.119***
	(4.554)	(9.045)	(17.608)
Primary roads	0.354***	-0.036	0.191***

	(0.031)	(0.084)	(0.044)
Secondary roads	-0.216	-21.565**	-0.090
	(0.149)	(10.273)	(0.292)
Tertiary roads	-0.310	0.030	1.768***
	(0.459)	(0.731)	(0.535)
Footway	74.193***	-40.288	4.217
	(10.341)	(50.395)	(26.379)
North	0.036	0.459	0.264
	(0.174)	(0.291)	(0.187)
South	0.544***	1.187***	-0.058
	(0.190)	(0.444)	(0.221)
West	-0.322*	-2.383***	-2.214***
	(0.181)	(0.350)	(0.218)
Late morning	-2.802***	-2.412***	-3.547***
	(0.120)	(0.357)	(0.152)
Early afternoon	-3.276***	-3.065***	-4.482***
	(0.131)	(0.375)	(0.173)
Late afternoon	-2.769***	-9.133***	-5.078***
	(0.125)	(1.147)	(0.225)
Evening	-0.440***	0.039	-0.302
	(0.158)	(0.444)	(0.189)
Monday	0.169	-1.108***	0.600***
	(0.122)	(0.254)	(0.144)
Tuesday	0.066	-1.265***	-0.238
	(0.117)	(0.283)	(0.154)
Wednesday	-0.019	-8.909***	-1.074***
	(0.112)	(0.872)	(0.154)
Thursday	-0.773***	-3.525***	-1.425***
	(0.130)	(0.390)	(0.174)
Saturday	-0.132	-21.755	-1.957***
	(0.108)	(570.878)	(0.166)
Sunday	-0.546***	-4.155***	-1.605***
	(0.113)	(0.438)	(0.171)
Latitude difference	-15.170***	18.024***	-12.527***
	(1.800)	(6.211)	(2.591)
Longitude difference	23.191***	97.735***	6.036
	(3.336)	(11.020)	(4.899)
Constant	1.755***	1.240	0.663
	(0.548)	(1.893)	(0.759)
Observations	11,024	11,024	11,024

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The base outcome is j=1(both monitors record WHO interim target III is met), P(y=2){US Embassy<37.5 and Portable monitor≥37.5}, P(y=3){US Embassy≥37.5 and Portable monitor <37.5}, P(y=4){US Embassy≥37.5 and Portable monitor ≥37.5}.

(b) Multinomial logit model estimation results with three categories

| | MNL (base outcome=1) | |
VARIABLES	P(y=2)	P(y=3)
Rain	0.09***	0.78***
	(0.01)	(0.13)
Lag rain	0.14***	0.01
	(0.02)	(0.11)
Wind speed	-0.19***	-0.47***
	(0.01)	(0.07)
Humidity	0.0004	-0.37***
	(0.02)	(0.13)
Temperature	-0.03**	0.28***
	(0.01)	(0.06)
High-density mixed residence	0.02***	0.01
	(0.03)	(0.01)
River buffer	0.05***	-0.30***
	(0.01)	(0.05)
Road network	-0.01	-0.03
	(0.01)	(0.03)
Open area	-0.02***	-0.03**
	(0.01)	(0.01)
Parking land use	0.06***	0.12**
	(0.01)	(0.06)
Commercial area	0.03***	-0.16***
	(0.01)	(0.03)
Public facility	-0.02***	-0.01
	(0.01)	(0.02)
Asphalt roads	1.25***	-0.98**
	(0.18)	(0.50)
Unpaved roads	0.52***	-0.01***
	(0.14)	(0.6)
North	0.39***	-0.73**
	(0.14)	(0.30)
South	0.76***	-0.29
	(0.15)	(0.40)
West	0.34**	-4.13***
	(0.15)	(0.69)
Late morning	-1.25***	-2.67***
	(0.09)	(0.37)
Early afternoon	-2.04***	-3.19***

	(0.11)	(0.41)
Late afternoon	-1.38***	-7.61***
	(0.10)	(1.11)
Evening	-0.28**	0.13
	(0.11)	(0.46)
Monday	-0.25**	-2.09***
	(0.11)	(0.29)
Tuesday	0.09	-1.08***
	(0.10)	(0.32)
Wednesday	0.16	-5.77***
	(0.10)	(0.80)
Thursday	-0.41***	-1.94***
	(0.11)	(0.37)
Saturday	0.58***	-19.02
	(0.10)	(354.24)
Sunday	-0.15	-4.52***
	(0.10)	(0.58)
Latitude difference	-13.80***	-18.85***
	(1.51)	(4.90)
Longitude difference	23.48***	25.80***
	(2.28)	(9.03)
Constant	0.31	2.12
	(0.49)	(2.68)
Observations	10,462	10,462

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The base outcome is j=1(both monitors record WHO interim target III is met or unmet), P(y=2) {US Embassy<37.5 and Portable monitor≥37.5}, P(y=3) {US Embassy≥37.5 and Portable monitor <37.5}

Table 3.A.6 Categories of the dependent variable in the multinomial logit model

Scenario (if we use the Multinomial model), y=j	US Embassy	Dylos 1700	Information difference	Frequency	Percent
j=1	<37.5	<37.5	No [both interim record target is met]	6265	55.85
j=2	<37.5	≥37.5	Yes [US Embassy record target is met, and Dylos records target is unmet]	3086	27.51
j=3	≥37.5	<37.5	Yes [US Embassy record target is unmet, and Dylos record target is met]	395	3.52
j=4	≥37.5	≥37.5	No [Both record target is unmet]	1472	13.12
Total				11,218	100

ii). Replicating the US embassy set up for locating a "new" fixed monitor.

Discrepancy d_{it} is defined as,

$$d_{it} = \rho + \beta X_i + \gamma W_t + \mu T_t + \alpha_0(LO_i - LO_{US})\delta_{i0} + \alpha_1(LA_i - LA_{US})\delta_{i1} \tag{1}$$

Where d_{it} refers to the discrepancy for the site i at time t, X_i is the site characteristics vector for site i, W_t is the weather variables vector for time t, T_t is the time of day and week vector for time t. LO refers to longitude, i is the site's subscript, n is the subscript for the nearest fixed monitor, and LA refers to latitude. The deltas are either +1 or -1 and match the signs of the terms in parenthesis. In other words, we are using the absolute value difference of longitude and latitude in the discrepancy equation to pick up the effect of location difference as we used the absolute difference in the regression to show the effect of moving away from the fixed site (i.e., the US Embassy).

Discrepancy squared for a given time and site is given by,

$$d_{it}^2 = (\rho + \beta X_i + \gamma W_t + \mu T_t + \alpha_0(LO_i - LO_{US})\delta_{i0} + \alpha_1(LA_i - LA_{US})\delta_{i1})^2 \tag{2}$$

We want to minimize the sum of expected discrepancies across sites for a relevant time.

$$\text{Min} \sum_{\substack{i \\ LO_{US}}} \sum_{\substack{t \\ US}} \text{E}(d_{it}^2) = (\rho + \beta X_i + \gamma W_t + \mu T_t + \alpha_0(LO_i - LO_{US})\delta_{i0} + \alpha_1(LA_i - LA_{US})\delta_{i1})^2 \tag{3}$$

As we move the longitude and latitude of the proposed site, we differentiate the above expression with respect to LO_{US} and LA_{US}, provided there is no location switches discrepancy changes and

the change for a marginal change in latitude and longitude for a given time and location is given

by the first order conditions in equations (4) and (5)

$$\frac{\partial \sum_i \sum_t E(d_{it}^2)}{\partial LO_{US}} = -2\alpha_0 \sum_i \sum_t \delta_{i0}(\rho + \beta X_i + \gamma W_t + \mu T_t + \alpha_0(LO_i - LO_{US})\delta_{i0} + \alpha_1(LA_i - LA_{US})\delta_{i1}) = 0$$

$$\frac{\partial d_{it}^2}{\partial LO_{US}} = -2\alpha_0 \sum_i \sum_t E(\delta_{i0} d_{it}) = 0 \tag{4}$$

$$\frac{\partial \sum_i \sum_t E(d_{it}^2)}{\partial LO_{US}} = -2\alpha_1 \sum_i \sum_t \delta_{i1}(\rho + \beta X_i + \gamma W_t + \mu T_t + \alpha_0(LO_i - LO_{US})\delta_{i0} + \alpha_1(LA_i - LA_{US})\delta_{i1}) = 0$$

$$\frac{\partial d_{it}^2}{\partial LO_{US}} = -2\alpha_1 \sum_i \sum_t E(\delta_{i1} d_{it}) = 0 \tag{5}$$

Since the sampling sites were chosen at random, to find the optimal location, we take the

expectation of the total discrepancy across all sites and relevant times, and since the derivatives

are linear, we take expected values of the derivatives across sites and time as we want to minimize

expected discrepancy. Hence the expected value of the derivatives across sites and times that we

want to get the minimum (close to zero in absolute terms) is given as,

$$E \left[\frac{\partial d_{it}^2}{\partial LO_{US}}\right] = \text{Min} \left[-2\alpha_0 \sum_i \sum_t E(\delta_{i0} d_{it}) \right] \tag{6}$$

$$E \left[\frac{\partial d_{it}^2}{\partial LA_{US}}\right] = \text{Min} \left[-2\alpha_1 \sum_i \sum_t E(\delta_{i1} d_{it}) \right] \tag{7}$$

Implementation of the method

To find the optimal location, we take the expectation of total discrepancy across all sites and

relevant times. We obtained the values of $(\gamma W_t + \mu T_t)$ at their means. For example, in a seven-day

week period, Thursday occurs one-seventh of the time; this applies to other days of the week.

Meanwhile, in a 16-hour day (6:00-24:00- we do not have data from the low-cost monitor for other

times of the day), the early morning period occupies 3 hours, so its weight is 3/16. We calculated the weights of other times of the day similarly to get weights.

The term $(\gamma W_t + \mu T_t)$ does not vary by site, and it is common to all sites. The weather (γ) and time variables (μ) coefficients are obtained from the discrepancy regression. For wind direction, we took the coefficients from the respective wind directions dummy from the regression, and the weights were calculated from the population weather data (population weather data is defined as the weather data for the years 2020 and 2021). We used population weather data for winds; 2.19% of the wind is to the north; therefore, the weight given to the north dummy is 0.0219; see Table 3.A.8 for a summary of weights given to weather and time variables in this appendix.

The term βX_i is obtained by taking the coefficient β from the discrepancy regression, and X_i is the respective measure of land use for each site; hence βX_i varies by site. This term captures the effect of land use types on the discrepancy. After obtaining all these, we weight $(\delta_{i0} d_{it})$ and $(\delta_{i1} d_{it})$ by the respective sites' sub-city population density, the weight of sub-city s is calculated as ($W_s = \dfrac{\text{Population density at sub−city s}}{\text{Population density of the Addis Ababa}} * \text{share of sub − city in total area})$ and the weight for each site in a sub-city s is given by

$W_{is} = \dfrac{W_s}{n_i}$ where n_i is the number of sites in a given sub-city; see Table 3.A.8 for weights given to each sub-city. The weighted sum of expected discrepancy for a marginal change in longitude and latitude is given in equations 8 and 9.

The weighted sum of expected discrepancy for a marginal change in longitude

$$W_{is} \, E\left[\frac{\partial d_{it}^2}{\partial LO_{US}}\right] = \; W_{is} \, \{[\text{-}2\alpha_0 \sum_i \sum_t E(\delta_{i0} d_{it})]\} = \text{-}2\alpha_0 \sum_i \sum_t E(W_{is} \, \delta_{i0} d_{it}) \tag{8}$$

The weighted sum of expected discrepancy for a marginal change in latitude

$$W_{is} \, E \, [\frac{\partial d_{it}^2}{\partial LA_{US}}] = \ W_{is} \, \{[-2\alpha_1 \sum_i \sum_t E(\delta_{i1} d_{it})]\} = -2\alpha_1 \ \sum_i \sum_t E(W_{is} \delta_{i1} d_{it}) \tag{9}$$

After obtaining all the calculated values, we adjust the latitude and longitude of the proposed site until the two weighted first-order conditions are close to zero or until it sets the discrepancy to the minimum possible if the calculated d_{it} is positive, it means that LO_{US} and LA_{US} needs to be increased, if the calculated d_{it} is negative then LO_{US} and LA_{US} Needs to be reduced. Zero discrepancies are impossible with finite sites as the model does not capture the entire land use types. We adjust longitude and latitude until the first two order conditions are close to zero or the gap is minimum and locate the new site at the point where it sets the discrepancy to a minimum.

We used a trial and error to find the optimal location; we started with a reasonable guess; a location that is farthest from the US embassy site since we want to reduce the bad (wrong) advice from the monitor (the discrepancy) and the discrepancy is affected by land use; therefore, we first took a location which is farthest from the US Embassy and makes the location adjustment from that point.

The final site of the new monitor location appears on a mixed land-use site in the middle of the study area. The population density of the new monitor location sub-city is the second highest of the whole city. We used distance land use variables as a sensitivity check to the model and the proposed site, which minimizes the weighted sum expected cost of the discrepancy that appears in the asphalt road site in the same sub-city to the results of the main model.
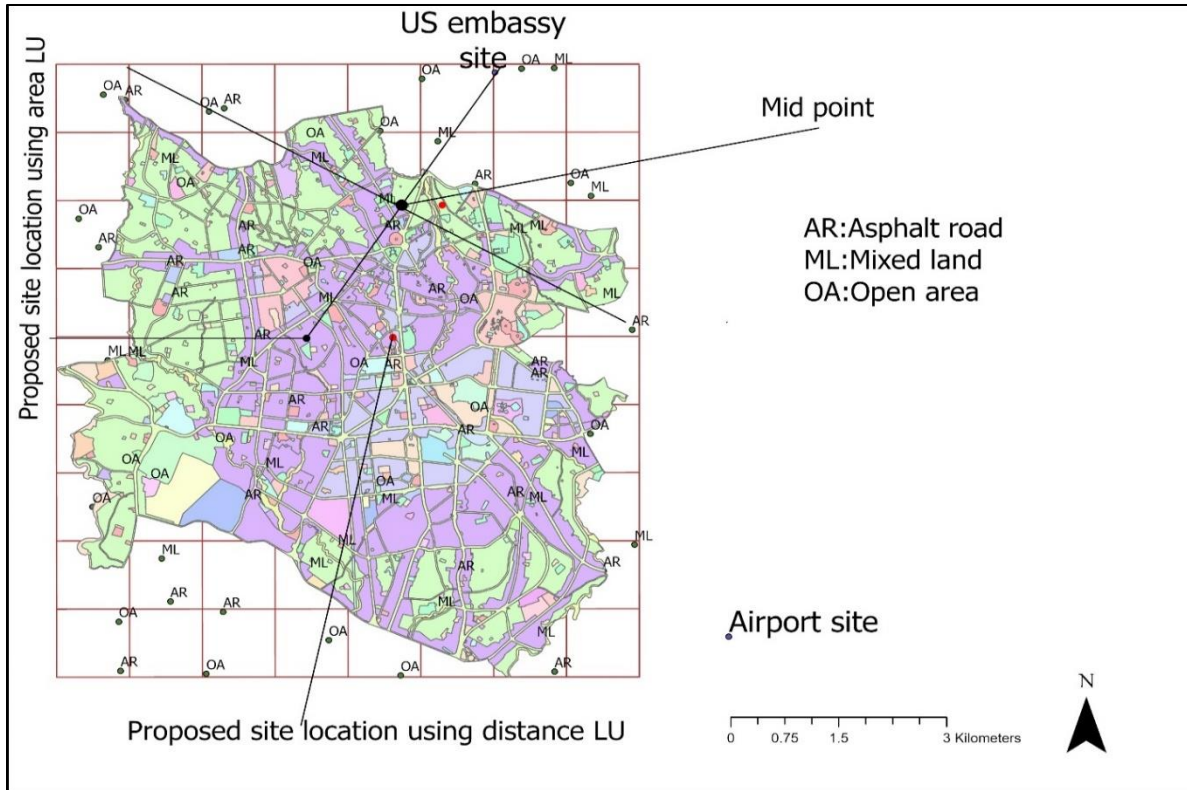
Note: Suggested location of the fixed monitor from the main model is in black dot, and the suggested monitor location is by altering the measurement type of the land use (using distance land use variables illustrated in red dot).

Source: Author`s illustration using ArcGIS Pro

Figure 3.A.1 Suggested optimal location for a fixed site in Addis Ababa
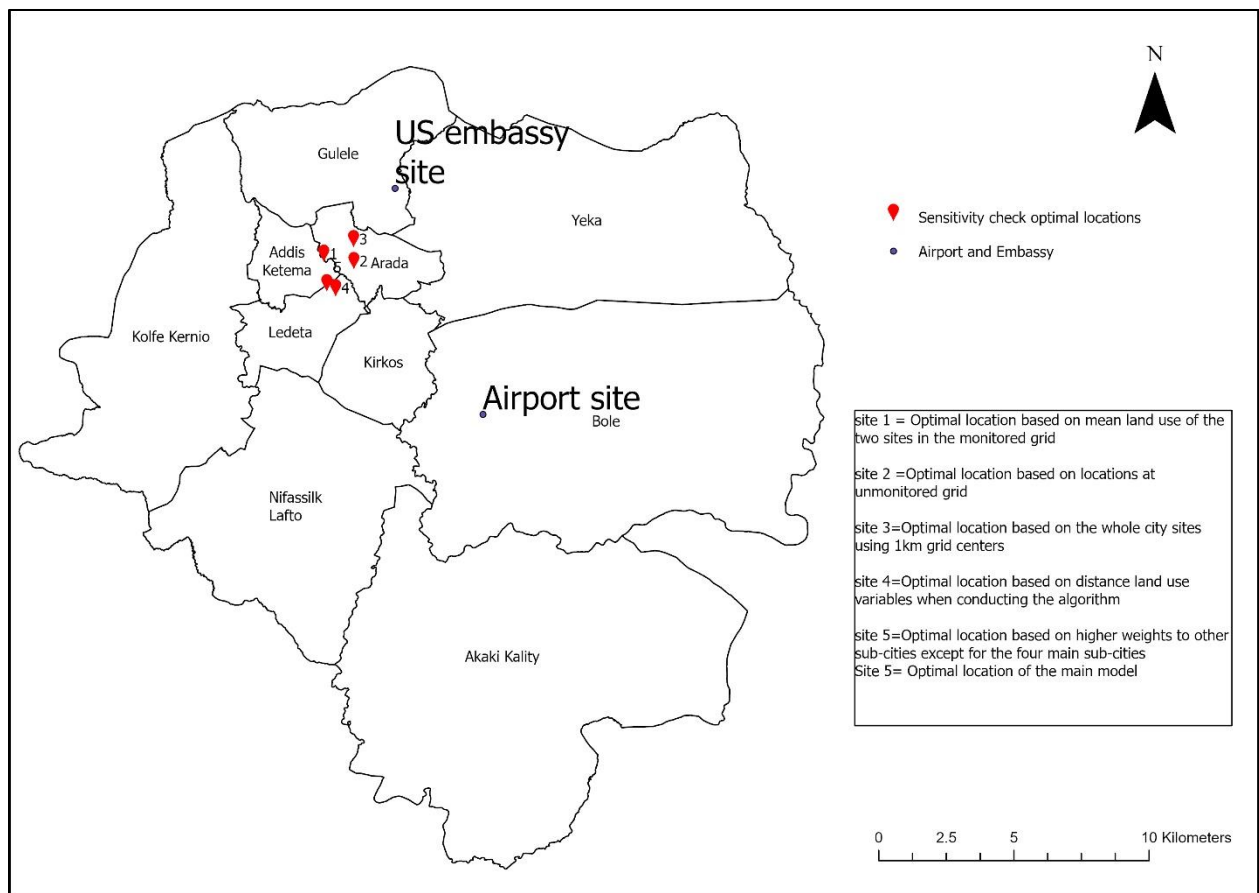
Sensitivity checks

We conducted five different sensitivity checks as a supplement to the main model; the results are shown in Table 3.A.7 and Figure 3.A.2.

1. We use the mean land use of the two sites in a grid. We calculate the average land use of the two sites in the monitored grid in a 100-meter buffer radius. To represent the two sites in this exercise, we used the midpoint of the two sites, and when calculating the distance from the fixed monitor, We used that midpoint.

2. For the second sensitivity check, we used locations in unmonitored grids. These grids were located between two consecutive monitored grids.

3. We selected sites at a center in a 1 km grid encompassing Addis Ababa. This was chosen due to its ability to capture land use variation in a previously unmonitored area, providing additional insight into air pollution levels in the wider region. During this exercise, we generated sites within a 1km grid using land use types measured within a 100m radius, which was consistent with the methodology used in the main model.

4. The fourth check involved changing the model specification, which altered the measurement type of the land use variables. As part of this sensitivity check, we used the distance to nearby land use types from monitoring sites in the regression and the algorithm.

5. The last check involved altering the weights outside the main sub-cities with higher population density to check the model's sensitivity to changes in population density. We gave weight higher than the four main sub-cities of the study area.

   Overall, the results of the entire sensitivity check were that all the sensitivity checks resulted in the optimal location for the fixed monitor in the four main sub-cities of the study area.

Table 3.A.7. Optimal locations of sensitivity check results

The method used to locate an optimal location	Latitude	Longitude	Description of its location	Distance from the current fixed monitor	It is in the four main sub-cities	Latitude sum of weighted discrepancy	Longitude sum of weighted discrepancy
The mean land use of the two sites in monitored grids is used	9.03444	38.74006	High-density mixed residence	3.7 km	Yes (Addis Ketema sub-city)	5.36	-0.03
Locations at the center of the unmonitored grid are used	9.032	38.75006	High-density mixed residence	3.29 km	Yes (Arada sub-city)	-1.51	2.00
Sites that include the whole of Addis Ababa in a 1km-by-1km grid at the center are used	9.0393	38.74996	High-density mixed residence	2.59 km	Yes(Arada sub-city)	0.92	1.744
Distance from monitored sites to the nearest land use is used as the main land use variable to do the algorithm	9.023	38.744	Asphalt road site	4.48 km	Yes (Lideta sub-city)	-1.37	-0.82
We gave more weights to other sub-cities in the analysis but not the four main sub-cities of the study area	9.024676	38.74106	High-density Mixed land use	4.48 km	Yes (Lideta sub-city)	-4.03	-0.14
Main Model result	9.024676	38.74106	High-density Mixed land use	4.48 km	Yes (Lideta sub-city)	-1.72	-1.35

 Note: Location of suggested fixed monitor location from the sensitivity analysis

Source: Author`s illustration using ArcGIS Pro

Figure 3.A.2 Map of sensitivity checks optimal locations.
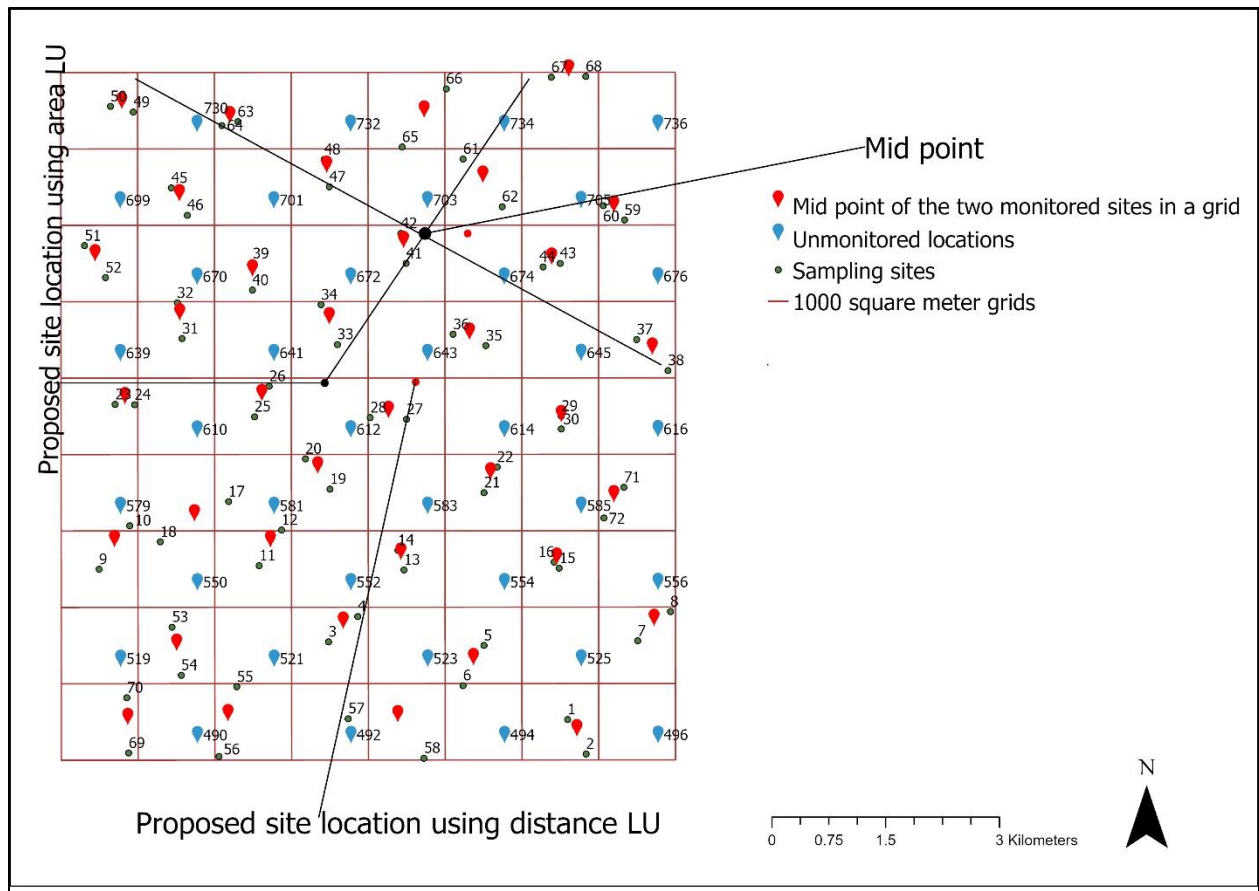
Note: The location of sites in unmonitored grids is shown in red, while the midpoint of the two monitored sites is shown in blue

Source: Author`s illustration using ArcGIS Pro

Figure 3.A.3 Location of sites in unmonitored grids and mid-point of the two monitored sites in a grid

Table 3.A.8 Weights of time and weather variables used to calculate the first-order conditions of

the discrepancy equation.

Variable	Observation	Mean (weight	Min	Max		Day	Weight
Rain	16,455	5.06687	0	104.43		Monday	0.14
Lag rain	16,454	5.067176	0	104.43		Tuesday	0.14
Wind speed	16,455	3.328075	1.21	8.34		Wednesday	0.14
Humidity	16,455	9.696551	2.32	13.12		Thursday	0.14
Temperature	16455	15.73346	10.12	21.12		Friday	0.14
						Saturday	0.14
Wind speed	Freq.	Percent	Cum.	Weight		Sunday	0.14
East	11,324	68.82	68.82	0.6882		Time of day	Weight
North	360	2.19	71.01	0.0219		Early morning (6:00-9:00)	0.1875
South	3,202	19.46	90.46	0.1946		Late morning) 10:00-12:00)	0.1875
west	1,569	9.54	100	0.0954		Early afternoon (13:00-15:00	0.1875
Total	16,455	100				Late afternoon 16:00-18:00)	0.1875
						Evening (19:00-22:00)	0.25

Weighting sub-cities for optimal location

The weight of sub-city s in Addis Ababa is given by

$$W_s = \frac{\text{Population density at subcity s}}{\text{Population density of Addis Ababa}} * \frac{\text{Area of sub-city s}}{\text{Total area of Addis Ababa}}$$

$$= \frac{\text{Total Population of sub-city s}}{\text{The Total area of sub-city s}} / \frac{\text{Total Population of AA}}{\text{Area of Addis Ababa}} * \frac{\text{Area of Sub-city s}}{\text{Area of Addis Ababa}}$$

$$= \frac{\text{Total Population Sub city s}}{\text{Total Population of AA}}$$

Since different sub-cities have different numbers of sites for each case in the sensitivity analysis then, for each site i in sub-city s, the weight can be calculated by

$$W_{is} = \frac{\text{Total Population Sub city s}}{\text{Total Population of AA}} * \frac{1}{ni}$$

Table 3.A.9. Calculated weight for the main model

Sub-city Name	Population in 2007	Land area in square km.	Population density per square km.	Number of Sites (ni)	W_s	W_{is} Main model
Addis Ketema	255,372	7.41	34,463.15	9	0.093217	0.010357
Arada	211,501	9.91	21,342.17	9	0.077203	0.008578
Bole	308,995	122.08	2,531.08	1	0.11279	0.037597
Gulele	267,624	30.18	8,867.10	8	0.097689	0.012211
Kirko	221,234	14.62	15,132.28	15	0.080756	0.005384
Kolfe Keranio	428,895	61.25	7,002.36	4	0.156557	0.039139
Lideta	201,713	9.18	21,973.10	12	0.07363	0.006136
Nefas Silk-Lafto	316,283	63.3	4,996.57	8	0.115451	0.014431
Yeka	346,664	85.46	4,056.44	4	0.12654	0.03163
Akaki Kality	181,270	118.08	1,535.14	0	0.066168	0
Total	2,739,551	521.47	5,253.51	72		

Table 3.A.10 Number of sites by sub-city for the main model and sensitivity analysis

Sub-city	No. sites, whole city case	No sites, unmonitored grid	No sites, main model
Akaki kality	142	0	0
Lideta	12	4	12
Arada	10	5	9
Kirkos	14	7	15
Addis Ketema	7	4	9
Gulele	35	2	8
Bole	127	1	3
Nefas silk lafto	64	7	8
Yeka	93	4	4
Kolfe keranio	74	2	4
Total	578	36	72

Table 3.A.11 Fixed effects estimation results of the difference

VARIABLES	PM2.5 Difference (No calibration)	PM2.5 Difference (Calibrated
Rain	-0.183	-0.154
	(0.605)	(0.569)
Lag rain	0.977	0.797
	(0.799)	(0.745)
Wind speed	-0.459	-0.531
	(0.517)	(0.482)
Humidity	-0.025	0.711
	(0.550)	(0.503)
Temperature	0.677***	0.696***
	(0.187)	(0.174)
High-density mixed residence	0.003***	0.003***
	(0.000)	(0.000)
River buffer	0.005***	0.005***
	(0.001)	(0.000)
Road network	-0.007***	-0.007***
	(0.001)	(0.001)
Open area	0.004***	0.004***
	(0.000)	(0.000)
Parking land use	-0.002***	-0.002***
	(0.001)	(0.000)
Commercial area	0.527***	0.495***
	(0.058)	(0.054)
Public facility	0.007***	0.006***
	(0.001)	(0.000)
Asphalt roads	0.157***	0.147***
	(0.019)	(0.017)
Unpaved roads	0.125***	0.117***
	(0.010)	(0.010)
North	-9.211**	-8.323*
	(4.352)	(4.186)
South	-5.756	-5.304
	(4.209)	(3.904)
West	-7.453*	-7.151*
	(4.213)	(4.001)
Late morning	21.628***	18.871***
	(3.197)	(2.934)
Early afternoon	24.000***	20.849***
	(3.225)	(2.961)
Late afternoon	22.906***	20.026***
	(3.605)	(3.316)

Evening	-2.293	-2.097
	(5.646)	(5.422)
Monday	-1.940	-1.762
	(5.007)	(4.746)
Tuesday	5.335	4.725
	(3.877)	(3.639)
Wednesday	0.299	-0.136
	(3.760)	(3.549)
Thursday	3.630	3.020
	(3.743)	(3.518)
Saturday	-5.454	-5.533
	(4.200)	(3.987)
Sunday	1.640	1.079
	(3.784)	(3.595)
Latitude difference	-174.696***	-147.228***
	(19.183)	(18.052)
Longitude difference	1,138.919***	1,023.700***
	(115.110)	(106.216)
Constant	-176.492***	-168.722***
	(17.068)	(15.843)
Site FE	Yes	Yes
Observations	10,462	10,462
R-squared	0.341	0.310

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.A.12 Discrepancy regression using only land use variables

| VARIABLES	PM2.5	PM2.5
High density mixed	3.917	1.969
	(3.532)	(3.023)
River buffer	-20.728*	-18.396*
	(11.725)	(9.834)
Open area	10.029**	8.208**
	(4.355)	(3.743)
Parking	27.686***	23.379***
	(6.555)	(5.189)
Commercial area	10.800**	7.839**
	(4.307)	(3.676)
Public facility	14.743**	12.740**
	(6.509)	(5.310)
Asphalt road	-503.938***	-447.350***
	(88.659)	(75.097)
Lattitude	353.208***	306.591***
	(39.034)	(33.841)
Longitude	139.930**	94.728*
	(60.709)	(51.891)
Constant	-29.071***	-22.675***
	(3.448)	(2.680)
Observations	10,462	10,462
R-squared	0.127	0.108

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Land use variables are expressed in percentages of total, the dependent variable in the first column is (PM2.5 fixed-PM2.5 Dylos) and the dependent variable in the second column is (PM2.5 fixed-PM2.5 Dylos claibrated)

Figure 4.A.1: Air quality and weather in Addis Ababa (2016-2022)
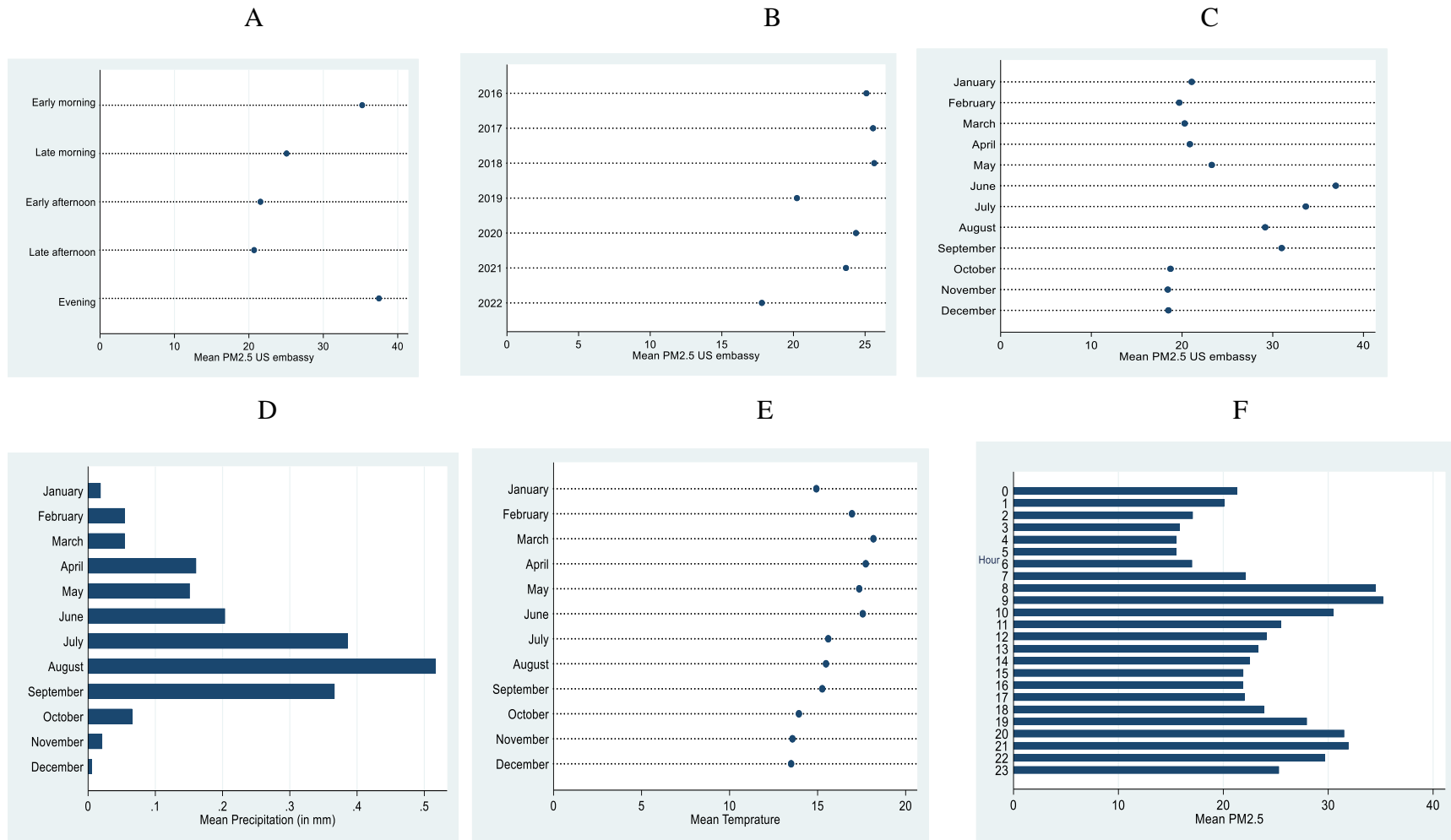
Note: All figures are for the years 2016-2022. Figure A, B, C, and F depict mean PM2.5 by the time of day, by year, by month, and by the hour, respectively. Figures D and E shows mean precipitation and mean temperature by month over the years 2016-202