UNIVERSITÉ JOSEPH KI-ZERBO **BURKINA FASO**

DOCTORAL SCHOOL OF INFORMATICS AND CLIMATE CHANGE

Unity-Progress-Justice

MASTER RESEARCH PROGRAM

SPECIALITY: INFORMATICS FOR CLIMATE CHANGE (ICC)

MASTER THESIS

Subject:

Fine particulate air pollution estimation in Ouagadougou using satellite aerosol optical depth and meteorological parameters

Presented July 2023 by:

Joe Adabouk AMOOLI

Major Supervisor

Prof. Daniel M. Westervelt

Co-Supervisor

Dr. Kwame Oppong Hackman

Academic Year 2022-2023

ACKNOWLEDGEMENTS

-I would like to thank the funders (The German Federal Ministry of Education and Research (BMBF)) and the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL) for offering me this master's research scholarship.

- I would like to thank the director (Prof. ZOUNGRANA Tanga Pierre) of "Ecole Doctorale Informatique et Changement Climatique (ED-ICC)" for his support during the course of the program.

- I would also like to thank the deputy director (Dr. COULIBALY Ousmane) of EDICC for his support, encouragement, and availability.

- I would also like to thank the scientific coordinator (Dr. ZOUNGRANA Benewindé) of EDICC for his guidance and feedback.

-A special thank you to my major supervisor (Prof. Daniel Westervelt) of Columbia University, U.S.A for his mentorship, availability, and continuous support in navigating challenges that arose during the research process.

-A big thank you to my co-supervisor (Dr. Kwame Hackman) of the WASCAL Competence Centre for his continuous guidance, suggestions, and support throughout the research process. –

-Many thanks to Dr. Bernard NANA of the Ecole Normale Supérieure for his guidance, and support in reviewing and providing feedback.

-I would also like to thank all my jury members for taking their time to evaluate and assess the quality of my thesis and providing feedback on areas that require improvement.

-I would like to thank the director (Prof. Ogunjobi Kehinde) of Research of the WASCAL Competence Centre for his contributions and for allowing me to benefit from his vast experience in air quality and aerosols meteorology.

-I would finally like to thank all my classmates for their continuous encouragement.

ABSTRACT

In this paper, $PM_{2.5}$ concentrations in Ouagadougou are estimated using satellite-based Aerosol Optical Depth and Meteorological Parameters. Firstly, Simple Linear Regression (SLR), Multiple Linear Regression (MLR), Decision Tree (DT), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) are developed using the available labeled data in the city. The XGBoost model outperforms all the models with a coefficient of determination (R^2) of 0.87 and a root-mean-square error (RMSE) of 15.8 µg/m^3. Given the outstanding performance of the supervised XGBoost model, it is upgraded by the incorporation of a semi-supervised algorithm to make use of the lots of unlabeled data in the city and allow for the extensive estimation of $PM_{2.5}$. The developed semi-supervised XGBoost model has an R^2 of 0.97 and an RMSE of 8.3 µg/m^3. The results indicate that the estimated $PM_{2.5}$ concentrations in the city are 2 to 4 times higher than the World Health Organization (WHO) 24-hour limit of 15 µg/m^3 in the rainy season and 2 to 22 times higher than the WHO 24-hour limit in the dry season. The results also reveal that the average annual estimated $PM_{2.5}$ concentrations are 11 to 14 times higher than the WHO average annual standard of 5 µg/m^3. Finally, the results reveal higher $PM_{2.5}$ concentrations in the center and industrial areas of the city compared to the other areas. There should be an improvement in traffic management in the central areas of the city and Industries should implement cleaner production methods.

Keywords: Air pollution; Fine Particulate Matter; Supervised Machine Learning; Fine Particulate Matter Spatial Distribution; Ouagadougou.

RÉSUMÉ

Dans cet article, les concentrations de $PM_{2,5}$ à Ouagadougou sont estimées à l'aide de la épaisseur optique des aérosols et des paramètres météorologiques par satellite. Tout d'abord, la régression linéaire simple (SLR), la régression linéaire multiple (MLR), l'arbre de décision (DT), la forêt aléatoire (RF) et l'amplification de gradient extrême (XGBoost) sont développées à l'aide des données étiquetées disponibles dans la ville. Le modèle XGBoost surpasse tous les modèles avec un coefficient de détermination (R^2) de 0,87 et une erreur quadratique moyenne (RMSE) de 15,8 µg/m^3. Compte tenu des performances exceptionnelles du modèle supervisé XGBoost, il est amélioré par l'incorporation d'un algorithme semi-supervisé pour exploiter les nombreuses données non étiquetées de la ville et permettre une estimation approfondie des $PM_{2,5}$. Le modèle XGBoost semi-supervisé développé a un R^2 de 0,97 et un RMSE de 8,3 µg/m^3. Les résultats indiquent que les concentrations estimées de $PM_{2,5}$ dans la ville sont 2 à 4 fois supérieures à la limite de 24 heures de l'Organisation mondiale de la santé (OMS) de 15 µg/m^3 pendant la saison des pluies et 2 à 22 fois supérieures à la limite de l'OMS 24 -limite d'heures en saison sèche. Les résultats révèlent également que les concentrations moyennes annuelles estimées de $PM_{2,5}$ sont 11 à 14 fois supérieures à la norme annuelle moyenne de l'OMS de 5 µg/m^3. Enfin, les résultats révèlent des concentrations de $PM_{2,5}$ plus élevées dans le centre et les zones industrielles de la ville par rapport aux autres zones. Il devrait y avoir une amélioration de la gestion du trafic dans les zones centrales de la ville et les industries devraient mettre en œuvre des méthodes de production plus propres.

Mots clés: Pollution de l'air; particules fines; Apprentissage automatique supervisé; Répartition spatiale des particules fines; Ouagadougou.

ACRONYMS AND ABBREVIATIONS

ANAM	:	Agence Nationale de la Météorologie
ANN	:	Artificial Neural Networks
AOD	:	Aerosol Optical Depth
BAM-1020	:	Met One Beta Attenuation Monitor
BC	:	Black Carbon
BiLSTM	:	Bidirectional Long Short-term Memory
BLH	:	Boundary Layer Height
BTEX	:	Benzene, Toluene, Ethylbenzene, and Xylene
CCD	:	Cold Cloud Duration
CHIRPS	:	Climate Hazards Group InfraRed Precipitation with Station data
CV	:	Cross-Validation
DNN	:	Deep Neural Network
DT	:	Decision Tree
EDXRF	:	Energy Dispersive X-ray Fluorescence
EMD	:	Empirical Mode Decomposition
GEE	:	Google Earth Engine
GWR	:	Geographic Weighted Regression
KNN	:	K-Nearest Neighbors
LST	:	Land Surface Temperature
MAE	:	Mean Absolute Error
MAIAC	:	Multi-Angle Implementation of Atmospheric Correction
MLR	:	Multiple Linear Regression
MODIS	:	Moderate Resolution Imaging Spectroradiometer
NO_2	:	Nitrogen Dioxide
PBLH	:	Planetary Boundary Layer Height
PM	:	Particulate Matter
PM_1	:	Particulate Matter with a diameter less than 1
PM_{10}	:	Particulate Matter with a diameter less than 10
$PM_{2.5}$	:	Fine Particulate Matter with a diameter less than 2.5

Precip	:	Precipitation
R^2	:	Coefficient of Determination
RF	:	Random Forest
RH	:	Relative Humidity
RMSE	:	Root-Mean-Square Error
RoI	:	Region of Interest
SLR	:	Simple Linear Regression
SO_2	:	Sulfur Dioxide
SVM	:	Support Vector Machine
Temp	:	Temperature
UNDP	:	United Nations Development Programme
US EPA	:	United States Environmental Protection Agency
USGCRP	:	United States Global Change Research Program
WD	:	Wind Direction
WHO	:	World Health Organization
WS	:	Wind Speed
XGBoost	:	eXtreme Gradient Boosting

SYNOPSIS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Background

Air pollution is an environmental risk affecting human health and has negative effects on the climate, biodiversity, and ecosystems. Increasing air quality will improve our environment and health, and aid development (Fisher et al., 2021). Nearly all of the world's population (99 %) breaths unhealthy air that exceeds World Health Organization (WHO) air quality standards. An estimated 7 million premature deaths annually are attributed to air pollution (WHO, 2021). In Africa in 2019, air pollution was the cause of 1.1 million deaths (WHO, 2021). Air pollution contributes to approximately 780,000 premature deaths annually in Africa (WHO, 2021). According to Cohen et al. (2005), fine particulate matter ($PM_{2.5}$) is the greatest health-relevant measure of urban air quality and is frequently used to determine international standards on air quality. $PM_{2.5}$ are particles with an aerodynamic diameter of 2.5 μm or less. The diameters of the bigger particles in the $PM_{2.5}$ size range would be approximately thirty times smaller than that of a human hair (Rushingabigwi et al., 2020).

The main sources of fine particles are the exhausts of cars, trucks, buses, and off-road vehicles (such as construction equipment and locomotives), as well as other processes that involve the burning of fuels like wood, heating oil, or coal, and natural sources like the dust storms from the Sahara Desert and forest fires (Rushingabigwi et al., 2020). In the atmosphere, fine particles can also result from the reaction of gases or droplets from sources like power plants. These chemical processes might take place hundreds of kilometers from the source of the pollutants (Rushingabigwi et al., 2020).

The 24-hour amount of $PM_{2.5}$ is taken into account when determining the health effects of exposure. The WHO guidelines exposure limit is 15 $\mu g/m^3$ for a 24-hour period and 5 $\mu g/m^3$ annually (WHO, 2021). According to the United States Environmental Protection Agency (US EPA), $PM_{2.5}$ at or below 12 $\mu g/m^3$ is regarded as healthy with little to no danger from exposure (US EPA, 2012). The air is deemed harmful if the quantity rises above 35 $\mu g/m^3$ over the course of a 24-hour period and can be problematic for persons who already have breathing problems (US EPA, 2012).

Despite the risks to public health posed by this condition, air quality measures have lately ceased or been halted, particularly in sub-Saharan Africa. The implementation of systematic PM data collection will enable the creation of programs to minimize the burden of air pollution on health as well as the formulation of urban planning and transportation policy in relation to air quality and health (Petkova et al., 2013). There is a link between daily ranges of particulate matter concentration and daily mortality, according to studies carried out in numerous locations around the world (Liu et al., 2019). Fine and ultrafine particulate matter appears to be linked to more severe diseases due to their ability to penetrate the deepest portions of the airways and more quickly enter circulation (Kelishadi & Poursafa, 2010).

Climate change and planetary warming might make things worse (USGCRP, 2009). Annual mean levels of coarse and fine particles published in the limited studies undertaken in Africa showed that pollution levels frequently exceed international recommendations (Petkova et al., 2013). Beyond its effects on health, PM has both cooling and warming effects on the planet's temperature (Solanki & Pathak, 2022), therefore research into this topic would help us better comprehend our climate system.

Problem Statement

Data on air quality are crucial for making decisions and assessing the effects of air quality on human health and the environment in the future. Data on air pollution would be greatly helpful in planning our cities, traffic patterns, and even in creating laws and regulations to lessen our carbon footprint (Babu Saheer et al., 2022). Few prior research in Ouagadougou have revealed that the city experiences excessive quantities of finer airborne particles, above WHO standards (Etyemezian et al., 2005; Lindén, 2011; Nana et al., 2012; Ouarma et al., 2020).

The major causes of air pollution in Ouagadougou are highly polluting traffic fleets with a high percentage of two-stroke motor vehicles, the widespread use of biomass burning for cooking, and frequently unregulated industry (Nana et al., 2012). Unpaved roads are a key source of road dust in the city during the dry season (Etyemezian et al., 2005). Another major source of airborne dust in the city is dust that has been transported from deserts and other arid regions (Etyemezian et al., 2005; Lindén, 2011). According to research, particulates in desert dust are hazardous to human health on a global scale. Particularly, the aerosolized dust from the desert is frequently linked to conditions like pneumonia and other health risks (Aili & Oanh, 2015). The urban climate

and air pollution in Ouagadougou are affected by seasonal changes in the regional climate (Lindén et al., 2012).

Despite the fact that particulate air pollution in Ouagadougou has more severe impacts, there are not many air monitoring stations. This is due to the limited financial means available to install these stations. Also, due to the few stations in the city, the study of the spatial distribution of $PM_{2.5}$ in the city is limited. In general, relatively few research on fine particle air pollution in Ouagadougou has been done due to a lack of data and monitoring stations. Understanding the correlation between satellite data, particularly AOD, meteorological parameters, and the concentration of $PM_{2.5}$ at the earth's surface in the city will help in developing models to improve the study of $PM_{2.5}$. However, surface $PM_{2.5}$ data in the city is sparse (small labeled data) but with lots of unlabeled data. Therefore, developing a model that makes use of the small labeled data (independent variables with corresponding $PM_{2.5}$ data) and the lots of unlabeled data (independent variables without corresponding $PM_{2.5}$ data) for estimating $PM_{2.5}$ would be very useful for extensive study of $PM_{2.5}$ and help in air quality decision-making in the city.

In this work, firstly, five models (simple linear regression (SLR), multiple linear regression (MLR), decision tree (DT), random forest (RF), and eXtreme Gradient Boosting (XGBoost)) would be built based on the available labeled data. The best-performing model would be upgraded by the incorporation of a semi-supervised algorithm to develop a semi-supervised model that makes use of the lots of unlabeled data with the small amount of labeled data in the city. The model would then be used with AOD and corrected satellite weather parameters at other locations with unavailable air monitoring stations in the city to estimate daily $PM_{2.5}$. The spatiotemporal variations of $PM_{2.5}$ in the city would be studied to understand their spatial distributions and growth over time. According to the literature research, Ouagadougou's air quality data challenges have not yet been solved using machine learning methods. Consequently, this is the initial effort.

Research Questions, Hypothesis, and Objectives

Research Questions

The main question behind this research is: **how can fine particulate air pollution concentrations in Ouagadougou be estimated?**

From the above main question, the following two (2) specific questions are obtained:

- **Specific 1**: How can we develop an effective model for estimating fine particulate air pollution from satellite-based aerosol optical depth and meteorological parameters using a small amount of labeled data and lots of unlabeled data in Ouagadougou?
- **Specific 2:** What are the Spatiotemporal variations of fine particulate air pollution concentrations in Ouagadougou?

Research Hypothesis

Based on the research questions, some hypotheses are made. So, each research question has a corresponding hypothesis.

- **Main**: Fine particulate air pollution concentrations in Ouagadougou can be estimated using satellite-based aerosol optical depth and meteorological parameters.
- **Specific 1:** An effective model can be developed for estimating fine particulate air pollution from satellite-based aerosol optical depth and meteorological parameters using a small amount of labeled data and lots of unlabeled data in Ouagadougou by the incorporation of a semi-supervised algorithm
- **Specific 2**: Fine particulate air pollution concentrations in Ouagadougou vary from season to season and are higher in the center and industrial areas.

Research Objectives

The objectives verify the stated hypotheses.

- **Main:** To estimate fine particulate air pollution concentrations in Ouagadougou using satellite-based aerosol optical depth and meteorological parameters.
- **Specific 1:** To develop an effective model for estimating fine particulate air pollution from satellite-based aerosol optical depth and meteorological parameters using a small amount of labeled data and lots of unlabeled data in Ouagadougou.
- **Specific 2:** To analyze the Spatiotemporal variations of fine particulate air pollution concentrations in Ouagadougou.

CHAPTER 1: LITERATURE REVIEW

Air pollution, especially $PM_{2.5}$, has become a pressing environmental and public health concern globally. In the case of Ouagadougou, rapid urbanization, industrial activities, and traffic emissions contribute to high levels of $PM_{2.5}$, which have adverse effects on both the environment and human health. The primary objective of this literature review is to critically examine previous studies that have employed satellite AOD data and meteorological variables to estimate $PM_{2.5}$ concentrations in general and in the context of Ouagadougou. The aim is to identify the strengths, limitations, and gaps in the existing knowledge and propose a methodology for accurate $PM_{2.5}$ estimation in the Ouagadougou.

1.1 Empirical Relationship Between AOD and $PM_{2.5}$

Several previous studies used the relationship between AOD and $PM_{2.5}$ to estimate $PM_{2.5}$ in areas without ground $PM_{2.5}$ monitoring stations (Wang & Christopher, 2003; Engel-Cox et al., 2004b; Gupta & Christopher, 2009; Kumar et al., 2007; Koelemeijer et al., 2006). Engel-Cox et al. (2004b) studied the relationships between MODIS satellite AOD readings and ground-based $PM_{2.5}$ measurements covering the period from 1 April to 30 September 2002 by calculating the correlation coefficient (r). The inverse of the amount of cloud cover was utilized to weight each observation's influence in each of their analyses to calculate the correlation. They discovered that correlations between ground-based particulate matter and MODIS aerosol optical depth were greater in the eastern and Midwest regions of the United States (east of 100° W). Also, they found that although the correlation was location-dependent, data in the western US were spotty and showed worse correlations. This fluctuation was caused by a mix of topography variability, regression artifacts, variations between the ground-based and column average datasets, and MODIS cloud mask and aerosol optical depth algorithms. They came to the conclusion that using satellite sensor data, such as that from MODIS, could significantly improve the monitoring of air quality at the regional and synoptic scales.

In Europe in 2003, Koelemeijer et al. (2006) compared the spatiotemporal variations of $PM_{2.5}$ with those of AOD observed by the MODIS satellite instrument. They discovered that the primary aerosol-generating locations in Northern Italy, Southern Poland, and the Belgium/Netherlands/Ruhr region, as well as specific significant towns and industrialized valleys,

were all clearly visible in the MODIS readings. They also discovered that there were clear differences between AOD and PM's seasonal variation. They discovered that the AOD, as determined by MODIS, clearly decreased in the winter across the majority of Europe. In different ways throughout Europe, the seasonal fluctuation in $PM_{2.5}$ was less pronounced than the AOD at various locations. As a result, there was little association between AOD and $PM_{2.5}$ over a year. When the AOD was split by the boundary layer height and, to a lesser extent, when it was corrected for aerosol growth with relative humidity, the association between PM and AOD was improved. They discovered a 0.6 correlation on average between $PM_{2.5}$ and the AOD. They came to the conclusion that satellite AOD observations could help with better tracking of $PM_{2.5}$ dispersion across Europe.

Also, Wang & Christopher (2003) investigated the relationship between hourly fine particulate mass ($PM_{2.5}$) measured at the surface at seven locations in Jefferson county, Alabama for 2002 and column AOD derived from the MODIS on the Terra/Aqua satellites. Their findings showed that the satellite-derived AOD and $PM_{2.5}$ had a strong association ($r = 0.7$), indicating that the majority of the aerosols were in the well-mixed lower boundary layer during the satellite overpass times. Also, they discovered that there was a strong agreement ($r > 0.9$) between the monthly mean $PM_{2.5}$ and MODIS AOD, with summer seeing the highest values due to increased photolysis. They found that due to higher traffic volume and constrained mixing depths in the morning (6:00–8:00 AM), $PM_{2.5}$ exhibits a distinct diurnal character. Using simple empirical linear relationships derived between the MODIS AOD and daily $PM_{2.5}$, they demonstrated that the MODIS AOD can be used quantitatively to estimate air quality categories as defined by the US EPA with an accuracy of more than 90 % in cloud-free conditions.

Furthermore, Hutchison (2003) used the relationship between satellite AOD and surface $PM_{2.5}$ to examine which continental haze from the northeast migrated into Texas and necessitated the issuance of health advisories for 150 counties in Texas. Also, they illustrated the limitations of using only ground-based observations to monitor air quality across Texas. These drawbacks included gradients in pollution concentration that depend on the location of the point source, the meteorology governing its transport to Texas, and its diffusion across the region, as well as the size of State borders, which can only be monitored with a large number of ground-based sensors. Their results demonstrated the capability of MODIS data and products to identify and monitor the

movement of pollutants. They came to the conclusion that MODIS serves as the foundation for creating advanced data products that, when combined with ground-based observations, will produce an efficient and precise pollution monitoring system for the whole state of Texas.

Moreover, Kumar et al. (2007) also examined the relationship between the AOD, estimated from satellite data at 5 km spatial resolution, and the mass of $PM_{2.5}$, monitored on the ground in Delhi Metropolitan. AOD and $PM_{2.5}$ were significantly positively correlated, according to their findings. They examined the time-space dynamics of air pollution in Delhi by using the relationship to predict surface air quality for past years.

In addition, van Donkelaar et al. (2010b) estimated global ambient fine particulate matter concentrations from satellite-based AOD. They discovered that estimates of long-term average $PM_{2.5}$ concentrations (1 January 2001 to 31 December 2006) at a resolution of roughly 10 km x 10 km pointed to a geometric mean $PM_{2.5}$ concentration of 20 $\mu g/m^3$ for the entire world that was population-weighted. According to their research, 38 % and 50 % of central and eastern Asia, respectively, had $PM_{2.5}$ levels over the World Health Organization's Interim Target-1 for Air Quality (35 $\mu g/m^3$ yearly average). Throughout eastern China, annual mean $PM_{2.5}$ concentrations were greater than 80 $\mu g/m^3$. They found significant geographic agreement between the satellite-derived estimate and ground-based in situ measurements ($r = 0.77$; slope = 1.07) as well as between the satellite-derived estimate and noncoincidental observations elsewhere ($r = 0.83$; slope = 0.86). The AOD retrieval, along with uncertainties in the aerosol vertical profile and sampling, was used to estimate the standard deviation of uncertainty in the satellite-derived $PM_{2.5}$, which was calculated to be 25 %. The average global uncertainty, population-weighted, was 6.7 $\mu g/m^3$.

Léon et al. (2021a) studied the correction between AOD and surface $PM_{2.5}$ at Cotonou, Benin, and Abidjan, Côte d'Ivoire and found a weekly $r = 0.75$ between mean AOD and measured $PM_{2.5}$. They used the relationship between the two variables to analyze the seasonal variability of Surface $PM_{2.5}$ from 2003-2019. Xue et al. (2017) constructed a three-stage model with a spatial resolution of 0.1° to estimate the daily $PM_{2.5}$ over China using data from satellite-derived AOD, and ground observations of $PM_{2.5}$. The cross-validation (CV) approach was used to gradually assess the performance of the three-stage model. Their findings demonstrated that there was good agreement between the fused estimator of $PM_{2.5}$ and the observational data (root- mean-square error (RMSE)= 23.0 $\mu g/m^3$ and Coefficient of determination (R^2)= 0.72).

The limitation of the above methods is that they did not take into account the meteorological influences on the AOD-$PM_{2.5}$ relationship but depending on the location, a simple linear regression can be a useful technique in explaining the variations of $PM_{2.5}$.

1.2 Combining AOD and Meteorological Data for PM2.5 Estimation

The relationship between the satellite-based AOD and air pollution monitoring on the ground can be influenced by a number of meteorological factors such as temperature, relative humidity, Precipitation, wind speed, and wind direction (Paciorek & Liu, 2010; Song et al., 2014; Kumar, 2010; Hu et al., 2013; Zheng et al., 2017). Hence using the AOD-$PM_{2.5}$ relationship alone for estimating $PM_{2.5}$ may not be able to explain accurately the variations of ground $PM_{2.5}$ concentrations (Kumar, 2010; Wang et al., 2019). Hence, many researchers have proposed new statistical and machine learning techniques using $PM_{2.5}$, AOD, and meteorological parameters to improve the estimation of surface $PM_{2.5}$.

1.2.1 Statistical Methods

Tian & Chen (2010) developed a semi-empirical model using MODIS AOD, $PM_{2.5,}$ and meteorological parameters including specific humidity, air pressure, air temperature, and boundary layer height (BLH) to estimate, at a regional level, the hourly concentration of ground-level $PM_{2.5}$ concurrent with satellite overpass. Their model was able to explain 65 % of the variation in ground-level $PM_{2.5}$ concentrations Their estimated mass concentrations of $PM_{2.5}$ were significantly correlated with the actual readings. Their model had an RMSE of 6.1 $\mu g/m^3$. It was discovered that adding ground-level temperature and relative humidity significantly increased the model's predictability.

Similarly, Gharibzadeh & Saadat Abadi (2022) used AOD data along with several effective meteorological variables such as temperature, relative humidity, wind speed, wind direction, and horizontal visibility to develop a multivariable linear regression model to estimate $PM_{2.5}$ concentrations over Ahvaz, Iran. Their results of the multivariable linear regression model showed that the model could predict 60 % of $PM_{2.5}$ changes in Ahvaz.

Furthermore, Wang et al. (2010) used a MODIS AOD, $PM_{2.5}$, and humidity correcting method to develop a linear model for estimating surface $PM_{2.5}$ in Beijing. The correlation between AOD and $PM_{2.5}$ improved with the R^2 increasing from 0.35 to 0.66 as a result of the humidity correction. The correlation between the satellite-estimated and $PM_{2.5}$ with the measurements was

R^2 = 0.47, and the bias was 6.49 %, according to validation against the in-situ measurements in Beijing. The R^2 between the estimated $PM_{2.5}$ and the observations increased to 0.66 when averaged over Beijing's urban region. Their findings suggested that the MODIS data may be used to monitor local air pollution by utilizing the relative humidity adjusting approach.

Benas et al. (2013) developed an MLR model for estimating $PM_{2.5}$ over the broader urban area of Athens, Greece using In-situ $PM_{2.5}$ measurements, MODIS AOD and related parameters; surface temperature and relative humidity. They evaluated each satellite-derived parameter's contribution and the effectiveness of linear relationships on a stepwise validation. The model's R^2 was roughly 0.7. They also calculated the seasonal mean $PM_{2.5}$ distributions, which showed an intra-annual fluctuation with greater values in the summer along with variances in $PM_{2.5}$ concentrations related to air pollution at the city center and at industrial districts.

Ma et al. (2014) estimated ground-level $PM_{2.5}$ from satellite-derived aerosol optical depth (AOD) and meteorological parameters in China using a spatial statistical model; geographic weighted regression (GWR) model. Their findings indicated that the performance of the model can be significantly enhanced by the meteorological data and land use information. Their model had an R^2 of 0.64 and an RMSE of 32.98 $\mu g/m^3$.

Also, Zhai et al. (2019) used the stepwise MLR model to relate $PM_{2.5}$ variations across China to relative humidity, temperature, wind speed, precipitation, and meridional velocity at 850 hPa (V850) as predictor variables. Their model explained about 50 % of the variance of surface PM2.5 concentrations, including 41-65 % for the five megacity clusters. Application to the $PM_{2.5}$ time series revealed that 6-year trends across China and in the megacity clusters were considerably influenced by meteorological variability. Removing meteorological variations as given by the MLR model also reduced the 235 uncertainty in the trend that could have been attributed to emission limits.

Additionally, in mainland China in 2019, He et al. (2021) clearly investigated the relationship between $PM_{2.5}$ and AOD and its potential influence elements, such as climatic variables and terrain. They found that stronger spatial correlations were mostly seen in northern and eastern China and that the linear 25 slope in the northern inland regions was often higher than that in other locations. They also discovered that, temporally, the $PM_{2.5}$-AOD association was most

pronounced in the winter and peaked in the midday and afternoon. They discovered that the $PM_{2.5}$-AOD correlation can be improved by taking both relative humidity and planetary boundary layer height (PBLH) into account. They observed large correlations between 400 and 600 meters, mainly in Sichuan, Shanxi, and Junggar basins. They came to the conclusion that in the majority of domains, such as the Tibetan Plateau and north-central China's Qinghai and Gansu, the Multi-Angle Implementation of Atmospheric Correction (MAIAC) 1-km 30 AOD can better represent the ground-level fine particulate matter.

Kanabkaew (2013) Used The relationship between AOD and hourly $PM_{2.5}$ over Chiangmai to develop an MLR model with ground-based meteorological observations. Their results revealed that the correlation between AOD and hourly $PM_{2.5}$ was improved significantly when corrected with relative humidity and temperature data. The model had an R^2 of 0.77. They then applied the model to smog data over Chiangmai in March 2007. The model performed reasonably with R^2 of 0.74. They concluded that the model applications would offer supplemental information to other regions with comparable circumstances but without air quality monitoring stations and reduce false alarms regarding the degree of air pollution associated with smog from extensive biomass burning.

1.2.2 Machine Learning Methods

In contrast to the traditional statistical methods, prediction technologies based on machine learning approaches have been shown to be the most effective instruments for studying air pollution (Joharestani et al., 2019; Kumar & Pande, 2022). Joharestani et al. (2019) used AOD, satellite and meteorological data, ground-measured $PM_{2.5}$, and geographical data to develop random forest, extreme gradient boosting, and deep learning machine learning approaches for $PM_{2.5}$ prediction in Tehran's urban area. The XGBoost model had the best performance obtained with $R^2 = 0.81$, Mean absolute error (MAE) = 9.93 $\mu g/m^3$, and RMSE = 13.58 $\mu g/m^3$. Kumar & Pande (2022) developed five machine Learning models, k-nearest neighbors, Gaussian Naive Bayes, Support Vector Machine, Random Forest, and XGBoost to investigate six years of air pollution from 23 Indian cities. They compared the output of their models to the accepted measures. The Support Vector Machine model had the lowest accuracy, while the Gaussian Naive Bayes model showed the highest accuracy. Also, when the performances of their models were assessed and contrasted through established performance parameters. The XGBoost model outperformed the models and achieved the highest linearity between predicted and observed data.

Also, they observed that in the pandemic year of 2020, there was a considerable decrease in practically all pollutants.

Also, McFarlane et al. (2021) developed an MLR and a random forest model using data from the Met One Beta Attenuation Monitor-1020 (BAM-1020) at the United States (U.S.) Embassy in Kampala, Uganda, along with data from low-cost $PM_{2.5}$ monitor, the PurpleAir, and meteorological parameters to predict corrected $PM_{2.5}$ concentration. The MLR and the random forest model had R^2 of 0.96 and 0.86 respectively and MAE of 3.4 $\mu g/m^3$ and 5.8 $\mu g/m^3$ respectively. Similarly, Lin et al. (2022) developed a Random Forest and eXtreme Gradient Boosting (RF-XGBoost) for estimating ground-level $PM_{2.5}$ in Guanzhong Urban Agglomeration using the MODIS AOD product, high density meteorological and topographic conditions, land use, population density, and air pollutions. The RF-XGBoost model had a good performance with an R^2 of 0.93 and an RMSE of 12.49 $\mu g/m^3$. The Guanzhong Urban Agglomeration had the worst pollution in the winters of 2018 and 2019, according to the RF-XGBoost model's output, as a result of the burning of coal for heating and bad climatic conditions. In the winter of 2019, they also observed that the air pollution situation remained dire, with more than 65 % of the study region meeting the mean $PM_{2.5}$ levels higher than 35 $\mu g/m^3$ and the maximum reaching 95.57 $\mu g/m^3$.

Similarly, Gupta et al. (2021) developed a random forest model for estimating $PM_{2.5}$ in Thailand using NASA's Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2) reanalysis data of aerosols and meteorology. For the validation data sets, the model had an R^2 between observed and estimated $PM_{2.5}$ that varied between 0.88 and 0.92 and an RMSE that ranged from 8.5 $\mu g/m^3$ to 10.5 $\mu g/m^3$ for the validation data spanning 10-folds. They also noted that the model underpredicted $PM_{2.5}$ levels at the hourly scale under extremely clean conditions ($PM_{2.5}$ <10 $\mu g/m^3$) and overpredicted $PM_{2.5}$ levels during excessive loading ($PM_{2.5}$ > 80 $\mu g/m^3$). The authors also observed that the daily mean $PM_{2.5}$ (24-hour) values closely mirrored day-to-day fluctuation, with high values observed during the winter months (November-February) and lower values observed during other seasons. They concluded that the trained model could reprocess the regional MERRA2 time series, and the bias-corrected data may be applied to other tasks like long-term trend analysis and health exposure research.

Using $PM_{2.5}$ data collected from 29 stations across Malaysia and MODIS AOD and weather data, Kamarul Zaman et al. (2017) developed MLR and artificial neural networks (ANN) models for estimating $PM_{2.5}$. In comparison to the MLR technique, which had R^2 = 0.66 and RMSE = 12.39 μg/m^3, their findings showed that the ANN using MODIS AOD_{550} provides superior accuracy with R^2 = 0.71 and RMSE = 11.61 μg/m^3. Also, The MODIS AOD_{550} was the key parameter for $PM_{2.5}$ according to their stepwise regression analysis done on the MLR approach. The RMSE was 13.61 μg/m^3 and R^2 was 0.59. They also discovered that the addition of meteorological parameters improved their model performance. Chen et al. (2018) used daily ground-level $PM_{2.5}$ measurements obtained from 1,479 stations across China from 2014-2016 with data on AOD and meteorological data to develop random forests model for estimating daily $PM_{2.5}$ concentrations. The daily random forest model has a substantially greater RMSE of 28.1 μg/m^3, and R^2 = 83 %, explaining the majority of the spatial variability in daily $PM_{2.5}$. They found that the model explained up to 86 % of the variation in average $PM_{2.5}$ at the monthly and annual time scales. They came to the conclusion that the machine learning method had better prediction capacity than earlier research when using the modeling framework and the most recent ground-level $PM_{2.5}$ measurements.

Additionally, Using ground-level data collected from 19 stations across China between 2017 and 2019 along with air temperature, specific humidity, sea level pressure, and wind speed, Fu et al. (2022) investigated the relationships between surface $PM_{2.5}$ and AOD and developed a random forest model for predicting $PM_{2.5}$. At 14 of the 19 sites, they discovered that specific humidity predominated the associations with normalized $PM_{2.5}$-AOD differences. They also observed that a low r of 0.49 was obtained between the predicted and observed $PM_{2.5}$ concentrations when only AOD was used as input in the random forest model. When they added specific humidity into their model, the r increased to 0.74, which was close to the r of 0.81 with three additional weather parameters. Their research demonstrated a significant decoupling between $PM_{2.5}$ and AOD and recommended considering humidity as a crucial factor in China for retrieving long-term $PM_{2.5}$ using AOD data.

Zhang et al. (2021) combined satellite AOD, meteorology, land use, and socioeconomic data, to develop a random forest model for estimating daily $PM_{2.5}$ concentrations at 1 km^2 resolution in and around Gauteng Province, South Africa. The daily readings from their model had

an overall cross-validation R^2 of 0.80 and an RMSE of 9.40 $\mu g/m^3$, indicating a satisfactory fit between model estimations and ground measurements.

Recently, many scientists have begun incorporating semi-supervised algorithm into their developed machine learning models for extensive $PM_{2.5}$ studies and enhancement of model's performance. Zhao et al. (2023) proposed a co-trained semi-supervised learning model combining the K-nearest neighbors (KNN) algorithm and deep neural network (DNN) in order to maximize the utilization of unlabeled samples and enhance the model's performance for fine-grained air quality analysis in China. Their model outperformed other models with a coefficient of determination between the predicted and true values of 0.98. For the purpose of estimating $PM_{2.5}$ concentrations in China, Zhang et al. (2021) developed a semi-supervised model. Their strategy used bidirectional long short-term memory (BiLSTM) neural networks and empirical mode decomposition (EMD). Their findings showed that the semi-supervised model, which had an RMSE of 6.8 $\mu g/m^3$ and an R^2 of 0.97, was more accurate than the other conventional LSTM-based model.

Similarly, a semi-supervised learning algorithm for forecasting $PM_{2.5}$ concentrations in Northeastern China was proposed by Jiang et al. (2022). Rich real-world data from 11 air quality monitoring stations in Shenyang and surrounding cities, meteorological data, and spatiotemporal information were all incorporated into the model. The experimental results demonstrated that the proposed model outperforms baseline methods in accuracy by 3 % to 18 % over a traditional multivariate linear regression, 1 % to 11 % over an MLR-ANN, and 21 % to 68 % over a support vector machine (SVM).

1.3 Previous $PM_{2.5}$ Studies in Ouagadougou

Etyemezian et al. (2005) assessed the in-situ measurements of suspended particulate matter concentrations during two campaigns to determine the airborne particle pollution levels in Ouagadougou. They used a portable instrument (AEROCET531S) to measure PM_1, $PM_{2.5}$, and PM_{10} at nine sites in 2018 and ten sites in 2019. They chose roadsides, in administrative facilities, secondary schools, and distant districts as their sites for the study. Their results showed that the PM_1 concentrations had no significant variation between days, seasons or sampling sites. Their findings demonstrated that there was no appreciable difference in PM_1 concentrations between days, seasons, or sampling locations. They noticed that the 24-hour $PM_{2.5}$ concentrations

frequently exceeded the levels advised by the WHO. Additionally, they noticed that during the dry season, $PM_{2.5}$ concentrations were higher than they were during the rainy season.

Also, Boman et al. (2009) determined the mass, black carbon (BC), elemental concentrations, and fluctuations in $PM_{2.5}$ at two sites in Ouagadougou. They performed their elemental analysis using energy dispersive x-ray fluorescence (EDXRF) spectroscopy. In majority of the samples, they identified Chlorine (Cl), Potassium (K), Calcium (Ca), Titanium (Ti), Manganese (Mn), Iron (Fe), Copper (Cu), Zinc (Zn), Bromine (Br), Rubidium (Rb), Strontium (Sr), and Lead (Pb). The particle mass concentration ranged from 27 $\mu g/m^3$ to 164 $\mu g/m^3$, and the BC ranged from 1.3 $\mu g/m^3$ to 8.2 $\mu g/m^3$. Additionally, they discovered that leaded gasoline had no impact on the particle concentrations. By comparing their findings to the elements in a soil sample, they were able to identify soil dust as a key component of the particles.

Furthermore, Lindén (2011) investigated the characteristics of the connections between Ouagadougou's metropolitan climate and air pollution. The author focused on the impacts of various land cover types while examining spatial differences in daily temperature and humidity trends during the early dry season. Additionally, the author looked at how atmospheric stability affected the patterns of intra-urban air temperature, the urban wind field, and the spatial variations of air pollution levels. The author assessed weather and air quality characteristics at set locations and while driving through regions with varying land use, activity, traffic density, and road surface. According to the author, the strongest intra-urban nocturnal cool islands in vegetated regions were responsible for Ouagadougou's thermal patterns. These islands were brought about by the vegetation's evaporative cooling throughout the nighttime hours. The author also observed that Ouagadougou's air pollution situation was marked by significant spatial variations, high pollution levels overall, and extreme levels of coarse particles, frequently exceeding WHO air quality guidelines in all areas. Important sources included transported dust, traffic, biomass burning, and re-suspension of road dust.

Moreover, Nana et al. (2012) quantified the concentrations of Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), BTEX (Benzene, Toluene, Ethylbenzene, and Xylene), and PM_{10} in Ouagadougou. Their findings showed that, aside from downtown, where levels were frequently above the standard, NO_2 concentrations in the city continued to be below the WHO-set limit. In general, the city's average SO_2 concentrations remained low. They also discovered that the city

had high quantities of BTEX. Additionally, they discovered that the levels of PM_{10} in the city frequently surpassed WHO guidelines. They noted that the daily PM_{10} concentrations in Ouagadougou exhibited two peaks, the first at roughly 7 a.m. and the second at about 7 p.m. They also studied the seasonal variations of $PM_{2.5}$ and found that the concentrations of PM_{10} were higher in the dry season and lower in the wet season. According to them, the PM_{10} was primarily made up of desert dust and dust that was re-suspended due to vehicle movement on unpaved roads.

Additionally, Ouarma et al. (2020) analysed PM_1, $PM_{2.5}$ and PM_{10} concentrations by hour, day and location in dry and rainy season 2018 and rainy season 2019 in Ouagadougou using measurements from the analyzer AEROCET 531S. They chose academic sites, sites in outlying neighborhoods, industrial sites, roadside locations with heavy traffic on paved roads, and administrative sites for the analysis. They observed that all the sites have PM concentrations that exceeded both the WHO and the US EPA guidelines except the academic site which was close to the required standard. They also found that combustion activities and resuspension processes had a substantial impact on the roadside, administrative, peripheral district, or industrial, leading to comparatively higher concentrations of $PM_{2.5}$ and PM_{10}. Traffic exhaust emission, which is a substantial source of fine and ultra-fine particles, had an impact on the traffic proximity locations. They also observed the PM concentrations in the dry season were higher than in the rainy season.

Overall, the literature review reveals that previous studies have successfully employed satellite AOD data and meteorological variables to estimate $PM_{2.5}$ concentrations in different regions. It has been observed that less attention is given to using statistical and machine learning techniques for estimating $PM_{2.5}$ in Ouagadougou. Also, not much attention is given to using satellite AOD and weather parameters for $PM_{2.5}$ estimation in the city. The long-term spatiotemporal variability of $PM_{2.5}$ in the city has not been explored. The main objective of this work is to use satellite AOD and meteorological parameters by applying statistical and machine learning techniques to develop an effective model for estimating $PM_{2.5}$ in the city of Ouagadougou.

CHAPTER 2: MATERIALS AND METHOD

The methodology section of this study outlines the approach used to estimate $PM_{2.5}$ concentrations in Ouagadougou. The estimation was based on the integration of satellite AOD with meteorological parameters. Two types of linear regression models (simple and multiple) and three nonlinear regression models (decision tree, random forest, and XGBoost) were developed based on the small amount of labeled data available in Ouagadougou to establish the relationship between the selected parameters and $PM_{2.5}$ levels. The XGBoost model based on its outstanding performance was upgraded by incorporating a semi-supervised algorithm to use the lots of unlabeled data in Ouagadougou. The meteorological parameters included temperature, wind speed, wind direction, relative humidity, and precipitation. These parameters were chosen based on their known influence on $PM_{2.5}$ concentrations, as documented in previous research.

2.1 Study Area

Burkina Faso is currently showing fast progress in human development, climbing from the 2nd lowest human development index in the world to the 9th position between 1970 and 2021 (UNDP, 2021). One effect of this development is the extremely rapid growth of urban areas. The capital, Ouagadougou (12°22 N, 1°31 W, 300 m above sea level), has grown from 800 000 inhabitants in the year 2000 to approximately 2.8 million in 2022 (UN, 2022). As a result of this urbanization, the urban area is expanding rapidly, with the proportionally fastest growth in the form of informal spontaneous settlements on the outskirts of the city. Informal spontaneous settlements in Ouagadougou have grown approximately 60 % since 2004, while the planned residential areas have grown approximately 30 % at the same time (Lindén et al., 2012). The growth in total length of paved roads has increased over the years, with the paving mainly taking place in high-income neighborhoods while the majority of the residential areas remain completely unpaved (Lindén et al., 2012).

Ouagadougou is considered one of the most polluted cities in Africa with its 24-hour $PM_{2.5}$ concentrations exceeding two to three times the WHO recommended (Ouarma et al., 2020). The city's particulate matter is primarily made up of desert dust, uncontrolled industry, extensive biomass burning, heavy road traffic, and dust particles that have been re-suspended due to vehicle movement on unpaved roads (Boman et al., 2009; Ouarma et al., 2020; Nana et al., 2012).

Ouagadougou is situated in the Sahel region's hot, semi-arid steppe climate. The weather is divided into two distinct seasons: a dry season from November to April, with an average of less than 100 mm of precipitation, and a wet season from May to October, with an average of 700 mm of precipitation (Agence Nationale de la Météorologie, ANAM 2007; Ouarma et al., 2020). The harmattan, which originates from the Sahara in the north, dominates the winds throughout the dry months (ANAM, 2007). The average yearly temperature is 29 °C. From March 8 to May 16, the hot season, with an average daily high temperature exceeding 38 °C, lasts for 2.3 months. In Ouagadougou, April is the hottest month of the year, with an average high of 39 °C and a low of 28 °C. From July 9 to September 18, the cool season, which has an average daily high temperature below 32.2 °C, lasts for 2.3 months. With an average low of 18 °C and a high of 33 °C, January is the coldest month of the year in Ouagadougou. Extreme seasonal variations in perceived humidity can be found in Ouagadougou. The 6.9-month period from April 9 to November 5 known as the "muggier season" is when the humidity is at its highest and is at least 25 % of the time uncomfortable. During the evenings and at night, wind speeds are often quite low, and atmospheric stability is high, which indicates that there is little ventilation of the urban air (Lindén, 2011).

Over the course of the year, Ouagadougou's average hourly wind speed shows significant seasonal variation. From November 26 to June 28 there are 7.1 months that are windier than others, with average wind speeds exceeding 2.86 m/s. With an average hourly wind speed of 3.76 m/s, January is the windiest month in Ouagadougou throughout the year. June 28 to November 26 is when things are most tranquil. In Ouagadougou, September is the calmest month of the year, with an average hourly wind speed of 1.97 m/s (ANAM, 2007).

The city is more polluted in the dry season than in the rainy season with $PM_{2.5}$ concentrations varying from area to area (Nana et al., 2012; Ouarma et al., 2020). $PM_{2.5}$ concentrations in the industrial and administrative areas of the city are highly influenced by combustion (Ouarma et al., 2020). Kossodo and Gounghin are the major industrial areas of the city and are home to several processing plants and factories (Gouba et al., 2021). In terms of road transportation in the city, the breakdown of the vehicle distribution is as follows; Motorized two-wheeled vehicles make up 74 % of the fleet, followed by private cars (18 %), buses (7 %), and heavy trucks (1 %) (Somda, 2018).

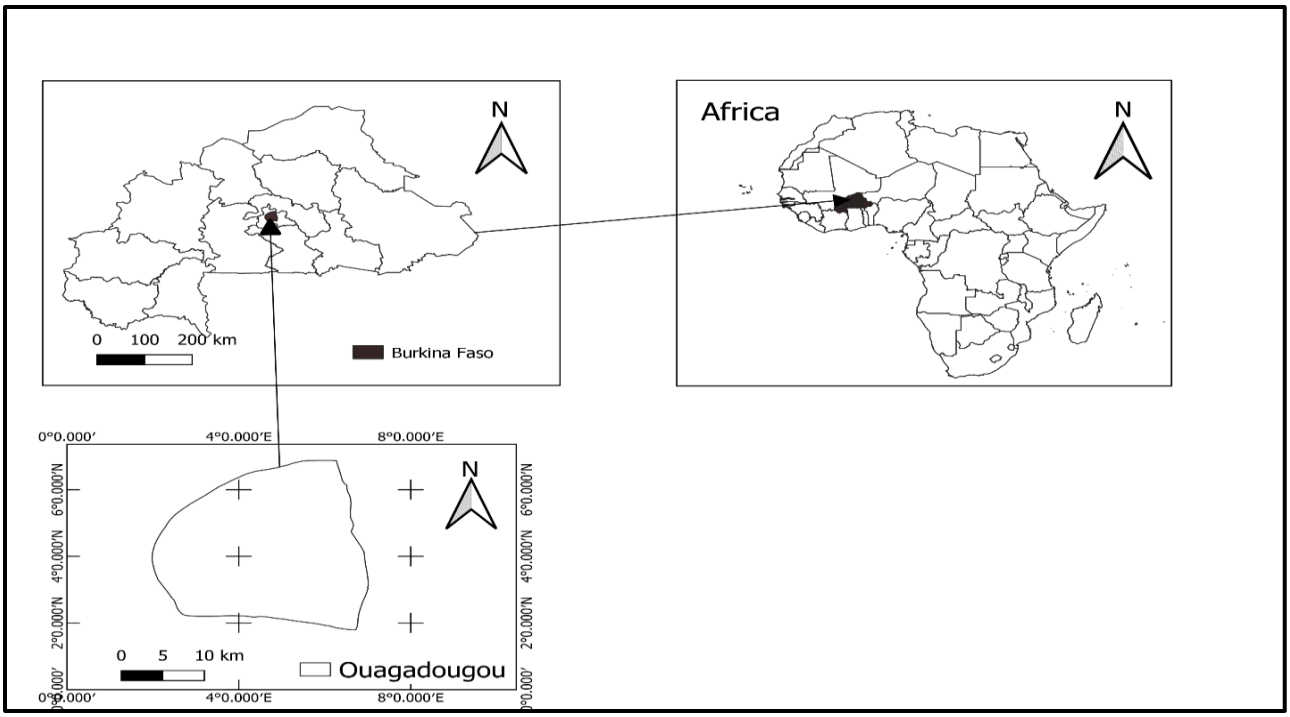

Figure 1: Location of Burkina Faso and Ouagadougou in Africa

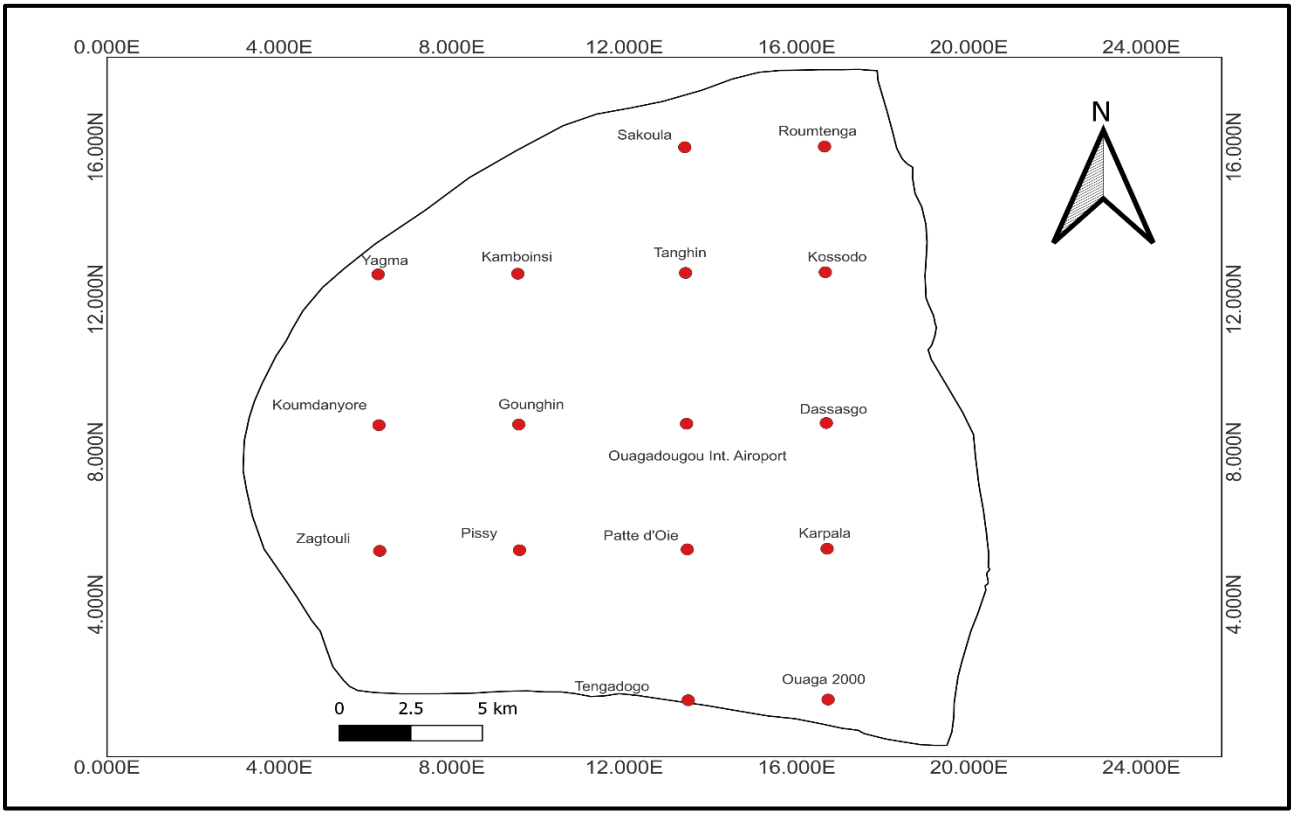

Figure 2: The sixteen (16) locations in Ouagadougou where the satellite data were extracted

2.2 Data Collection

2.2.1 $PM_{2.5}$

$PM_{2.5}$ concentrations are usually monitored using aerosol samplers or $PM_{2.5}$ monitors. These devices operate identically, drawing air inside before calculating the amount of $PM_{2.5}$ present on a filter. The values are expressed as micrograms ($\mu g/m^3$) per cubic meter of air. The $PM_{2.5}$ in this work was collected from the U.S. Embassy air quality station instrument (BAM-1020) in Ouagadougou. The BAM-1020 Instrument measures and reports $PM_{2.5}$ levels with high accuracy on an hourly basis. The Hourly $PM_{2.5}$ data at the station was available for the period January 2022 to December 2022. Therefore, data for this period was obtained.

The BAM-1020 Instrument uses the principle of beta ray attenuation to measure the mass concentration of $PM_{2.5}$ in ambient air. The beta-attenuation approach is one of the most popular real-time techniques for measuring ambient particulate matter since it provides continuous measurement while needing little operator attention (Shukla & Aggarwal, 2022). A $_{14}C$ element (<60 μCi) in the BAM-1020 emits a continuous stream of beta particles (high-energy electrons) directed at the filter tape. The ambient particulate matter collected on the filter tape weakens the beta rays, and the mass loading on the filter tape has an inverse relationship with the signal loss seen by the BAM-1020 scintillation counter before and after collection. The BAM-1020 calculates particle concentrations in ambient air using mass data and flow observations. The BAM-1020 calculates and reports these concentrations as hourly averages in units of ug/m^3 or mg/m^3.

2.2.2 MODIS AOD

AOD refers to the measurement of aerosols dispersed throughout an air column from the Earth's surface to the top of the atmosphere. AOD informs us of the amount of direct sunlight that these aerosol particles keep from reaching the Earth. A value of 0.01 indicates an exceptionally clean environment, whereas a value of 0.4 indicates foggy air conditions. When there is severe pollution, an AOD value greater than 1 indicates a high concentration of aerosols in the atmosphere. AOD levels between 1 and 3 are high in some natural events like wildfires and sandstorms. The greatest AOD that MODIS sensors may report is 4 (Li et al., 2021). During the "harmattan" season, when biomass/urban pollution aerosol combines with coarse mode dust, fine mode aerosols predominate the daily aerosol optical influence (Ogunjobi et al., 2008).

In this work, the AOD was extracted from the MODIS Terra and Aqua MAIAC Daily Level 2 Aerosol Product at spatial resolutions of 1 km × 1 km for Ouagadougou for the period 2000-2022 using Google Earth Engine (GEE) and Python Programming. The MODIS sensors aboard National Aeronautics and Space Administration (NASA) Terra satellite track the AOD at daily frequency globally, and atmospheric retrievals at 36 spectral bands from the visible (VIS) to near-infrared (NIR) ranging from (0.41-15.0 μm) with a spatial resolution ranging between 0.25 km and 1 km by using two distinct algorithms over land and ocean (Levy et al., 2010). To extract the AOD in GEE, the vector shapefile for Ouagadougou (the Region of Interest (RoI)) and the raster images for AOD were imported. The coordinates of the sixteen (16) different locations in Ouagadougou were extracted from Ouagadougou's vector shapefile into a new shapefile in order to have daily AOD at different areas in the city. The sixteen (16) areas in Ouagadougou were chosen based on the geographical distribution of the city. The new shapefile was then imported into GEE and used to download the daily AOD data. The daily AOD values were obtained from different images of the collection at 550 nm (AOD_{550}) which were tagged by the date on which the location was observed by the MODIS satellite.

A function was defined to reduce the pixel values contained within the RoI to a single statistic by averaging them. This function was then mapped over all the images in the collection to derive the average AOD for the region for the selected duration. The resulting time series was then exported as comma separated values (CSV) file to Google Drive. The AOD values are reported on a scale of 0.001. Therefore, the CSV file was then imported into a Google Colab notebook, and a function was defined to divide all the values by 1000 to rescale the values to a scale of 1. Since several observations can be made on the same day by the MODIS satellite, the values were averaged to get a single number for each day.

2.2.3 Ground-based Meteorological Parameters

The Ground-based meteorological Parameters used in this work were obtained from ANAM, Burkina Faso. ANAM is the appropriate national authority for meteorology and climate issues in Burkina Faso. It is also the recognized national service provider for meteorological and climatic data. Daily weather parameters including temperature, relative humidity, Precipitation, wind speed, and wind direction were collected for the period 2000-2022, from the ANAM weather station at Ouagadougou International Airport.

2.2.4 Satellite Weather Parameters

Since the ground-based weather observations from the ANAM station at Ouagadougou International Airport alone might not be a good representation of all the different areas in Ouagadougou, Satellite weather observations including temperature, relative humidity, Precipitation, wind speed, wind direction were complimented. The satellite weather observations were extracted for the 16 areas in the city where the AOD was extracted, which included areas where the ANAM station did not cover and the area where the station covered (Ouagadougou International Airport).

2.2.4.1 Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) Satellite Precipitation

CHIRPS is a quasi-global rainfall dataset spanning more than 30 years. To construct gridded rainfall time series for trend analysis and seasonal drought monitoring, CHIRPS combines in-situ station data with 0.05° resolution satellite imagery (de Sousa et al., 2020). The data set spans from 1981 to the near present. Two main data sets are present. The first is almost universal and spans the globe from 50° N to 50° S. The second includes regions of the Middle East and Africa. It encompasses the region between 20° W and 55° E, and between 40° N and 40° S. Data on a 0.05° grid at monthly, pentad, and daily times steps are included in the worldwide data set.

Furthermore, data at a 0.10° grid and a 6-hour time step are included in the Africa data set. The anticipated correlation between the precipitation for a given pixel and that from the neighboring stations is used in the technique for integrating CHIRPS with station measurements. The CHIRP fields are used to estimate these correlations. It also uses a second correlation value, which is meant to be an estimation of the correlation between the CHIRPS values and the "actual" precipitation at each pixel. This correlation, which is calculated from correlations between CHIRPS pixel values and gridded station observations, is given a value of 0.5. The next step is to calculate bias ratios using the closest five stations. A weighted average is then used to integrate these ratios into a single adjustment factor, with the weights being the squares of the correlation coefficients. To construct adjusted-CHIRP, these correction factors are multiplied by the CHIRP data. The original (unadjusted) CHIRP and the adjusted CHIRP are mixed in the final phase. The ratio of CHIRP and adjusted-CHIRP to be combined is calculated using the square of the correlation between CHIRP and "actual" rainfall as well as the projected correlation of the closest station. This last stage results in the CHIRPS product (Funk et al., 2015).

The pentad (5-day) and monthly time scales are used to merge station data with CHIRP, and the pentads are afterward rescaled so that the total number of pentads in a month equals the monthly values. The monthly fields and pentads are combined to create a daily version. Daily cold cloud duration (CCD) percentages are used by the daily CHIRPS to distinguish between rain and no-rain events, and the related pentad rainfall is subsequently distributed across the daily rain events proportionally to CCD.

In this work, daily precipitation was extracted from CHIRPS in Google Earth Engine using JavaScript code. The Ouagadougou vector shapefile and the shapefile containing the coordinates of the 16 areas in the city were used to extract the data in order to have daily precipitation data for these 16 areas. During the extraction, the CHIRPS 5 km x 5 km spatial resolution was resampled to 1 km x 1 km spatial resolution using the resampling nearest neighbor interpolation method, so it has the same spatial resolution as the MODIS AOD. The data was extracted for the period 2000-2022.

2.2.4.2 MODIS Land Surface Temperature

The generalized split-window approach and the day/night algorithm are used to recover the Land Surface Temperature (LST) and Emissivity daily data at 1 km pixels and 6 km grids, respectively. The split-window approach divides the lower boundary air surface temperature, atmospheric column water vapor, and band 31 and band 32 emissivities into tractable sub-ranges for optimal retrieval. The surface temperature, the surface's emissivity and reflectivity, atmospheric emission, the absorption and scattering of thermal radiation from the surface, as well as the sun's radiation during the day, all affect the thermal infrared signature that satellite sensors detect. LST is retrieved from MODIS thermal channel data for the entirety of the Earth's land surface, encompassing evergreen and deciduous forests and shrubs, crop and grasslands, inland waterbodies, snow and ice, barren lands with exposed soil, sands and rocks, and urban areas (Wan, 2013).

The day/night algorithm uses pairs of day and night MODIS measurements in seven thermal infrared (TIR) bands to extract daytime and nighttime LSTs and surface emissivities. LSTs, quality evaluations, observation times, view angles, and emissivities make up the product. The MOD11 L2 swath product yields the temperature value. Some pixels may contain several observations where the clear-sky requirements are satisfied above 30 degrees latitude. The pixel

value is the average of all qualifying observations when this happens. The MODIS bands 31 and 32 and six observation layers are also included, together with the daytime and nighttime surface temperature bands and associated quality indicator layers. The MODIS LST product based on thermal infrared data will only be accessible in clear sky conditions, it should be noted.

In this work, the temperature was extracted from MODIS Daytime Land Surface Temperature (LST_Night_1km) and the Nighttime Land Surface Temperature (LST_Day_1km) bands in Google Earth Engine for the period 2000-2022. The Ouagadougou vector shapefile and the vector shapefile containing the coordinates of the 16 areas in the city were used to extract the data according to my region of interest. A function was created to convert the temperature from Kelvin (K) to degree Celsius (°C) by converting the image to float and multiplying by a scale of 0.02 and then subtracting 273.15 from each value. To get the average temperature per day, the mean of the LST_Night_1km band and LST_Day_1km band was calculated using the expression;

$$T = \frac{T_d + T_n}{2} \quad (1)$$

Where T is the average temperature per day in °C

T_d is the LST_Day_1km in °C

T_n is the LST_Night_1km in °C

2.2.4.3 ERA5-Land Daily Aggregated - ECMWF Climate Reanalysis

ERA5 is the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis of the entire world's climate. A global full and consistent dataset is produced through reanalysis, which mixes model data with observations from all across the world. ERA5 supersedes ERA-Interim, which it succeeded. Seven (7) ERA5 climate reanalysis parameters are gathered in ERA5 DAILY: 2 m air temperature, 2 m dewpoint temperature, total precipitation, mean sea level pressure, surface pressure, 10m u-component of wind, and 10m v-component of wind. These values are provided for each day. Also, using the hourly 2 m air temperature data, the daily minimum and maximum air temperatures have been determined. Figures for daily sums of precipitation are provided. Daily averages are supplied for all other parameters.

ERA5-Land is a reanalysis dataset that offers an improved resolution compared to ERA5 and a consistent view of the evolution of land characteristics over multiple decades. Replaying the land portion of the ECMWF ERA5 climate reanalysis led to the creation of ERA5-Land. Reanalysis utilizes the rules of physics to merge model data with observations from all across the world into a globally complete and consistent dataset. Reanalysis generates data that covers a number of decades in the past and gives a precise account of the historical climate. All 50 variables that are included in the climate data store are included in this dataset (Muñoz-Sabater et al., 2021). The ERA5-Land Daily Aggregated dataset is at a spatial resolution of 9 km x 9 km. The asset is a daily aggregation of hourly assets from the ECMWF ERA5 Land model that includes both flow bands and non-flow bands. The first hour of the following day's data, which contains the aggregated sum of the previous day, is collected to produce flow bands, while the entire day's hourly data is averaged to create non-flow bands. This approach differs from the daily data generated by Copernicus Climate Data Repository, where flow bands are also averaged, and the flow bands are identified with the "_sum" identifier. Daily ERA5-Land aggregated data from July 1963 to three months from real-time are accessible.

In this work, the relative humidity, wind speed, and wind direction were extracted from the daily ERA5-Land aggregated data in GEE for the period 2000-2022. The 9 km x 9 km spatial resolution of the daily ERA5-Land was resampled to 1km x 1km spatial resolution using the resampling nearest neighbor interpolation method to allow the data to have the same spatial resolution as the resampled CHIRPS precipitation, the MODIS AOD, and the land surface temperature. The Ouagadougou vector shapefile containing the coordinates of the 16 areas in the city was used to extract the data to the study region. Since relative humidity is not a direct product from satellite observations, it was calculated from the 2 m air temperature and 2 m dewpoint temperature (Lawrence, 2005). The 2 m air temperature is the temperature to which the air, at 2 meters above the surface of the Earth, would have to be cooled for saturation to take place. it serves as an indicator of air humidity. The 2 m dew point temperature is determined by interpolating between the Earth's surface and the lowest model level while taking the atmospheric conditions into consideration. The following expression was used to compute the Relative humidity;

$$RH = exp(\frac{17.269 \times T_d}{273.3 + T_d}) - \frac{17.269 \times T}{237.3 + T}) \times 100 \quad (2)$$

where RH is the relative humidity in percentage

T_d is the dewpoint_temperature_2m in °C

T is the temperature_2m in °C

The wind speed and wind direction were computed from the u and v components of wind (10 m u-component of wind and 10 m v-component of wind) (Weber, 1991). The 10 m u-component of wind is the eastward component of the 10 m wind. It is the horizontal speed in meters per second of air traveling in the direction of the east at a height of 10 meters above the Earth's surface. The 10 m v-component of wind is the Northward component of the 10 m wind. It is the air's horizontal speed, measured in meters per second, at a height of ten meters above the Earth's surface as it moves in the direction of the north. The Pythagorean Theorem expression below was used to compute the wind speed

$$WS = \sqrt{(u^2 + v^2)} \quad (3)$$

where WS is the wind speed in m/s

u is the u-component of wind at 10 m

v is the v-component of wind at 10 m

The wind direction was also computed using the following trigonometric expression;

$$WD = mod(180 + (\frac{180}{3.14}) \times atan2(v, u), 360) \quad (4)$$

where WD is the wind direction in degrees

u is the u-component of wind at 10 m

v is the v-component of wind at 10 m

Table 1: Summary of the satellite weather parameters downloaded

Satellite Parameter	Spatial resolution	Temporal resolution	Source
Precipitation	Resampled at 1 km	Daily	CHIRPS
Temperature	1 km	Daily	MODIS
Relative Humidity	Resampled at 1 km	Daily	Era5-Land
Wind Speed	Resampled at 1 km	Daily	Era5-Land

Wind Direction	Resampled at 1 km	Daily	Era5-Land

2.3 Data Processing and Analysis

2.3.1 $PM_{2.5}$

The hourly $PM_{2.5}$ data was used to compute the average $PM_{2.5}$ concentrations according to local time for the period it was collected. This analysis was done to determine which times (hours) of the day were the $PM_{2.5}$ concentrations high and which times were the concentrations low at the air quality station location (U.S. embassy, Ouaga 2000). This helped to draw some conclusions about the possible internal sources of $PM_{2.5}$ at the location and in the city. The 24-hour $PM_{2.5}$ concentrations were averaged to obtain daily $PM_{2.5}$ concentrations which was then used for the models' development.

It should be noted that $PM_{2.5}$ concentrations were analyzed following both the WHO and the US EPA air quality guidelines.

2.3.2 Observed and Satellite Weather Parameters

The relationship between the observed weather parameters (temperature, relative humidity, precipitation, wind speed, wind direction) and the satellite weather parameters at the Ouagadougou International Airport was determined using Pearson correlation. The Pearson correlation coefficient measures the linear correlation between two sets of data (Barnston, 1992). It is the ratio between the covariance of two variables and the product of their standard deviations; as a result, it is effectively a normalized measurement of the covariance, with the result always falling between 1 and 1. Values that are close to +1 or -1 indicate a strong relationship (Schober & Schwarte, 2018).

It is given by the expression;

$$r = \frac{N\varepsilon xy - (N\varepsilon x)(\varepsilon y)}{\sqrt{[N\varepsilon x^2 - (\varepsilon x)^2][N\varepsilon y^2 - (\varepsilon y)^2]}} \quad (5)$$

Where:

r = Pearson correlation coefficient

N = the number of pairs of scores

εxy = the sum of the products of paired scores

εx = the sum of x scores

εy = the sum of y scores

εx^2 = the sum of squared x scores

εy^2 = the sum of squared y scores

The slope of the line denotes whether there is a positive or negative linear relationship between the variables. A Positive correlation exists between the variables if the line slopes upward. This implies that if the value of one variable increases, the value of the other variable will also increase. A negative correlation shows a slope that is downward. This indicates that an increase in one variable causes a decrease in another variable's value.

The satellite weather parameters showed strong correlations with the observed weather parameters but with few biases. Because of these strong correlations, simple linear regression models were developed to correct the satellite data and reduce the bias. The days where they were missing values for satellite parameters were removed along with their corresponding ground observed values from the dataset before the data was used for the model's development. The observed parameters served as the dependent variables and the satellite parameters served as the independent variables. Each observed parameter was then fitted with its corresponding satellite parameter and the linear equation between them was obtained. A simple linear regression model describes the relationship between two variables by fitting a line through them. The simple linear regression model helps to estimate how a change in the dependent variable explains a change in the independent variable (Altman & Krzywinski, 2015). The simple linear regression models were developed following the equation below;

$$Sat_{corri} = \beta_i \times Sat_i + \beta_{0i} \quad (6)$$

Where:

Sat_{corri} is the corrected satellite weather parameter, Sat_i is the satellite weather parameter, and β_i is the regression coefficient, β_{0i} is the intercept.

The models were evaluated based on the R^2 and the RMSE. The R^2 is a statistical metric that depicts how much variation in a dependent variable results from an independent variable. It is

a measure of how well the regression line approximates the actual data, and hence, the goodness of fit of a model (Miles, 2014). The range of R^2's value is between 0 and 1. If the value is 0, it indicates that the independent variable is not able to account for changes in the dependent variable. A value of 1 indicates, however, that the independent variable completely accounts for the variation in the dependent variable. So, if a model's R^2 is 0.50, it means that its inputs can account for around half of the observed variation. The closer the R^2 value to 1, the good the model. It is represented mathematically as follows;

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\varepsilon(y_i - \hat{y}_i)^2}{\varepsilon(y_i - \bar{y})^2} \quad (7)$$

The sum squared regression (SSR) is the sum of the residuals squared and the total sum of squares (SST) is the sum of the data's deviations from the mean.

RMSE is the residuals' standard deviation (prediction errors). The distance between the data points and the regression line is measured by residuals, and the spread of these residuals is measured by RMSE. In other words, it provides information on how tightly the data is clustered around the line of best fit. In simple terms, RMSE is the square root of the mean of the square of all of the errors (Barnston & G., 1992).

$$RMSE = \sum_{i=0}^{n} \frac{(\hat{y}_i - y_i)^2}{n} \quad (8)$$

$\hat{y}_i$ are predicted values

y_i are observed values

N is the number of observations

Given the same climatic zone, these models were then applied to correct all the satellite data for the other fifteen (15) locations in Ouagadougou where ground observation data was not available.

2.3.3 Statistical Regression Analysis

2.3.3.1 Simple Linear Regression

A Simple Linear Regression was developed from the daily (24 h averaged) $PM_{2.5}$ data from the air quality monitoring station at the U.S. embassy (Ouaga 2000) and the satellite AOD observed at the same location. This analysis was done in Python and in Excel using the data analysis tool.

Before the model development, Pearson correlation was done to determine the Pearson correlation coefficient between the $PM_{2.5}$ and the AOD variable and determine their strength and direction. It was noted that there are sample gaps in both the satellite AOD and the ground $PM_{2.5}$ data, so days, where both values were not available, were removed from the dataset along with their corresponding corrected satellite weather data before the model's development and validation. 80 % of the data was used for model development and 20 % was used for model testing. The model was developed based on the $PM_{2.5}$-AOD linear equation proposed by Engel-Cox et al. (2004).

$$PM_{2.5} = \beta_0 + \beta_{AOD} \times AOD \quad (9)$$

where $PM_{2.5}$ is the mass concentration ($\mu g/m^3$), β_0 is the intercept and β_{AOD} is the regression coefficient of the AOD.

The model was evaluated using the R^2 and the RMSE statistical metrics.

2.3.3.2 Multiple Linear Regression

MLR is used for Modeling the linear relationship between the explanatory (independent) variables and response (dependent) variables (Jobson, 1991). It assumes that there is a linear relationship between the dependent variables and the independent variables. A precise estimation of the degree of influence each independent variable will have on the outcome variable may be made using the data on the many variables once each independent factor's ability to predict the dependent variable has been established. The model constructs a linear relationship that best approximates each of the discrete data points as a straight line (Jobson,1991). It is represented as;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{in} x_{in} + \varepsilon \quad (10)$$

Where for i = n observations:

y_i = dependent variable

x_i = explanatory variables

β_0 = y-intercept

β_n = slope coefficient for each explanatory variable

ε = model's error term

To understand the influence of weather parameters on $PM_{2.5}$, the MLR model was developed. The AOD and the corrected weather parameters (temperature, relative humidity, precipitation, wind speed, and wind direction) at Ouaga 2000 were used as input to develop the MLR model. 80 % of the data was used to develop the model and 20 % was used to test the model's performance. Before the model development, each parameter was correlated with the surface $PM_{2.5}$ to determine their relationship. This was done using Pearson's correlation technique

Song et al. (2014) modified the simple linear equation (9) by introducing meteorological parameters to obtain a multivariable equation for estimating $PM_{2.5}$ as follows;

$$PM_{2.5} = (\alpha + \varepsilon_1) + (\beta_1 + \varepsilon_2) \times AOD + (\beta_2 + \varepsilon_3) \times TEMP + (\beta_3 + \varepsilon_4) \times RH + (\beta_4 + \varepsilon_5) \times WS \quad (11)$$

where TEMP is temperature (°C); RH is relative humidity (%); WS is wind velocity (m/s); α and β are fixed coefficients; and ε is a random error. In developing the MLR model in this work, equation (11) was further modified by adding precipitation and wind direction to form equation (12) since research has shown that these variables have an influence on surface $PM_{2.5}$ concentrations.

$$PM_{2.5} = (\alpha + \varepsilon_1) + (\beta_1 + \varepsilon_2) \times AOD + (\beta_2 + \varepsilon_3) \times TEMP + (\beta_3 + \varepsilon_4) \times RH + (\beta_4 + \varepsilon_5) \times WS + (\beta_5 + \varepsilon_6) \times Precip + (\beta_6 + \varepsilon_7) \times WD \quad (12)$$

Where:

TEMP is the temperature (°C); RH is relative humidity (%); WS is wind speed (m/s), Precip is precipitation (mm); WD is wind direction (degrees); α and β are fixed coefficients; and ε is a random error.

The Significance F of the model and the P values of the independent parameters were also determined to test if the model was a good model and the significance of each independent parameter respectively. The model was also evaluated using the R^2 and the RMSE metrics.

2.3.4 Machine Learning Models Development and Validation

From the Pearson correlation coefficients in the statistical regression analysis, it was observed that some of the independent variables are not well linearly correlated with $PM_{2.5}$, which

means that the relationship between some of these variables and $PM_{2.5}$ is not directly linear. This might have affected the optimal performance of the developed MLR model. Multiple linear regression does not give its optimal performance if the independent variables are not strongly correlated with the dependent variables (Uyanık & Güler, 2013).

Hence, three supervised non-linear models (decision tree, random forest, and XGBoost) were developed at the same location (Ouaga 2000) where the simple linear regression and multiple linear regression models were developed. The decision tree model and the random forest model were developed using the DecisionTreeRegressor class and the RandomForestRegressor class respectively from the scikit-learn machine learning library in Python. The XGBoost model was developed using the XGBRegressor class from the XGBoost library in Python. To determine how only AOD can be a predictor of $PM_{2.5}$ in the non-linear models, only AOD was first used as an input to estimate $PM_{2.5}$. These models assume that the relationship between the AOD, the weather parameters, and the $PM_{2.5}$ is non-linear. In all the models, randomly selected 80 % of the data was in the models' development and 20 % for the models' testing.

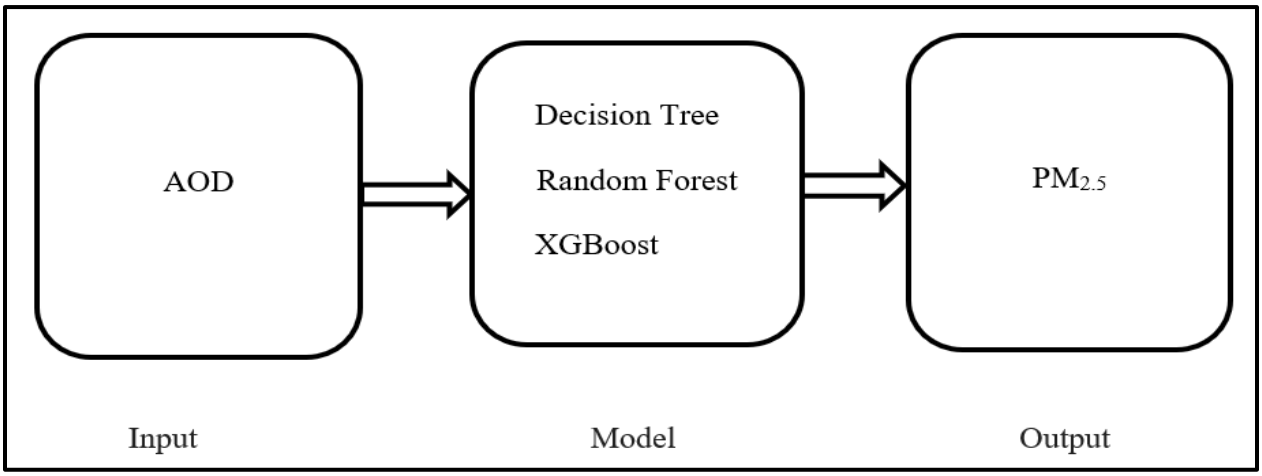

Figure 3*:* AOD Parameter alone as model input

To understand the influence of weather parameters on surface $PM_{2.5}$ and the $PM_{2.5}$-AOD relationship, the corrected weather parameters were then introduced into the models. In all the models, randomly selected 80 % of the data was in the models' development and 20 % for the models' testing.

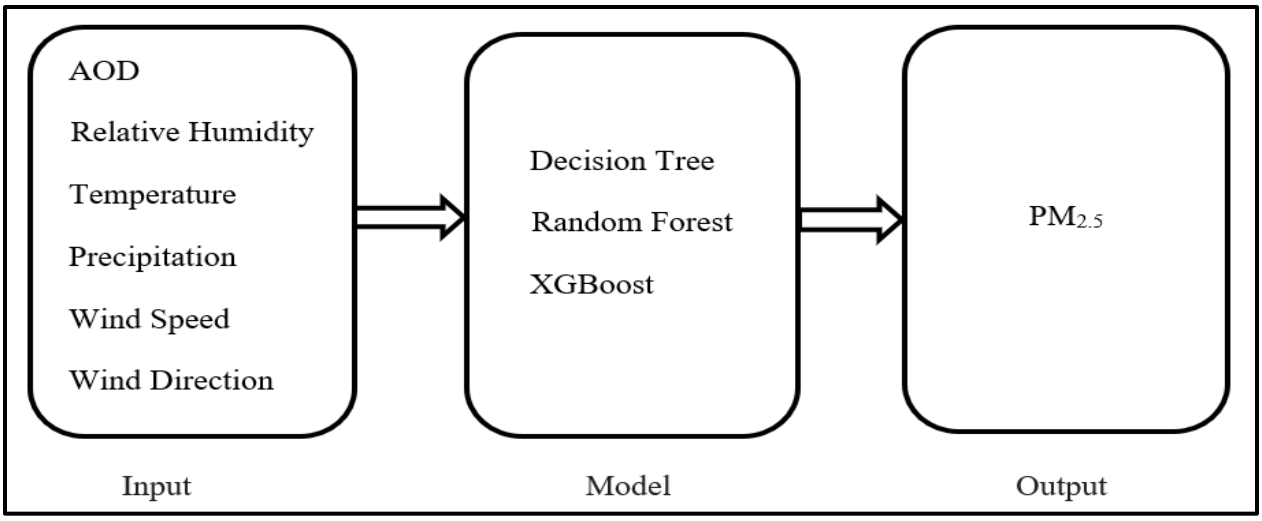

Figure 4*:* AOD and weather Parameters as model input

2.3.4.1 Decision Tree

DT is a non-parametric supervised learning method that can be applied to classification and regression issues (Song & Lu, 2015). It is a tree-structured classifier, where internal nodes stand in for the dataset's features, branches for the decision rules, and each leaf node for the outcome. Two nodes—the Decision Node and the Leaf Node—make up a decision tree. In contrast to Leaf nodes, which represent the decisions' output and have no more branches, Decision nodes are used to make any decision and have several branches. A decision tree only poses a question and divides the tree into subtrees according to the response (Yes/No). it has both classification and Regression Tree algorithms.

In this work, the regression tree algorithm was used. A regression tree is essentially a decision tree used for regression, which predicts continuous-valued outputs rather than discrete outputs**.** It can be useful where the relationship between the variables is found to be non-linear (Song & Lu, 2015). The model was first developed using only AOD as input. This was done to determine how AOD alone would explain the variations of $PM_{2.5}$ in the DT model. After that, all the weather parameters were then introduced into the model to determine how all these parameters combined would explain the variations of $PM_{2.5}$.

The model was developed from the following hyperparameters; ccp_alpha = 0.01, max_depth = 5, max_features = 'sqrt', min_samples_split = 2, random_state = 42, splitter = 'best'.

These hyperparameters were set after using the grid search technique which indicated that these were the optimal hyperparameters for the model. Before a machine-learning algorithm is applied to a dataset, hyperparameters are the parameters that are specifically established to regulate the learning process. The grid search is the simplest approach for hyperparameter tuning. A discrete grid was created within the hyperparameter domain, then all possible combinations of these grid's values experimented.

The model was evaluated using the R^2 and the RMSE from a 5-fold cross-validation technique. The dataset was split into 5 folds and the training and the testing were done on each one. One fold was taken into consideration for testing during each run, with the remaining folds being used for training as iterations continued. This was done to evaluate the model's ability when given new data. The feature importance of the model was computed to determine which feature is important in the model in explaining the variations of $PM_{2.5}$

2.3.4.2 Random Forest

RF is a supervised ensemble learning technique for classification, regression, and other problems that work by building a large number of decision trees during the training phase (Breiman, 2001). For classification tasks, the outcome of the random forest is the class selected by most trees. For regression tasks, the mean prediction of the individual trees is returned. Random Forest's popularity has grown due to how simple and adaptable it is and how well it can handle classification and regression issues. It typically performs very well for problems involving non-linear relationships**.** A random forest algorithm is made up of many decision trees. The random forest algorithm trains its "forest" through bagging or bootstrap aggregating. An ensemble meta-algorithm called bagging increases the precision of machine learning algorithms (Breiman, 2001).

Based on the predictions of the decision trees, the random forest algorithm determines the result. It makes predictions by averaging or averaging out the results from different trees. The accuracy of the result grows as the number of trees increases. The steps used by the RF algorithm are as follows;

A). Pick p data points at random from the training set

B). Create a decision tree based on these p data points

C). Build number N of trees and repeat the A and B steps

D). Predict the value of y for a new data point using each of the N tree trees, then give the new data point the average of all the anticipated y values.

The model was developed using the Grid Search method to get the optimal hyperparameters. The following hyperparameters gave the model its best performance. max_depth =7, n_estimators = 50, max_features = 'sqrt', ccp_alpha=0.01, min_samples_split = 4, min_samples_leaf=2.

Five-fold cross-validation was performed to evaluate the model, its average was used to obtain the R^2 and RMSE of the model. The feature importance of the model was calculated and ranked in descending order.

2.3.4.3 XGBoost

XGBoost is an ensemble learning method that uses more precise approximations to identify the optimum tree model. It is an enhanced distributed gradient boosting library created for effective and scalable machine learning model training (Chen & Guestrin, 2016). XGBoost is one of the effective algorithms in gradient descent that has a linear model algorithm and a tree learning algorithm. Hence, it can handle non-linear relationship problems (Aditya Sai Srinivas et al., 2019). Supervised machine learning, which uses data from various aspects of xi to predict a target variable yi, can be implemented using XGBoost. Because of its speed and precision in predicting outcomes, XGBoost is frequently used by authors to solve various regression and classification issues.

One of XGBoost's distinguishing features is its effective handling of missing values, which enables it to handle real-world data with missing values without necessitating a lot of pre-processing. XGBoost is quite adaptable and enables fine-tuning of many model parameters to enhance performance (Chen & Guestrin, 2016). The model was developed from the following set of hyperparameters using grid search.

base_score = 0.5, booster = gbtree, callbacks = None, colsample_bylevel = None, colsample_byn ode = None, colsample_bytree = 0.5, early_stopping_rounds = None, enable_categorical = False, eval_metric = None, feature_types = None, gamma = 0.4, gpu_id = None, grow_policy = None, i mportance_type = None, interaction_constraints = None, learning_rate =0.1, max_bin = None, m ax_cat_threshold = None, max_cat_to_onehot = None, max_delta_step = None, max_depth = 7,

max_leaves = None, min_child_weight = 7, missing = nan, monotone_constraints = None, n_estimators = 100, n_jobs = None, num_parallel_tree = None, predictor = None, random_state = 42. The model was evaluated using a five-fold cross-validation, the R^2 and the RMSE were then computed. The plot_importance function from the XGBoost library was then used to plot the feature importance of the model.

2.3.4.4 Semi-supervised XGBoost Model

Semi-supervised learning is a machine learning technique that uses a small portion of labeled data and lots of unlabeled data to train a predictive model. Semi-supervised learning uses a combination of supervised and unsupervised learning techniques to train a model (Zhu, 2005). It can be used for both classification and regression problems. It uses clustering techniques on the unlabeled data to better depict the underlying data distribution and more accurately generalize to new unseen samples.

Based on the training principle, semi-supervised learning can be divided into Inductive and Transductive. In the inductive training principle, the semi-supervised model is trained to keep the rules observed during the training process so it can generalize well to new unseen data whereas in the transductive training principle, the semi-supervised model is trained to solve the problem at hand and forget the rules observed during its training, hence does not generalize well to new unseen data and always require re-running of the algorithm. Semi-supervised learning often outperforms the traditional unsupervised learning since it takes advantage of the labeled data to facilitate its learning process (Zhang et al., 2021; Chen et al., 2022). Semi-supervised learning has proven effective in areas with fewer labeled data or in areas where labeled data is expensive (Zhu, 2005).

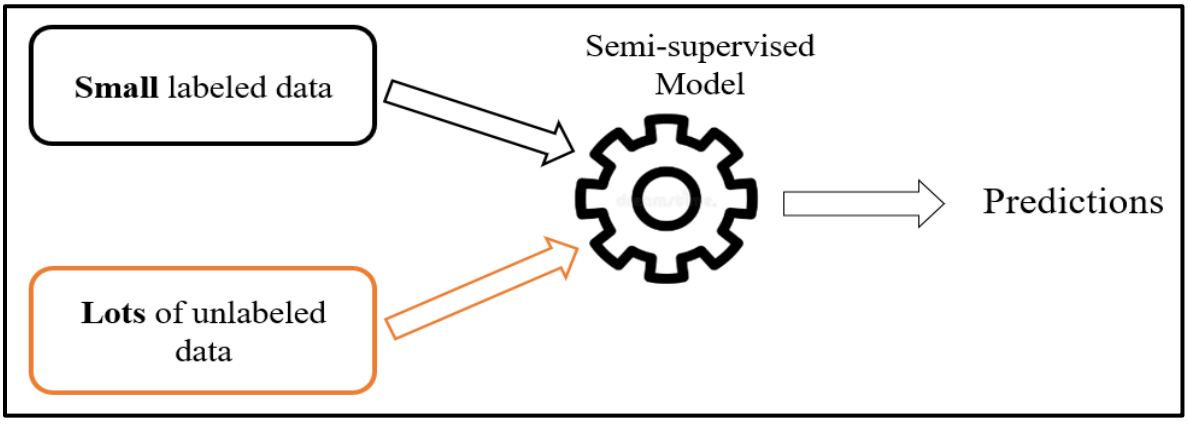

Figure 5: Semi-supervised learning in a Nutshell

In this work, Since after cross-validation XGBoost outperforms DT, RF and the other statistical models in estimating $PM_{2.5}$, the model is upgraded by incorporating a semi-supervised algorithm into it to develop a semi-supervised XGBoost model which allows the model to learn from both the small amount of labeled data available and the lots of unlabeled data in previous years and make predictions. When $PM_{2.5}$, AOD, and meteorological parameters are all available, they are considered labeled data and when AOD and meteorological parameters are available without $PM_{2.5}$, they are considered unlabeled data. Ouagadougou has a small amount of labeled data and lots of unlabeled data, hence the need for the semi-supervised XGBoost model. The model uses techniques of supervised learning on the labeled data and clustering techniques of unsupervised learning on the unlabeled data to learn the structures and patterns of the data, and make accurate predictions. The algorithm divides the unlabeled data into clusters and applies the clustering assumption; points in the same cluster are likely to have the same output. Guided by the labeled data it makes predictions. The model was trained using the inductive training principle so it can generalize well to new unseen data.

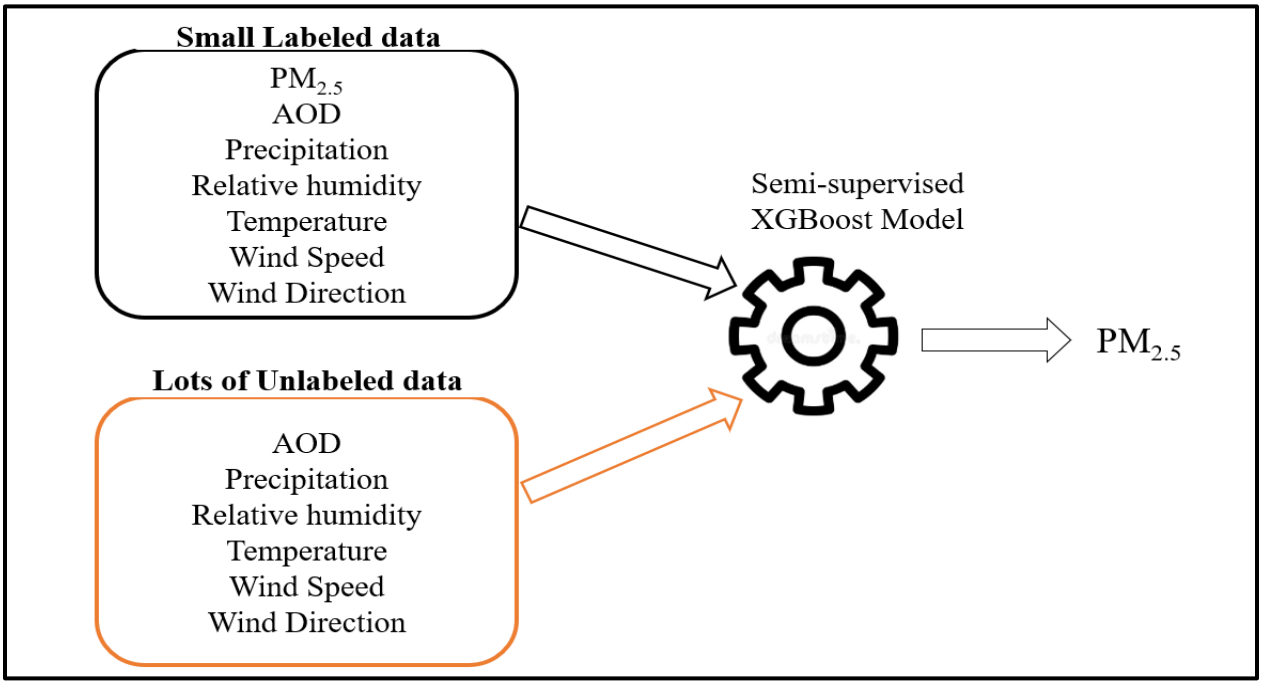

Figure 6: Semi-supervised XGBoost

80 % of the labeled data was combined with the unlabeled data and 20 % of the labeled data was on hold for model testing. The model was evaluated using the R^2 and RMSE metrics on the 20 %

labeled data. A five fold cross-validation technique was used to assess the model's generalization performance.

2.3.5 Estimation of $PM_{2.5}$ in Areas without $PM_{2.5}$ Data in Ouagadougou

The semi-supervised XGBoost model was applied to estimate $PM_{2.5}$ in the remaining fifteen (15) areas of the city where $PM_{2.5}$ data was not available since these areas are in the same climatic zone as where the model was trained. The estimation was done for the period in which the model was trained (2000-2022) to study the growth and distribution of $PM_{2.5}$ for this period.

The estimated daily $PM_{2.5}$ concentrations were analyzed to determine which months have days with high $PM_{2.5}$ concentrations and which months have days with low concentrations at the location. The averaged monthly variations of the estimated $PM_{2.5}$ were also analyzed to determine which months have the highest $PM_{2.5}$ concentrations and which months have the lowest concentrations. The yearly averaged variations of the estimated $PM_{2.5}$ were studied to determine which years had the highest $PM_{2.5}$ concentrations and which years had the lowest.

2.3.6 Spatial Distribution of $PM_{2.5}$

The estimated $PM_{2.5}$ was then plotted spatially using the quantum geographic information system (QGIS) software and the Inverse Distance Weighted (IDW) interpolation. This was done to study the distribution of $PM_{2.5}$ in the city. IDW is a specific kind of deterministic approach for multivariate interpolation using a known scattered set of points (Liu et al., 2021). A weighted average of the values available at the known points is used to determine the values allocated to the unknown points. The representative traditional interpolation techniques for PM estimation are IDW and kriging (Li & Heap, 2011).

IDW is quick and simple to use, and it has been used in many research to evaluate the trend and distribution of PM (Liu et al., 2009). According to some studies, IDW is more appropriate for PM estimation than the Kriging-based approach (Li et al., 2016). The seasonal (dry season and rainy season) distribution of estimated $PM_{2.5}$ was studied. The min, the mean, and the max of $PM_{2.5}$ distributions for the dry and rainy seasons of 2000-2005, 2006-2011, 2012-2017, and 2018-2022 were studied to determine the most polluted areas in Ouagadougou over these intervals (long-term). The reason for analyzing data over every 6 years is to give more information on the variability of $PM_{2.5}$ at the different areas. Also, the mean distribution of the estimated $PM_{2.5}$ for the

dry season and rainy season of each year was studied to determine polluted areas in the city over short-term.

In summary, this methodology section outlined the approach used to estimate $PM_{2.5}$ concentrations in Ouagadougou, leveraging satellite AOD data and meteorological parameters. By incorporating temperature, wind speed, wind direction, relative humidity, precipitation, and AOD as predictors, we developed a comprehensive framework to enhance the accuracy of $PM_{2.5}$ estimation. Also, the spatial distribution $PM_{2.5}$ in the city is studied using the IDW technique in QGIS. Moving forward, the subsequent sections of this work will present the results and analysis derived from our methodology, providing a deeper understanding of $PM_{2.5}$ levels in Ouagadougou and their association with the identified predictors.

CHAPTER 3: RESULTS AND DISCUSSION

The results and discussion section of this study presents the findings from methodology. This section aims to analyze and interpret the obtained results, providing insights into the relationship between the selected predictors and $PM_{2.5}$ levels in the region. the results delve into the outcomes of the statistical models, the machine learning models, and the spatial distribution analyses conducted in the study. Moreover, the discussion provides a deeper analysis and interpretation of the results and draws meaningful conclusions from the results, synthesizing the key findings and their significance.

3.1 Hourly profile of $PM_{2.5}$ at Ouaga 2000

Figure 7 shows the hourly profile of $PM_{2.5}$ at the Ouaga 2000. There are two peaks observe, one in the morning and one in the evening but with different magnitudes. The evening peak is higher with $PM_{2.5}$ concentrations at about 105 $\mu g/m^3$ whilst the morning peak is lower with $PM_{2.5}$ concentrations at about 70 $\mu g/m^3$. The morning peak is observed between 5:00 am and 9:00 am and the evening peak is observed between 4:00 pm and 11:00 pm. The peak in the morning is due to morning vehicle traffic when people are going to work and the evening peak is due to evening vehicle traffic when people are returning home after work. This means that vehicle emissions have a great impact on the concentrations of $PM_{2.5}$ in the area.

Additionally, the high concentrations of $PM_{2.5}$ in the evening might be due to the dynamics of the boundary layer (BL), which produce high dilution rates during the day and low dilution rates after sunset (Lee et al., 2019). However, there is no visible peak between 12: 00 noon and 3:00 pm, as it can be expected since people go out for lunch during this time. Nana et al. (2012); Ouarma et al. (2020) observed similar peaks at administrative sites in Ouagadougou. These peaks are also similar to peaks observed by McFarlane et al. (2021) in Kinshasa-Brazzaville. They observed peaks in the morning at 8: 00 am and in the evening at 8:00 pm.

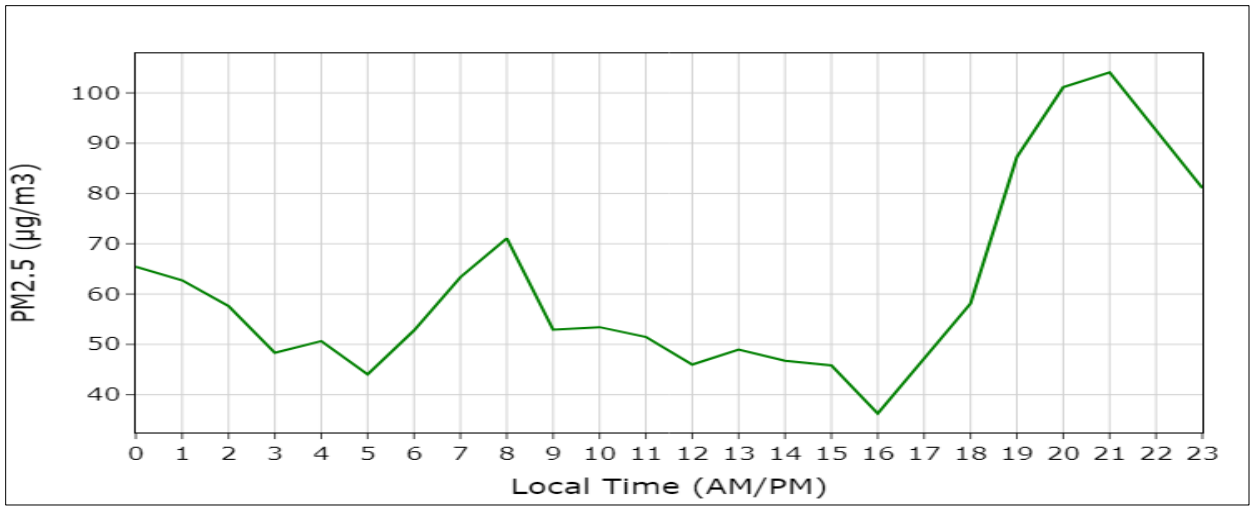

Figure 7: Hourly profile of $PM_{2.5}$ at the Ouaga 2000

3.2 Observed $PM_{2.5}$ and MODIS AOD at Ouaga 2000

Figure 8 shows the relationship the 24-hour average (daily) $PM_{2.5}$ and MODIS AOD at Ouaga 2000. The daily surface $PM_{2.5}$ shows a strong relationship with the daily MODIS AOD with a Pearson correlation coefficient of 0.72. This strong correlation is due to the fact that the city is mostly cloudless and most of the $PM_{2.5}$ in the city comes from the Sahara desert. This correlation is similar to the correlation observed by Léon et al. (2021b). They observed of 0.75 between mean weekly AOD and surface $PM_{2.5}$ in Cotonou, Benin and Abidjan, Côte d'Ivoire. This similarity in correlations is because the surface $PM_{2.5}$ in the region is mostly from the sahara desert and also given the similarities in climatic conditions.

Malings, Westervelt et al. (2020) found similar correlations in Rwanda but however observed different correlations in Pittsburgh. Koelemeijer et al. (2006) found a correlation of 0.6 between MODIS AOD and $PM_{2.5}$ over Europe. Similarly, van Donkelaar et al. (2010b) found a correlation of 0.77 between surface $PM_{2.5}$ and MODIS AOD over Eastern China. These correlations show the usefulness of the MODIS AOD in estimating surface $PM_{2.5}$ concentrations.

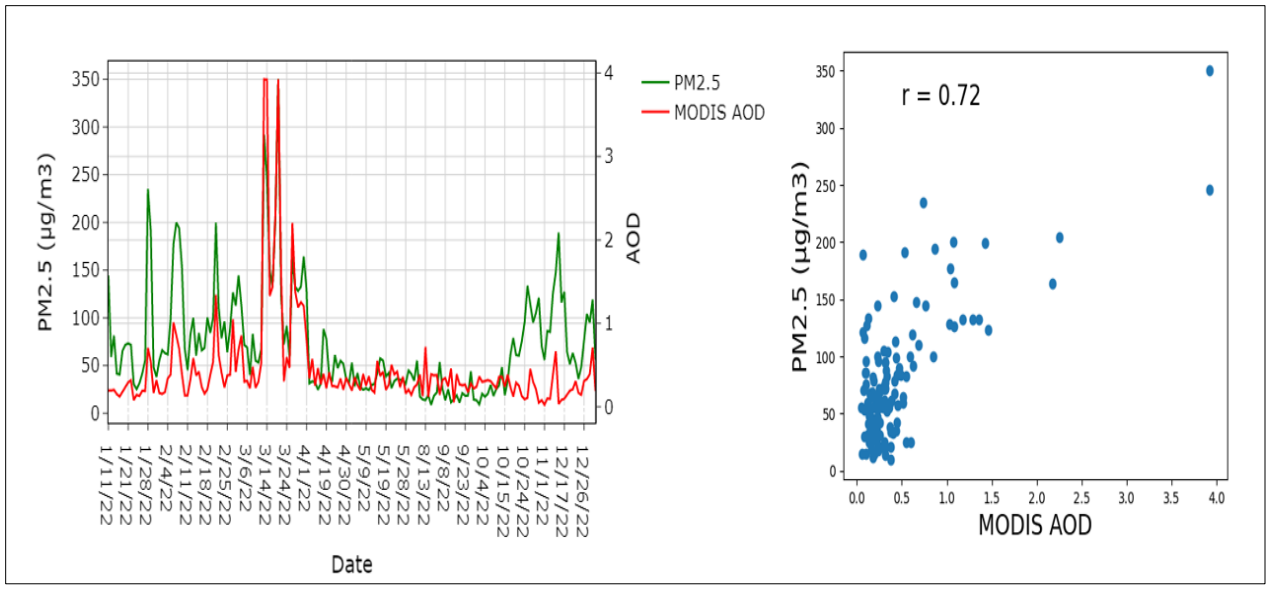

Figure 8: $PM_{2.5}$ and AOD

3.3 Observed and Satellite weather parameters at Ouagadougou International Airport

Figure 9 shows the trend of the relationship between observed precipitation and CHIRPS satellite precipitation at Ouagadougou International Airport. **Table1** shows the Pearson correlation coefficients (r) between observed and satellite weather parameters at the same location. The observed precipitation and the CHIRPS satellite precipitation are following the same trends and are strongly correlated with a Pearson correlation coefficient of 0.87. These findings are similar to the findings by Plessis & Kibii (2021), they had Pearson correlation coefficients of 0.77 between observed precipitation and CHIRPS precipitation over South Africa.

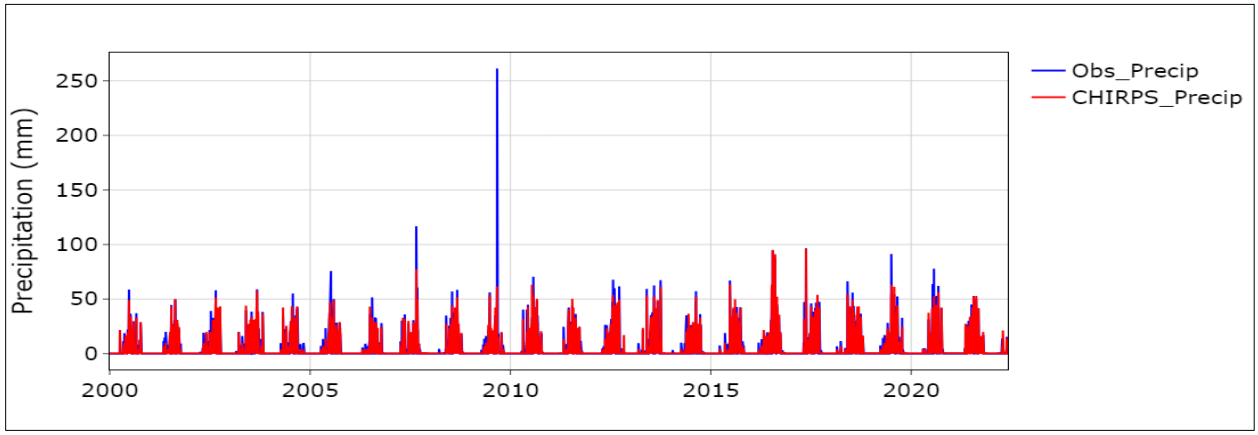

Figure 9*:* Observed precipitation and CHIRPS precipitation

Figure 10 shows the trend of the relationship between observed temperature and MODIS satellite temperature at Ouagadougou International Airport. The observed temperature and the MODIS satellite temperature have similar trends and are strongly correlated with a Pearson correlation coefficient of 0.92. Shen & Leptoukh (2011) found similar correlations (r=0.93) between MODIS land surface temperature and observed temperature over central and eastern Eurasia.

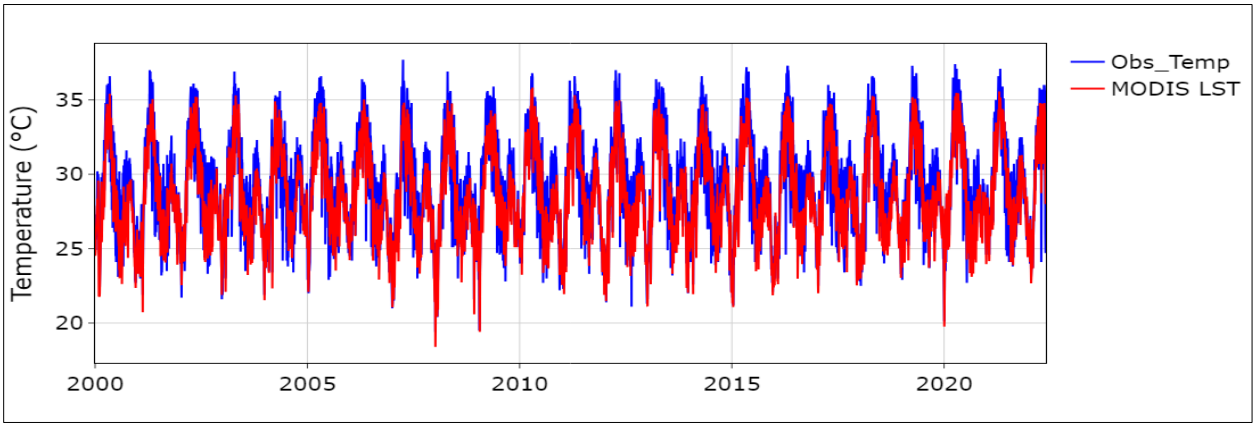

Figure 10: Observed Temperature and MODIS temperature

Figure 11 shows the trend of the relationship between observed relative humidity, wind speed, wind direction and Era5-Land reanalysis relative humidity, wind speed, wind direction at Ouagadougou International Airport. The observed weather parameters and Era5-Land reanalysis parameters show similar trends and are strongly correlated with their Pearson correlation coefficients ranging from 0.89 to 0.96. These Pearson correlation coefficients are similar to what was found by Assamnew & Mengistu Tsidu (2023); Gleixner et al. (2020), they had Pearson correlation coefficients ranging from 0.90 to 0.96 between Era5 weather parameters and observed weather parameters over East Africa.

The strong correlation between observed weather parameters and satellite weather parameters is due to the fact that Ouagadougou is mostly cloud-free throughout the year except in the rainy season.

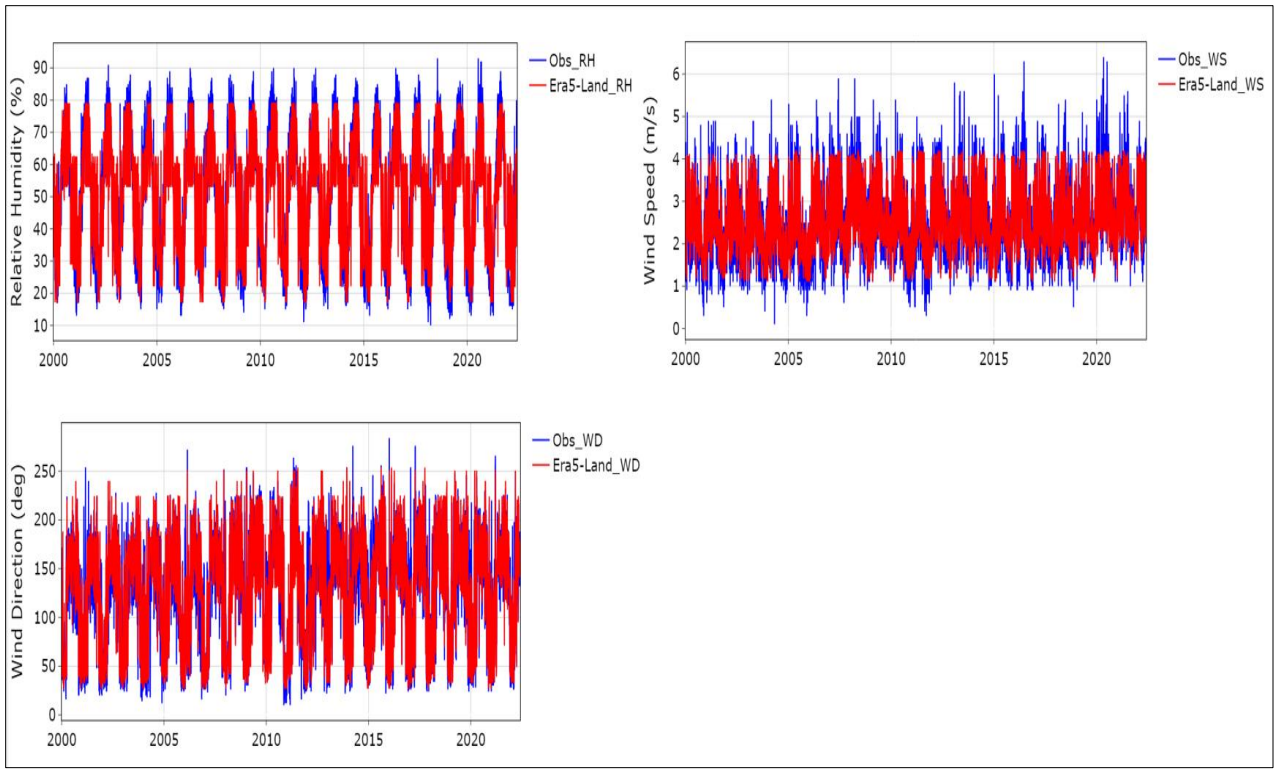

Figure 11: Observed parameters and Era5-Land Parameters

Table 2*:* Pearson correlation coefficients between observed and satellite weather parameters at Ouagadougou International Airport.

Parameters	**r**
Observed precipitation and CHIRPS precipitation (resampled 1km resolution)	0.87
Observed relative humidity and Era5-Land relative humidity (resampled 1km resolution)	0.96
Observed temperature and MODIS Land surface temperature	0.92
Observed wind speed and Era5-Land wind speed (resampled 1km resolution)	0.93
Observed wind direction and Era5-Land wind direction (resampled 1km resolution)	0.89

Based on the strong Pearson correlation coefficients, simple linear regression models are developed as shown in **Figure 12** for correcting satellite data. The R^2 of the models range from 0.75 to 0.92 with RMSE also ranging from 0.3 to 22.2, indicating that the models are able to explains adequately the variations between observed weather parameters and satellite weather parameters. These models are applied to correct the satellite weather data of the other fifteen (15) areas in the city without ground weather observations. The corrected satellite weather parameters are then used along MODIS AOD to develop the models for estimating $PM_{2.5}$ in Ouagadougou.

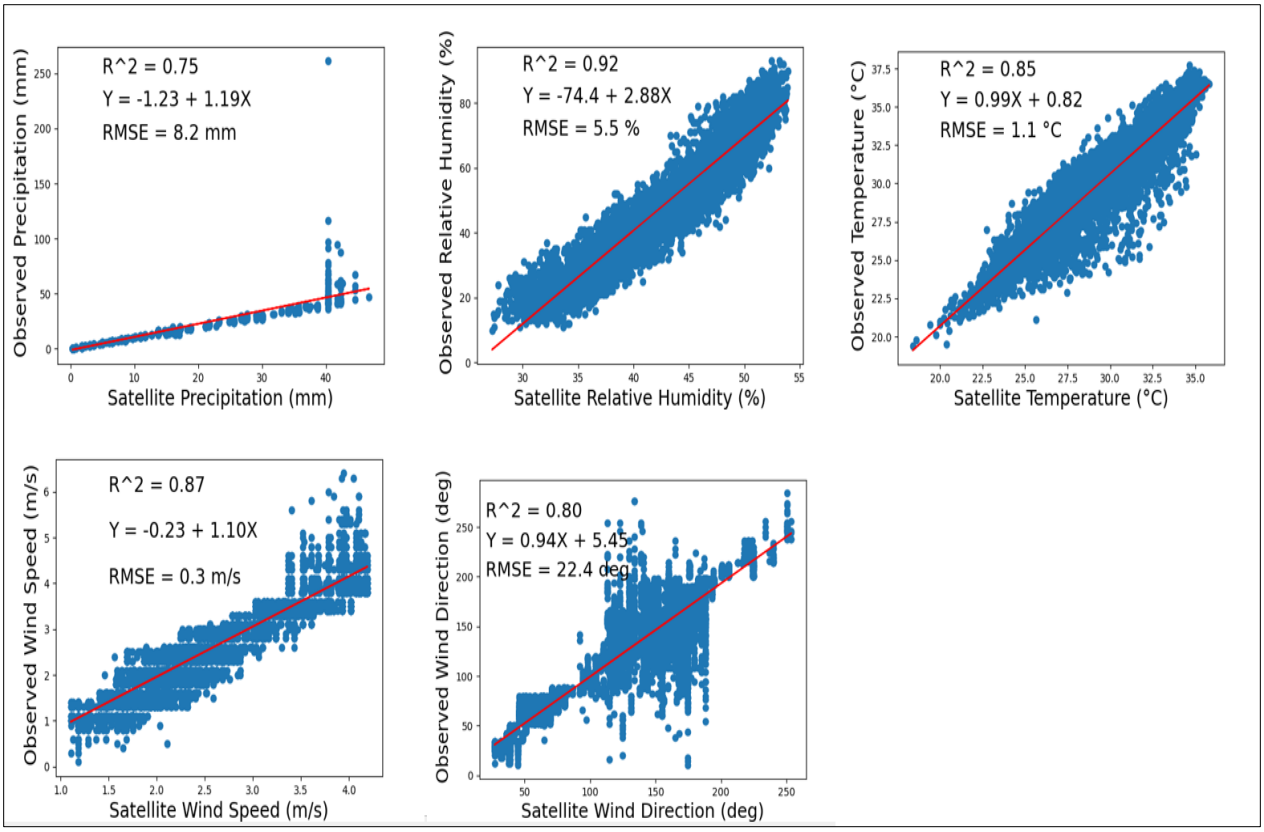

Figure 12: Simple linear models for correcting satellite data

3.4 Observed $PM_{2.5}$ and corrected satellite weather parameters at Ouaga 2000

Figure 13 shows the relationship between surface $PM_{2.5}$ and the corrected CHIRPS precipitation. $PM_{2.5}$ shows a weak negative correlation (r=-0.26) with CHIRPS precipitation. The negative correlation means that as precipitation increases, $PM_{2.5}$ decreases and vice versa. This is consistent to what was explained by Tai et al. (2012). They found that wet deposition serves as the primary sink for atmospheric particulate matter, hence increases in precipitation will result in declines in particle concentrations. The weak correlation means that most of the variations of $PM_{2.5}$

are not directly explained by precipitation. These findings are also consistent with the findings by Westervelt et al. (2016), they observed a correlation of -0.41 between precipitation and $PM_{2.5}$ concentrations in the upper midwest and parts of eastern United States.

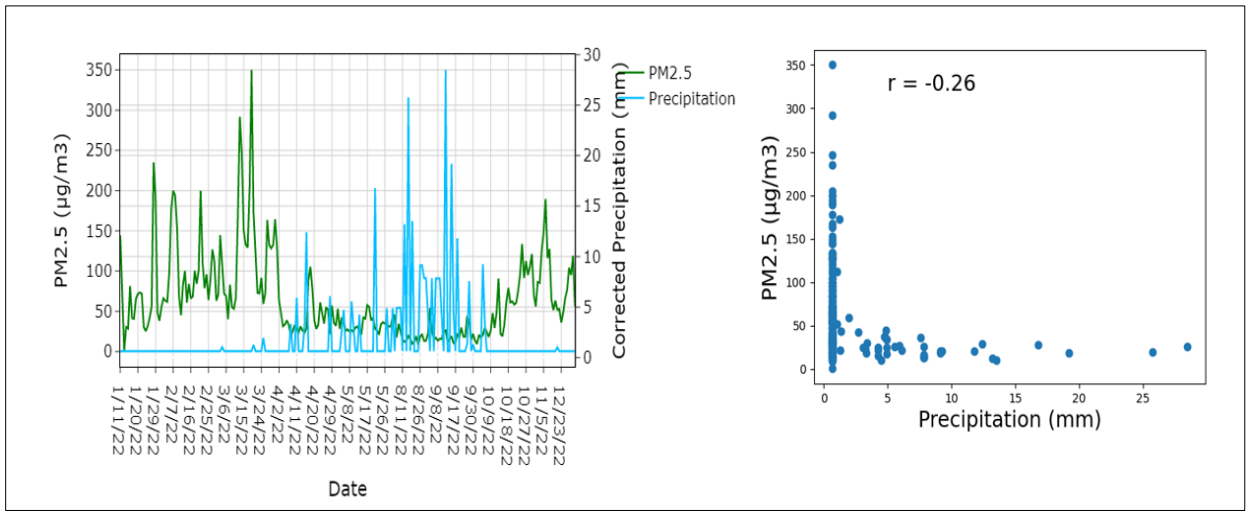

Figure 13: $PM_{2.5}$ and Corrected CHIRPS Precipitation

Figure 14 shows the relationship between surface $PM_{2.5}$ and corrected MODIS land surface temperature. The relationship between these two variables is weak (r=0.11). Temperature is positively correlated with surface $PM_{2.5}$, meaning an increases in temperature will lead to some increase in $PM_{2.5}$ concentrations. This is consistent with the findings observed by Westervelt et al. (2016), they observed a correlation of 0.07 between temperature and $PM_{2.5}$ concentrations over eastern and Midwest United States. Also, Tai et al. (2012) observed a correlation of 0.1 to 0.4 between temperature and gaseous pollutants in some parts of United States. They further explained that heat worsens air pollution by causing reactions between atmospheric particles like nitrogen oxide and oxygen, which create ozone and break down primary particles into even smaller, more dangerous particles. However, the weak correlation observed in our findings indicate that most of the variations in the concentrations of $PM_{2.5}$ are not directly explained by temperature.

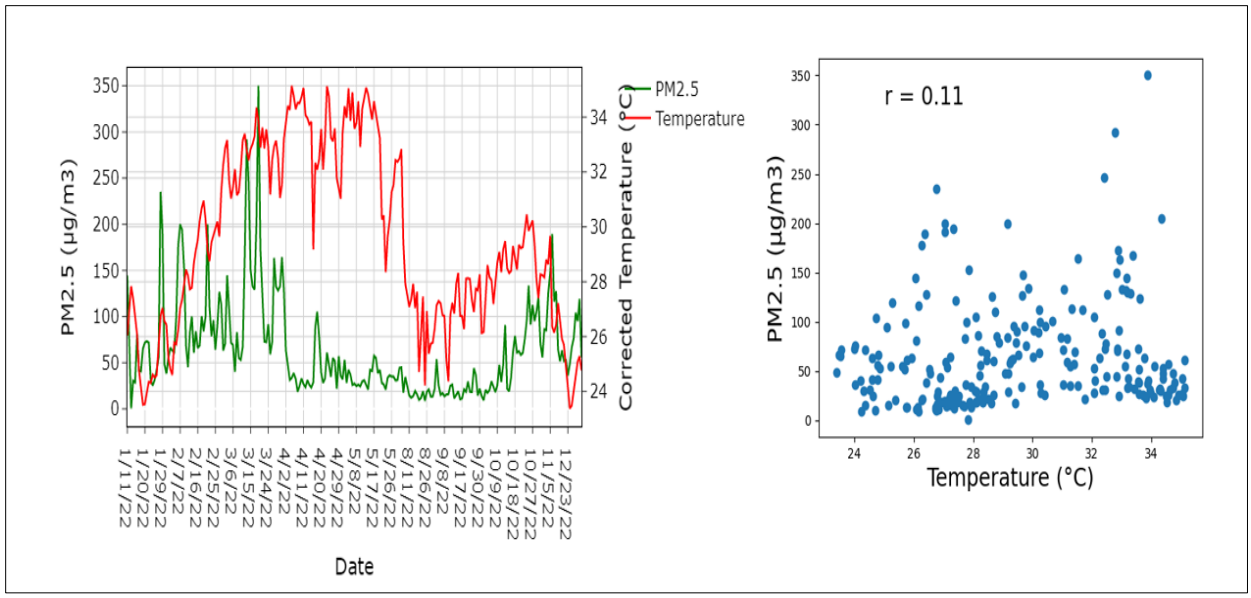

Figure 14: $PM_{2.5}$ and corrected MODIS LST

In **Figure 15**, the relationship between surface $PM_{2.5}$ and corrected Era5-Land relative humidity, wind speed, and wind direction is shown. All these variables are negative correlated with surface $PM_{2.5}$. Relative humidity has the strongest negative correlation (r=-0.55) followed by wind direction(r=-0.33) and wind speed(r=-0.04). The negative correlations implies that an increase in any of these variables will result in a decrease of $PM_{2.5}$. These findings are similar to the find observed by Islam et al. (2023) in Bangladesh. They found a correlation of -0.16 between $PM_{2.5}$ and relative humidity, -0.28 between $PM_{2.5}$ and wind speed, and -0.18 between $PM_{2.5}$ and wind direction. Sirithian & Thanatrakolsri (2022) also found a correlation of -0.72 between $PM_{2.5}$ and relative humidity, and -0.03 between $PM_{2.5}$ and wind speed in northern Thailand. High Relative humidity leads to $PM_{2.5}$ flocculation, followed by gravity settling similar to wet deposition by precipitation hence decreasing the concentrations of $PM_{2.5}$. Ouarma et al. (2020) found that during their PM measurement periods in Ouagadougou, when relative humidity was high (70 % in August and 67 % in September), PM concentrations were low. Also, Lou et al. (2017) found that high humidity (70-90 %) had significant influence on reducing $PM_{2.5}$ concentrations in Yangtze River Delta, China. Low wind speed leads to a stagnant atmosphere favoring $PM_{2.5}$ accumulation. On the other hand, high wind speeds promote $PM_{2.5}$ dissipation. It is important to note that high wind speed can also generate and transport dust.

Similarly, the direction of the wind can determine the concentration of $PM_{2.5}$. If the wind direction is away from an area with $PM_{2.5}$ sources, the $PM_{2.5}$ particles are transported away from that area resulting in lower $PM_{2.5}$ concentrations in that area and hence negative correlation. The weak correlations of the parameters with $PM_{2.5}$ imply that most of the changes in concentrations of $PM_{2.5}$ are not directly explained by the parameters.

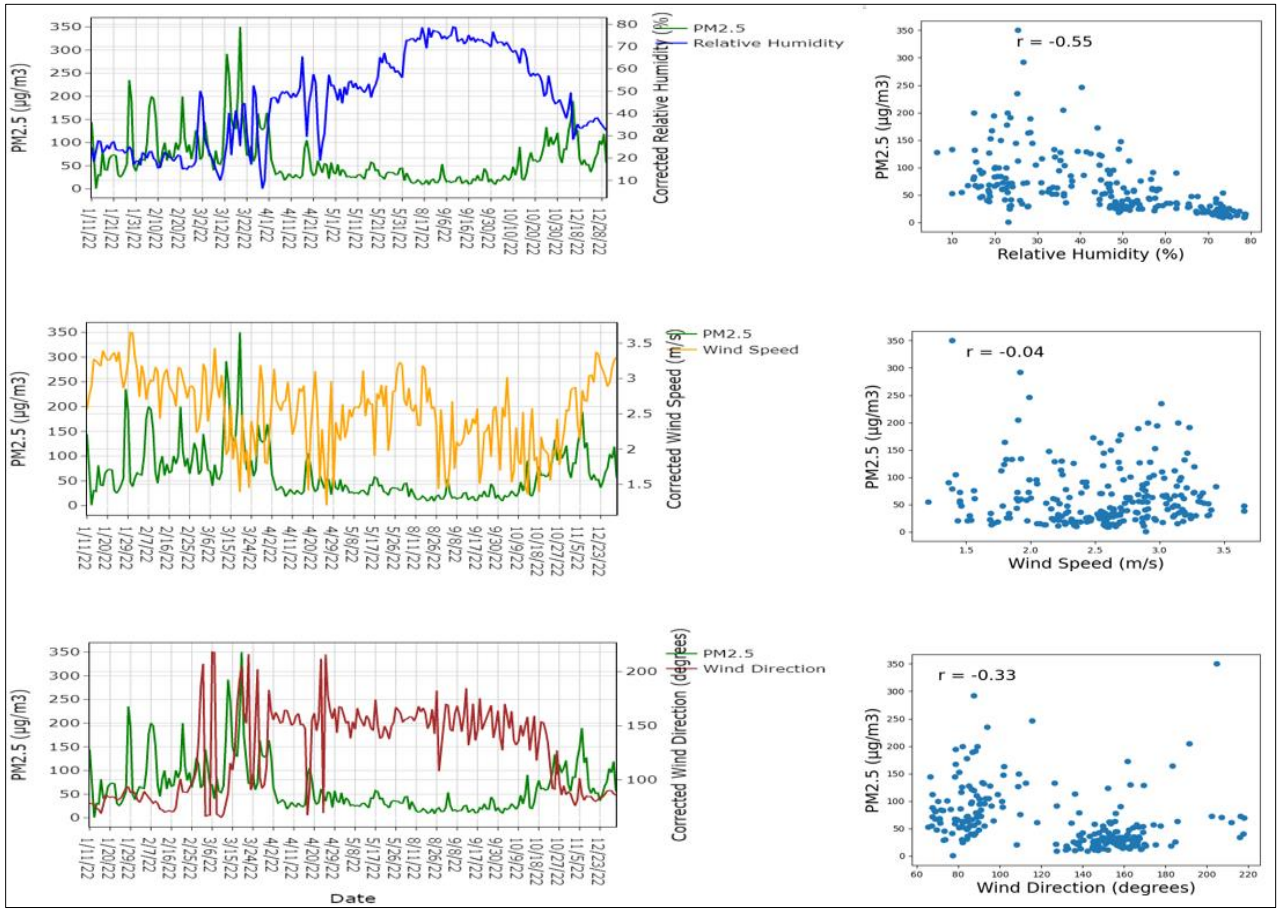

Figure 15: $PM_{2.5}$ and corrected Era5-Land parameters

Table 3*:* Summary of the Pearson correlation coefficients between $PM_{2.5}$ and AOD and corrected satellite weather parameters at Ouaga 2000.

Parameters	**r**
AOD-$PM_{2.5}$	0.72
Relative Humidity-$PM_{2.5}$	-0.55
Temperature-$PM_{2.5}$	0.11
Precipitation-$PM_{2.5}$	-0.26
Wind Speed-$PM_{2.5}$	-0.04
Wind Direction-$PM_{2.5}$	-0.33

3.5 Statistical Regression Models

Two models are developed; SLR model and MLR model. $PM_{2.5}$ and AOD at the Ouaga 2000 are used to developed the SLR model as shown in **Figure 16**, to determine how MODIS AOD can be a predictor of surface $PM_{2.5}$. The simple linear regression model developed is:

$$PM_{2.5} = 75.61 \times AOD + 38.36 \quad (13)$$

The model has an R^2 of 0.52, indicating that MODIS AOD explains about half of the variations of surface $PM_{2.5}$. The RMSE of the model is 38.3 μg/m^3 and the significance F is 1.23 x 10^{-22} (less than $\alpha = 0.05$). These findings are similar to the findings by van Donkelaar et al. (2010b). They obtained R^2 of 0.59 between MODIS AOD and $PM_{2.5}$ over Eastern China. Wang & Christopher (2003) obtained R^2 of 0.49 at seven locations in Jefferson County, Alabama. Koelemeijer et al. (2006) also obtained R^2 of 0.36 between AOD and $PM_{2.5}$ at some locations in Europe.

To understand the influence of meteorological parameters on surface $PM_{2.5}$, the corrected satellite weather parameters; relative humidity, temperature, precipitation, wind speed, and wind direction at the same location (Ouaga 2000) are introduced into the linear equation to produce the multiple linear regression model below

$$PM_{2.5} = 174.11 + 69.27 \times AOD - 1.65 \times T - 1.00 \times RH - 14.04 \times WS - 1.54 \times Precip - 0.08 \times WD \quad (14)$$

The MLR model has an R^2 of 0.67 and an RMSE of 33.7 μg/m^3 with a significance F of 8.85 x 10^{-30} (less than $\alpha = 0.05$) as shown in **Figure 16**. One explanation for this is that the inclusion of

meteorological parameters leads to better prediction of surface $PM_{2.5}$, the MLR model explains 0.67 of the variations of surface $PM_{2.5}$ with smaller RMSE compared to the SLR model. These findings are very consistent with the findings of Tian & Chen (2010). Their regression model was able to explain 0.65 of the variations of ground-$PM_{2.5}$ after adding relative humidity and temperature. Also, Gharibzadeh & Saadat Abadi (2022) multiple linear regression model explained about 60 % (R^2 of 0.6) of the changes in $PM_{2.5}$ over Ahvaz, Iran. Similarly, Ma et al. (2014) multiple linear regression model predicted 0.64 the variations of surface $PM_{2.5}$ in China with RMSE of 32.98 $\mu g/m^3$.

Though the findings clearly show that the addition of meteorological parameters into the SLR model to produce the MLR model improves the model's performance, this model assumes a linear relationship between the parameters and $PM_{2.5}$. From the Pearson correlation performed on the parameters and the $PM_{2.5}$, it is observed that most of the variations of $PM_{2.5}$ are not directly explained by the parameters (less Pearson correlation coefficients, nonlinear relationships) and this resulted in the model explaining just 0.67 the variations of surface $PM_{2.5}$.

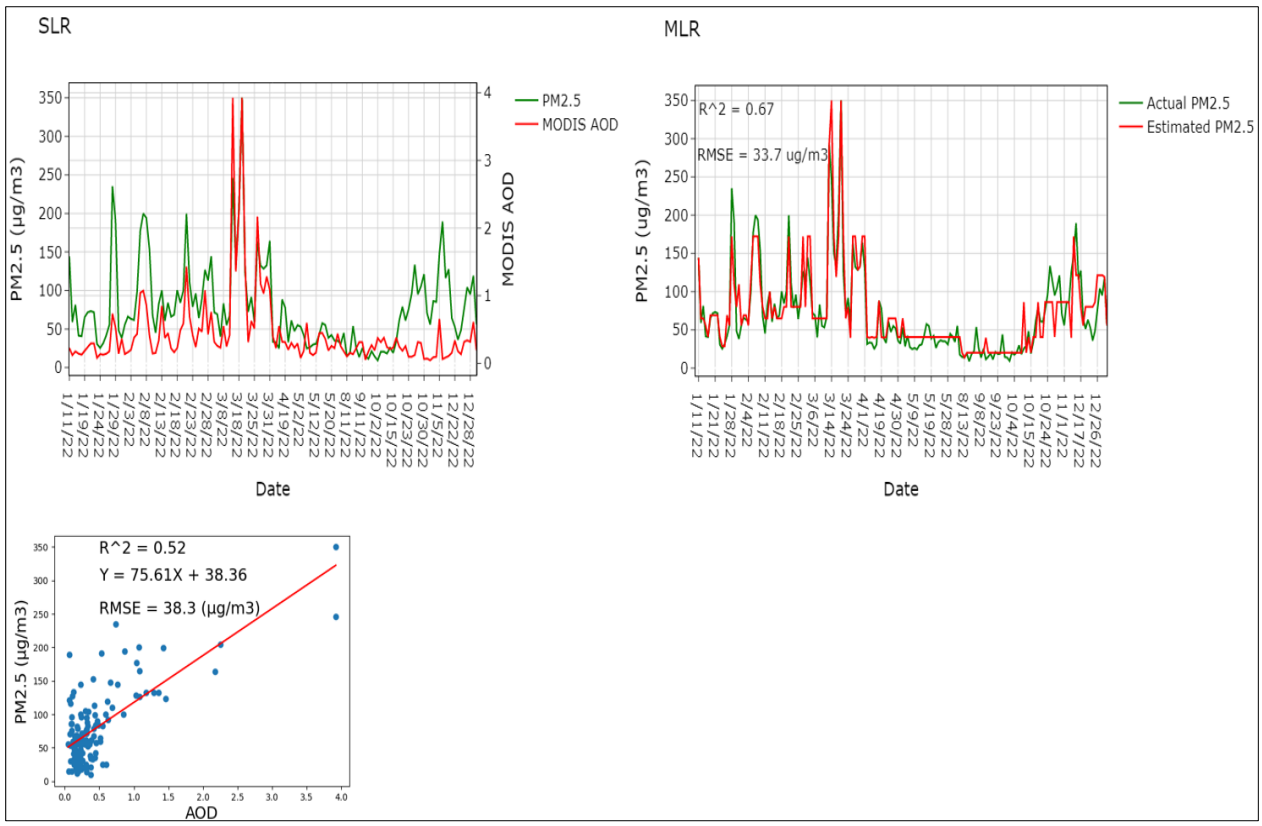

Figure 16: Statistical Regression Models

3.6 Machine Learning Models

3.6.1 Only AOD parameter as input in models

Three nonlinear machine learning models (DT, RF, and XGBoost) models are developed at the same location (Ouaga 2000) where the SLR and MLR models are developed. The nonlinear models are developed first using only AOD as input parameter. This is done to determine how only MODIS AOD can be a predictor of $PM_{2.5}$ in the nonlinear models. As shown in **Figure 17**, the DT, RF, and XGBoost model has R^2 of 0.56, 0.58, and 0.54 respectively and RMSE of 39.5 $\mu g/m^3$, 42.1 $\mu g/m^3$, and 48.9 $\mu g/m^3$ respectively. The performance of these models with only AOD is better than the SLR model with only AOD. The models explain more than half of the variations of surface $PM_{2.5}$. However, their RMSE are high compared to SLR model. Fu et al. (2022) observed similar performance across China when only AOD was used in their random forest model, their model had R^2 of 0.49.

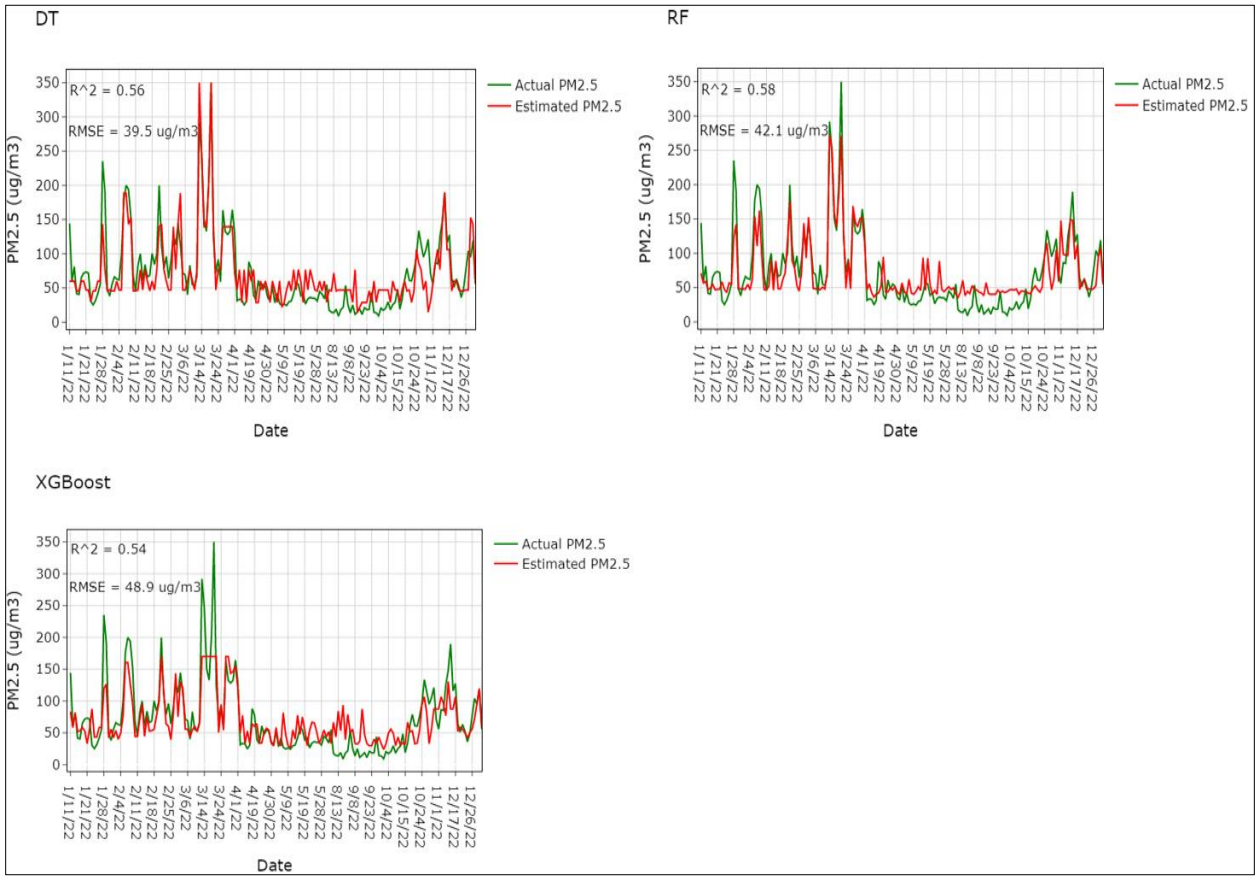

Figure 17: Only MODIS AOD as input parameter in nonlinear models

3.6.2 All parameters as input in models

To understand the influence of meteorological factors on surface $PM_{2.5}$, the corrected satellite meteorological parameters; relative humidity, temperature, precipitation, wind speed, and wind direction are added into the models. As shown in **Figure 18,** when the meteorological parameters are introduced into the models, their performance increases significantly. The XGBoost model outperform all the models explaining 0.87 of the variations of surface $PM_{2.5}$ with lower RMSE of 15.8 $\mu g/m^3$, followed by the random forest model explaining 0.85 of the variations of surface $PM_{2.5}$ with RMSE of 16.6 $\mu g/m^3$ and then the decision tree model explaining 0.70 of the variations of surface $PM_{2.5}$ RMSE of 34.0 $\mu g/m^3$. These findings clearly show that the nonlinear models perform better than the linear models in the estimation of surface $PM_{2.5}$. These findings are consistent with the findings by Joharestani et al. (2019), they found that the XGBoost outperformed all their models in estimating $PM_{2.5}$ in Tehran's urban area with R^2 of 0.81 and RMSE of 13.58 $\mu g/m^3$.

McFarlane et al. (2021) random forest model for correcting low-cost sensors $PM_{2.5}$ data in Kampala, Uganda, had similar performance (R^2 of 0.86). Also, Zhang et al. (2021) random forest model for estimating surface $PM_{2.5}$ around Gauteng Province, South Africa gave similar performance (R^2 of 0.80 and RMSE of 9.4 $\mu g/m^3$) after the addition of meteorological data.

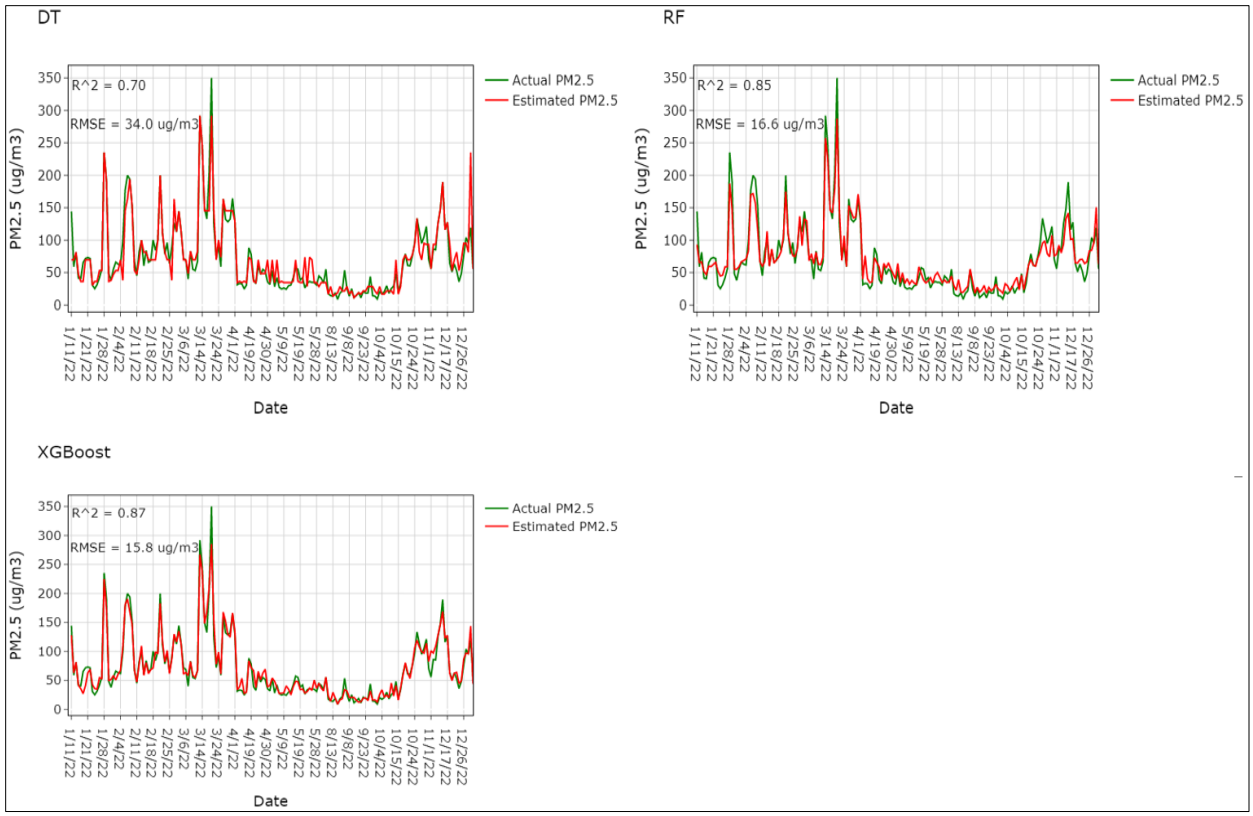

Figure 18: All parameters as input in models

Figure 19 shows the feature importance in the DT, RF, and XGBoost model. In all the models, MODIS AOD is the most important feature in $PM_{2.5}$ estimation, followed by relative humidity and temperature. Precipitation is the less important parameter in the models' estimation. This means that the impact of precipitation in explaining the variability of $PM_{2.5}$ is less significant compared to AOD, relative humidity, temperature, wind speed, and wind direction.

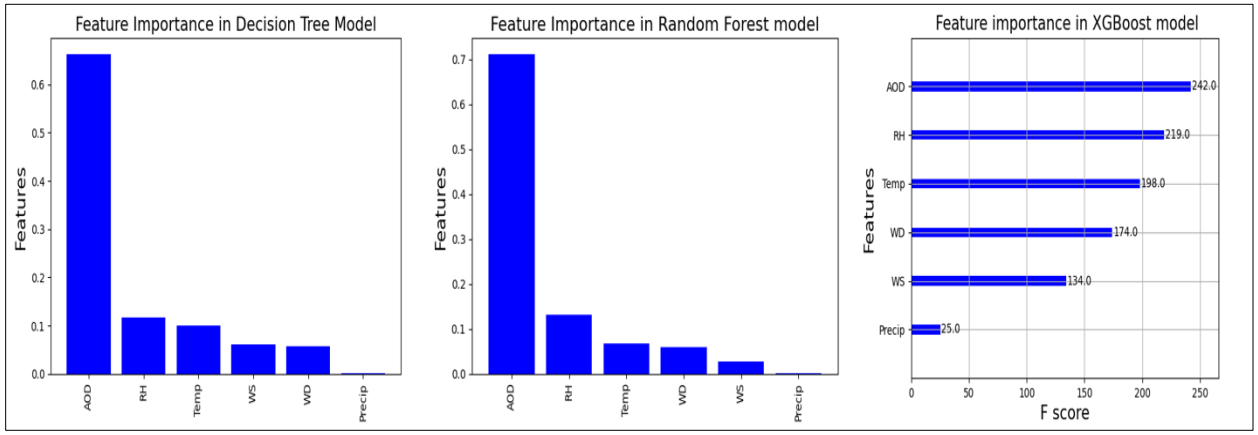

Figure 19: Nonlinear models feature importance

3.6.3 Semi-supervised XGBoost model

Figure 20 shows the performance of the semi-supervised XGBoost model. The model has an R^2 of 0.97 and an RMSE of 8.3 μg/m^3 after a five fold cross-validation, indicating that the model explains 0.97 of the variations of $PM_{2.5}$ with lower RMSE. These findings are similar to the findings by Bougoudis et al. (2016), their semi-supervised ANN model explained about 0.9 of the variations of air pollutants in Athens, Greece. Similarly, the semi-supervised KNN model proposed by Zhao et al. (2023) in China had R^2 of 0.97. **Figure 20** also shows the estimation of $PM_{2.5}$ by the model on the whole data (labeled and unlabeled) after testing.

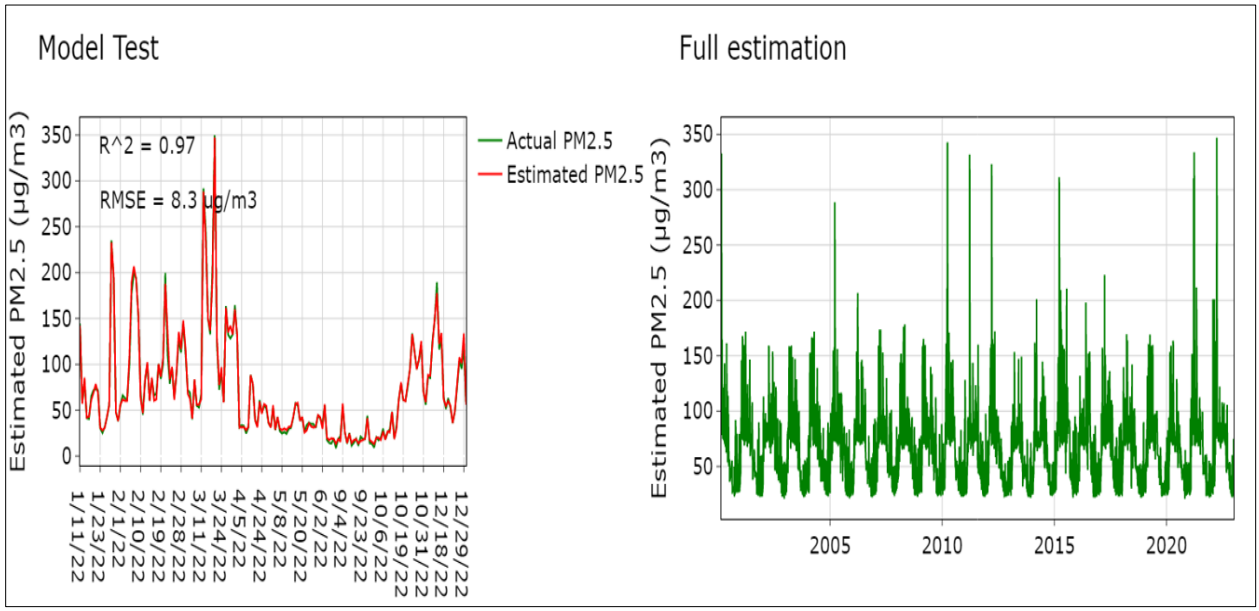

Figure 20: Semi-supervised XGBoost Model performance

Table 4: Summary of the performance of all models

Model	Input	R^2	RMSE (μg/m^3)
SLR	AOD	0.52	38.3
MLR	AOD and weather parameters	0.67	33.7
DT	AOD	0.56	39.5
DT	AOD and weather parameters	0.70	34.0
RF	AOD	0.58	42.1
RF	AOD and weather parameters	0.85	16.6
XGBoost	AOD	0.54	48.9

XGBoost	AOD and weather parameters	0.87	15.8
Semi-Supervised XGBoost	AOD and weather parameters	0.97	8.3

3.7 PM2.5 estimation in other areas of Ouagadougou

In order to analyze and study the distribution of $PM_{2.5}$ in the whole city, the semi-supervised XGBoost model is applied to estimate surface $PM_{2.5}$ in the other 15 areas in the city of Ouagadougou, since these areas are in the same climate zone as Ouaga 2000. The $PM_{2.5}$ estimation was done for the same period the model was trained so the growth and distribution of $PM_{2.5}$ in Ouagadougou are studied.

3.7.1 Average of daily and monthly trend of estimated $PM_{2.5}$ in Ouagadougou

In **Figure 21,** the average of daily estimated $PM_{2.5}$ from the 16 areas (including Ouaga 2000) in Ouagadougou for the period 2000-2022 is shown. It is observed that in the whole city, days in the dry season (November to April) are associated with high $PM_{2.5}$ concentrations and day days in the rainy season (May to October) are associated with low $PM_{2.5}$ concentrations and it is the same trend repeating every year. March has days with higher $PM_{2.5}$ concentrations reaching a maximum of 350 $\mu g/m^3$ whereas August has days with $PM_{2.5}$ concentrations as low as 16 $\mu g/m^3$. The lower $PM_{2.5}$ concentrations in the rainy season are due to wet deposition and gravitational settling of the particles caused by precipitation whereas the higher $PM_{2.5}$ concentrations in the dry season are due to the dust from the Sahara desert transported by Harmattan winds which reaches its peak in February-March. Also, dust from unpaved roads and biomass burning are the major contributors to the higher concentrations in the dry season (Nana et al., 2012). These variations are consistent with our second hypothesis that $PM_{2.5}$ concentrations in Ouagadougou vary from season to season.

The monthly trend increases from September till March when the harmattan reaches its peak and then decreases till August when precipitation is frequent. The estimated daily $PM_{2.5}$ concentrations in the whole of Ouagadougou on average in every rainy season is 2 to 4 times higher than the WHO 24-hour limit of 15 $\mu g/m^3$ and in every dry season, the estimated $PM_{2.5}$ concentrations on average are 2 to 22 times higher the WHO 24-hour limit. However, most of the days in the rainy season meet the US EPA recommended limit of 35 $\mu g/m^3$ per day (US EPA,

2012). These findings are consistent with the findings from Nana et al. (2012); Ouarma et al. (2020) during the 2018-2019 measurement campaigns in Ouagadougou. They observed higher concentrations of $PM_{2.5}$ in the dry season and lower PM2.5 concentrations in the rainy season with the $PM_{2.5}$ concentrations in the rainy season being 2 to 3 times higher than the WHO recommended limit at many sites.

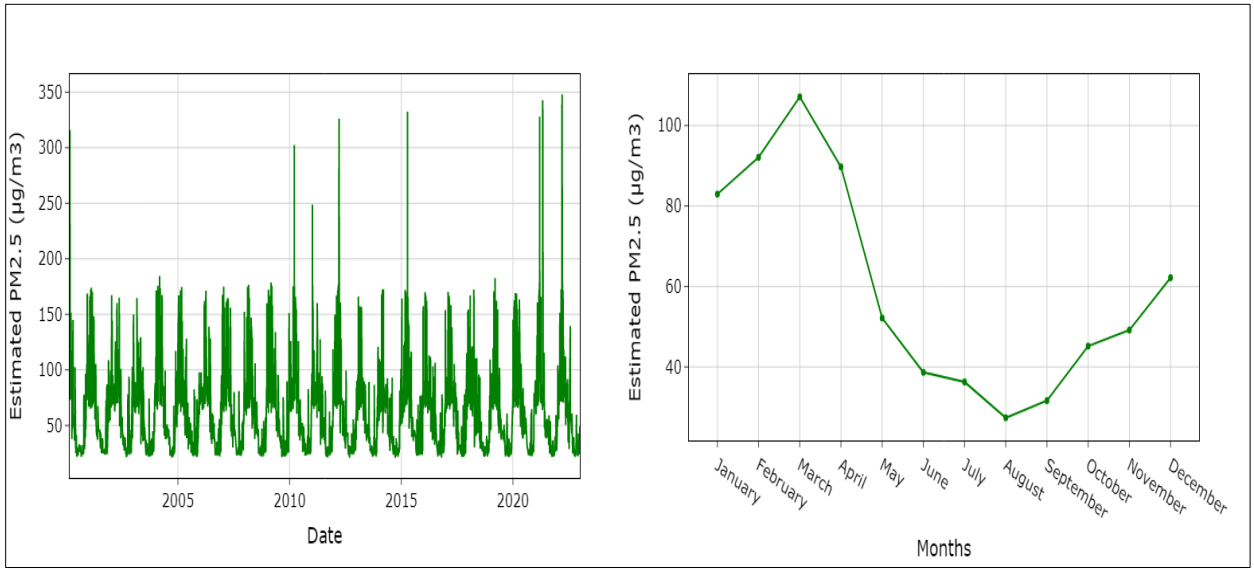

Figure 21: Average of the daily and monthly trend of estimated $PM_{2.5}$ in Ouagadougou

3.7.2 Average yearly trend of estimated $PM_{2.5}$ in Ouagadougou

Figure 22 shows the average yearly trend of estimated $PM_{2.5}$ in Ouagadougou. The yearly trend of $PM_{2.5}$ in the city is not direct, it increases and decreases but in general there is slight increasing trend. This result is in agreement with the findings by Ouarma et al. (2020), they found that particulate matter concentrations varied throughout their study period. The trend largely depends on the intensity of dust from the Sahara desert and the variability of weather conditions in a given year (Lindén et al., 2012). The higher the intensity of dust from the Sahara desert in a given year, the higher the concentrations of $PM_{2.5}$ in that year. It also depends on the intensity of emissions from unpaved roads, heavy traffic, and industrial activities in a given year (Boman et al., 2009; Nana et al., 2012; Ouarma et al., 2020). Great increases in the trend are observed in 2004, 2010, and 2015 whilst great decreases are observed in 2002, 2003 and 2020.

The observed low levels of $PM_{2.5}$ concentrations in Ouagadougou during the years 2002 and 2003 can be attributed to the relatively minimal presence of cars and industrial activities within

the city during that period. The limited number of cars on the roads during that time meant a lower release of particulate matter from vehicle exhausts, which is a significant contributor to $PM_{2.5}$ concentrations in urban areas. Additionally, the minimal presence of industrial activities, such as factories and manufacturing plants, resulted in a small amount of pollutants released into the atmosphere.

In 2020 where there was lockdown due to Covid-19, it is observed that the $PM_{2.5}$ trend decreased significantly. This is because of the reduced in industrial activities and heavy traffic due to the lockdown. The findings are similar to the findings observed by McFarlane et al. (2021). They observed lower $PM_{2.5}$ concentrations in 2020 in Kinshasa and Brazzaville due to the Covid-19 lockdown. Similarly, Shi & Brasseur (2020) observed lower $PM_{2.5}$ concentrations in China during the Covid-19 quarantine period.

From 2000-2022, the average annual estimated $PM_{2.5}$ concentrations range from 58.2 $\mu g/m^3$ to 72.1 $\mu g/m^3$, which is 11 to 14 times higher than the WHO guidelines of 5 $\mu g/m^3$ annually and 4 to 6 times higher than the U.S. EPA guidelines of 12 $\mu g/m^3$ annually. These results are consistent with IQAir 2022 report which states that annual $PM_{2.5}$ concentrations in Ouagadougou are over 10 times the WHO guidelines (IQAir, 2022). These results indicate very poor air quality in Ouagadougou.

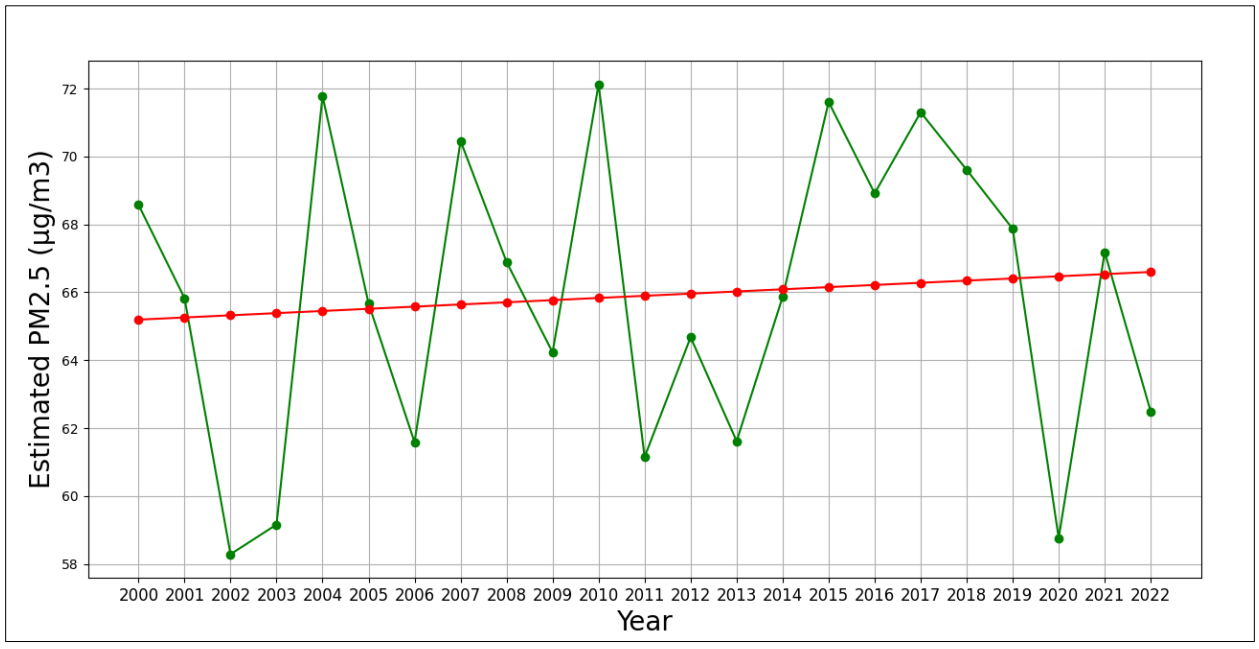

Figure 22: Average yearly trend of estimated $PM_{2.5}$ in Ouagadougou

3.8 Spatial Distribution of $PM_{2.5}$ in Ouagadougou.

3.8.1 Dry season 2000-2005

Figure 23 shows the spatial distribution of estimated $PM_{2.5}$ in dry season 2000-2005 in Ouagadougou. The minimum estimated $PM_{2.5}$ concentrations in all the areas ranged between 40 $\mu g/m^3$ and 55 $\mu g/m^3$ and the maximum estimated $PM_{2.5}$ concentrations were all above 105 $\mu g/m^3$. The mean estimated $PM_{2.5}$ concentrations in all the areas of the city were between 70 $\mu g/m^3$ and 80 $\mu g/m^3$ except in Yagma, Tengadogo, and Ouaga 2000 where the mean estimated $PM_{2.5}$ concentrations were slightly higher, between 80 $\mu g/m^3$ and 85 $\mu g/m^3$. Yagma and Ouaga 2000 $PM_{2.5}$ concentrations reached 90 $\mu g/m^3$ and 100 $\mu g/m^3$ respectively in the dry season of 2000 as shown in **Figure 31** in appendices. Yagma, Tengadogo, and Ouaga 2000 are towns at the periphery of Ouagadougou, the slightly higher mean $PM_{2.5}$ concentrations in these areas could depict greater emissions from unpaved roads. Another reason could be emissions from agricultural activities. The high $PM_{2.5}$ concentrations in every area of the city in the dry season are majorly due to the dust transported from the Sahara desert by the Harmattan winds from November to March.

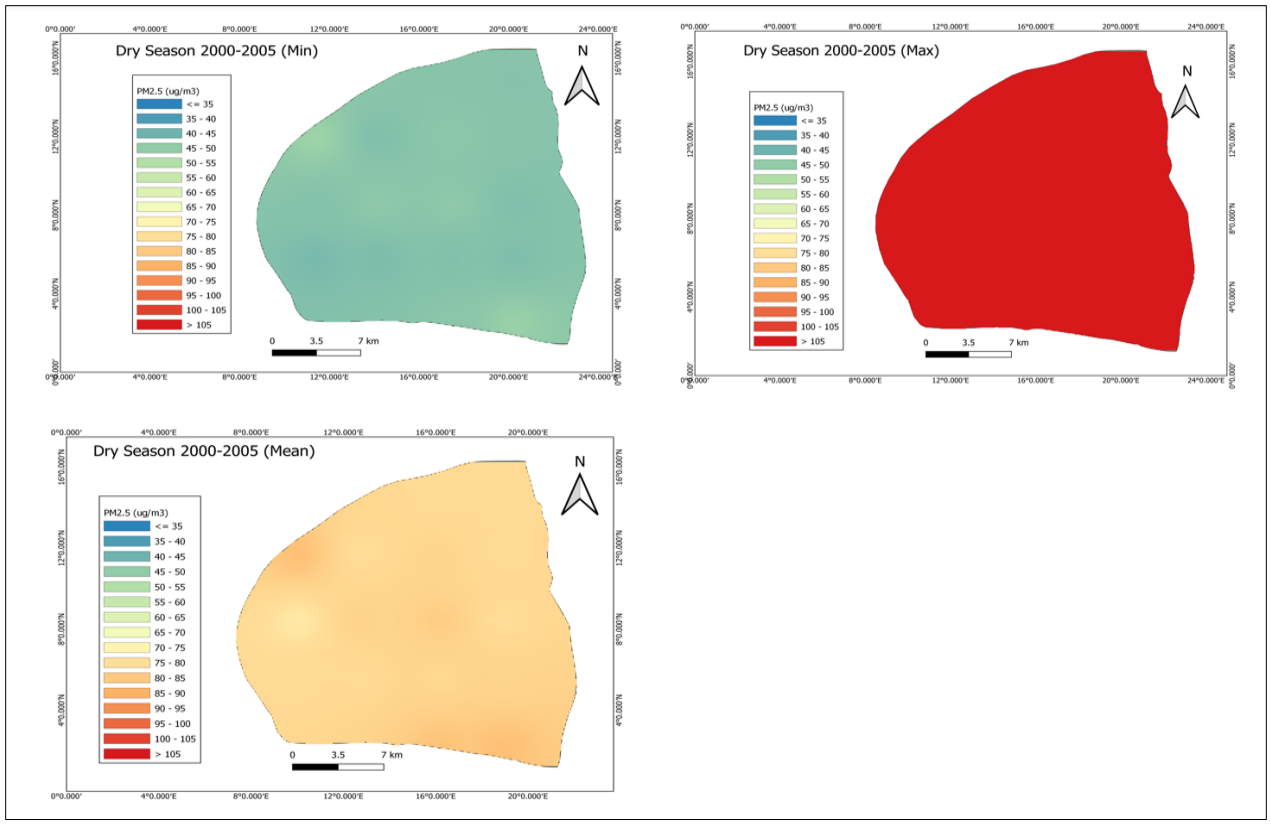

Figure 23: Spatial distribution of estimated $PM_{2.5}$ in dry season 2000-2005

3.8.2 Rainy Season 2000-2005

Figure 24 shows the spatial distribution of estimated $PM_{2.5}$ in rainy season 2000-2005 in Ouagadougou. Generally, the estimated $PM_{2.5}$ concentrations in the rainy were lower compared to the dry season. The minimum estimated concentrations in all the areas were below or equal to 35 $\mu g/m^3$. The maximum estimated concentrations of $PM_{2.5}$ in all the areas were between 50 $\mu g/m^3$and 60 $\mu g/m^3$. The mean estimated $PM_{2.5}$ concentrations in all areas of the city were within the same range, between 35 $\mu g/m^3$ and 40 $\mu g/m^3$. In rainy season of 2000 and 2001, Yagma and Ouaga 2000 had extremely lower mean PM2.5 concentrations below 35 $\mu g/m^3$ as shown in **Figure 31 and Figure 32** in appendices. The lower estimated $PM_{2.5}$ concentrations clearly explains the influence of precipitation on the concentrations of $PM_{2.5}$. Precipitation leads to wet deposition which removes $PM_{2.5}$ particles in the atmosphere. Also, vegetation growth in the rainy season tend to absorb and remove pollutants in the air hence lowering $PM_{2.5}$ concentrations.

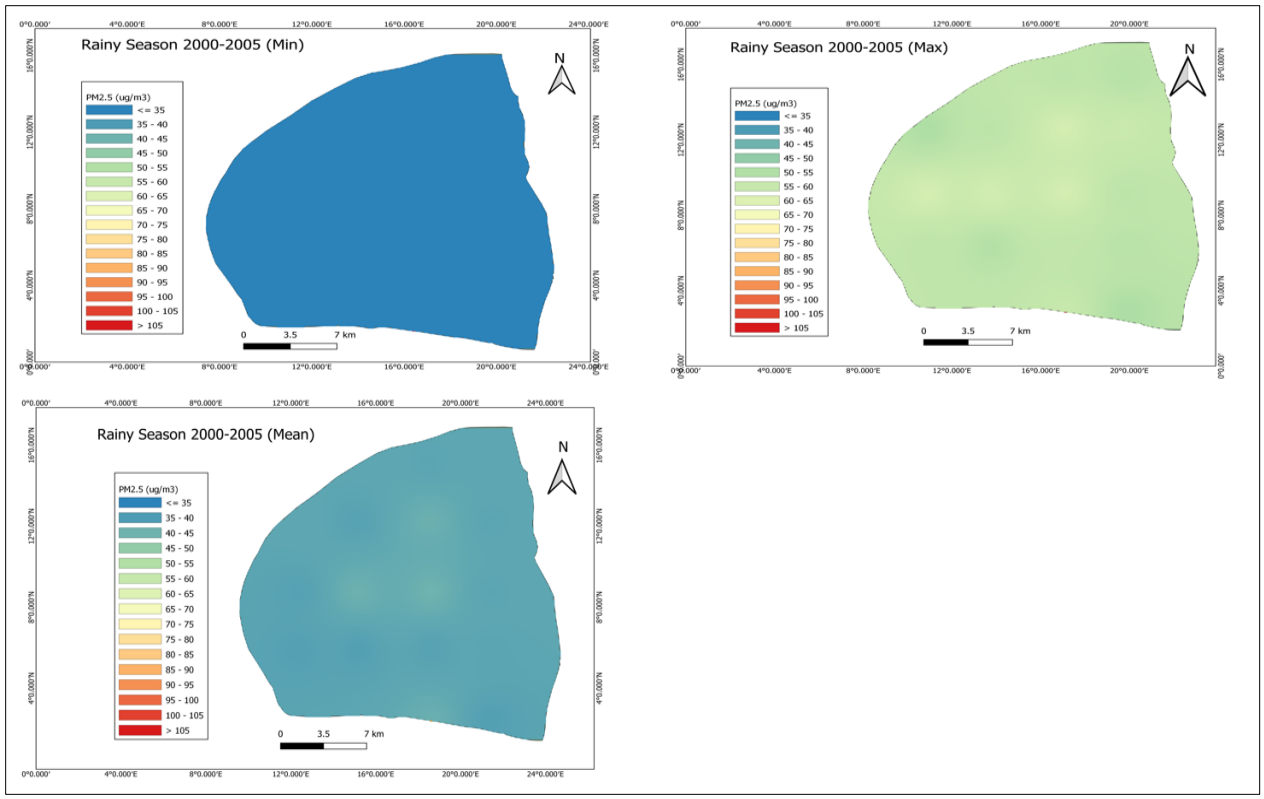

Figure 24: Spatial distribution of estimated $PM_{2.5}$ in rainy season 2000-2005

3.8.3 Dry season 2006-2011

Figure 25 shows the spatial distribution estimated $PM_{2.5}$ in dry season 2006-2011 in Ouagadougou. The minimum estimated $PM_{2.5}$ concentrations were between 40 $\mu g/m^3$ and 60 $\mu g/m^3$ for all the areas in Ouagadougou. However, the maximum estimated $PM_{2.5}$ concentrations in each area of the city were all above 105 $\mu g/m^3$. The high concentrations of $PM_{2.5}$ in the dry season are largely due to the dust from the Sahara desert transported by the Harmattan winds. The mean estimated $PM_{2.5}$ concentrations in the city were between 70 $\mu g/m^3$ and 75 $\mu g/m^3$ (about 6 % decrease from their $PM_{2.5}$ levels in the dry season of 2000-2006) except Gounghin, Ouagadougou International Airport, Kamboinsi, and Tanghin which had slightly higher mean $PM_{2.5}$ concentrations between 75 $\mu g/m^3$ and 80 $\mu g/m^3$ (same levels as dry season of 2000-2006).

Ouagadougou International Airport, Kamboinsi, and Tanghin are home to many commercial activities with high volumes of traffic which contribute to higher levels of $PM_{2.5}$ in these areas. Gounghin is one of the most established industrial zones in the city, home to many processing plants and factories hence emissions from industrial combustion activities and heavy trucks contribute to the higher $PM_{2.5}$ levels in the area. The 6 % decrease in $PM_{2.5}$ concentrations in other areas of the city largely depended on the intensity of dust transported from the Sahara desert and increment of paved roads within these areas.

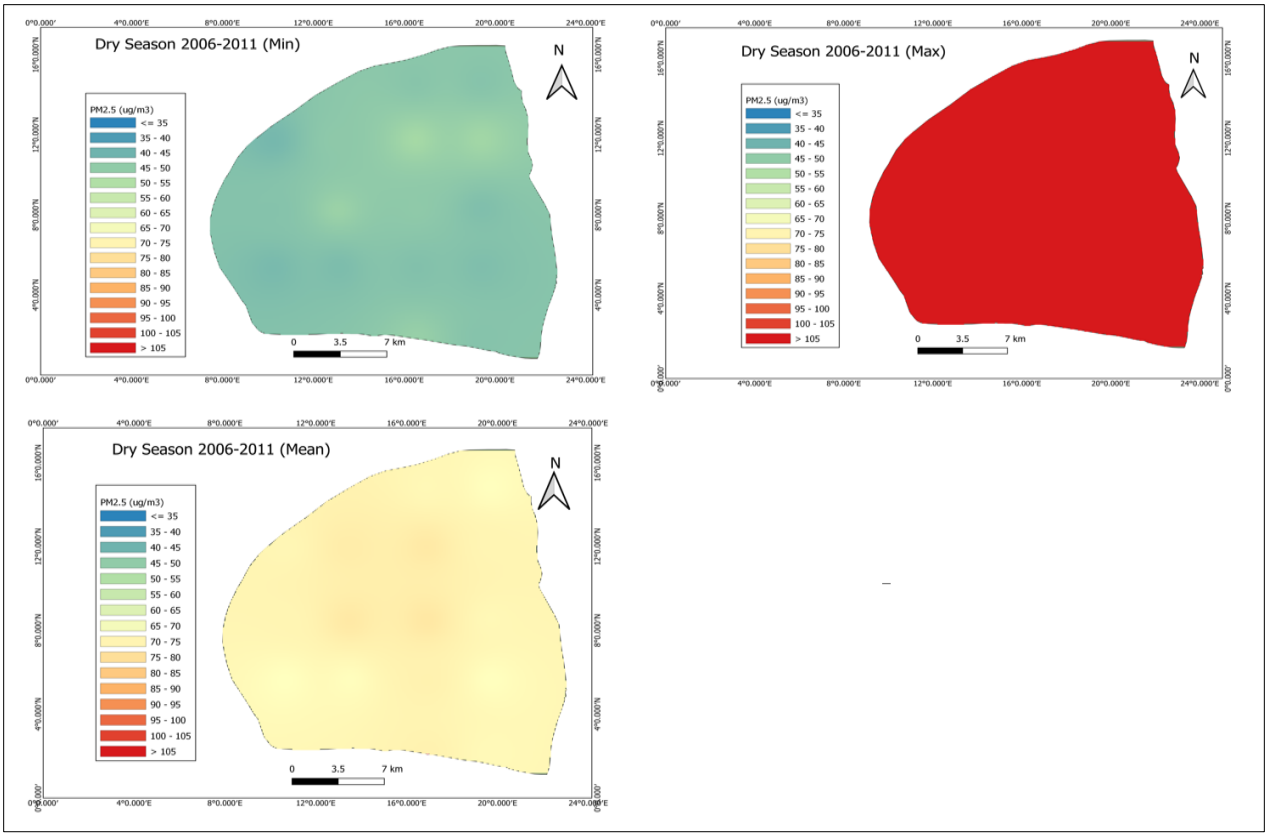

Figure 25: Spatial distribution of estimated $PM_{2.5}$ in rainy season 2006-2011

3.8.4 Rainy season 2006-2011

Figure 26 shows the spatial distribution of estimated $PM_{2.5}$ in rainy season 2006-2011 in Ouagadougou. $PM_{2.5}$ concentrations in the rainy were lower with minimum concentrations below 40 $\mu g/m^3$ and maximum $PM_{2.5}$ concentrations between 55 $\mu g/m^3$ and 65 $\mu g/m^3$ in all areas of the city except in Ouagadougou International Airport and Kossodo where maximum $PM_{2.5}$ concentrations were between 65 $\mu g/m^3$ and 70 $\mu g/m^3$. The mean estimated $PM_{2.5}$ were between 35 $\mu g/m^3$ and 45 $\mu g/m^3$ in all areas except in Gounghin, Kossodo, Ouagadougou International Airport, and Kamboinsi where the mean $PM_{2.5}$ concentrations were between 55 $\mu g/m^3$ and 60 $\mu g/m^3$ (about 50 % increment from the $PM_{2.5}$ levels in the rainy season of 2000-2005). Kossodo is a rapidly growing industrial area in the city. It is a home to plastics manufacturing, metalworking, textile production, and food processing. These activities lead to the high $PM_{2.5}$ concentrations in the area. These findings are in agreement with the findings by (Nana et al., 2012), they observed high concentrations of gaseous pollutants at industrial sites (Kossodo and Gounghin) and downtown.

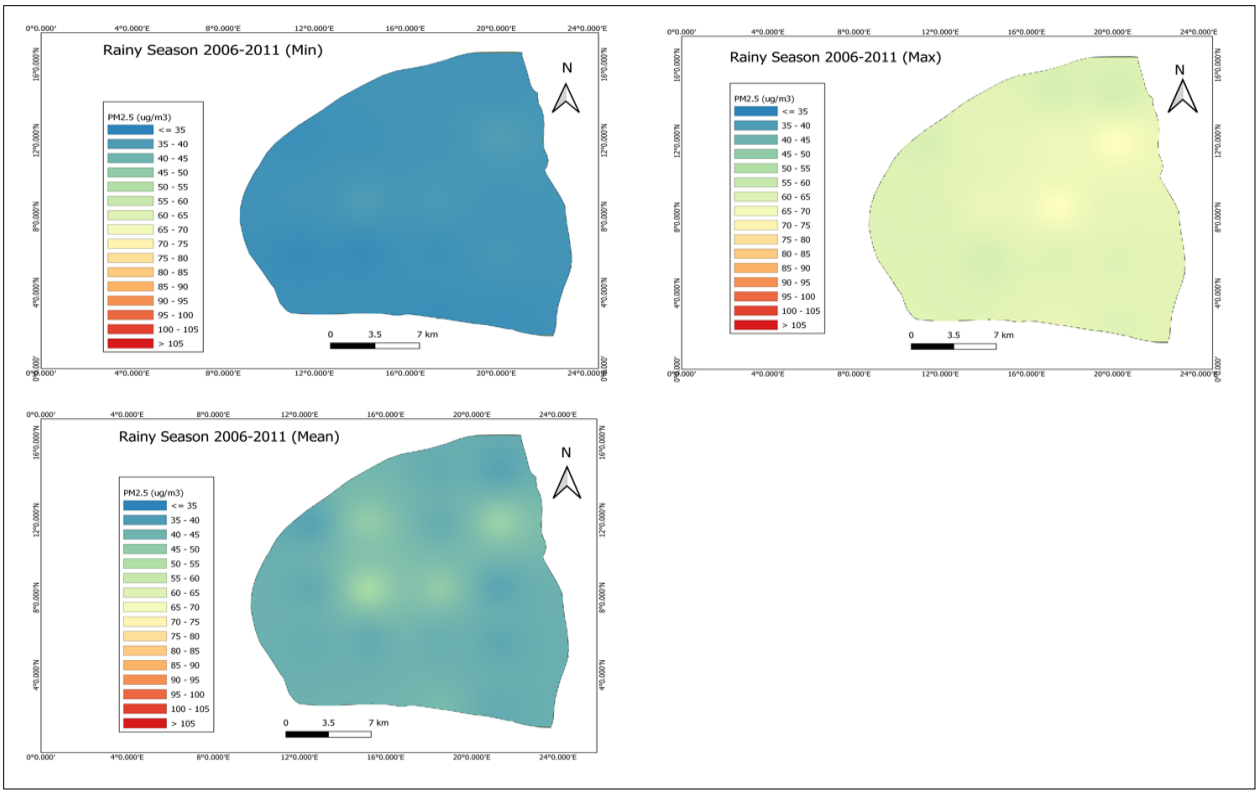

Figure 26: Spatial distribution of estimated $PM_{2.5}$ in rainy season 2006-2011

3.8.5 Dry season 2012-2017

Figure 27 shows the spatial distribution of estimated $PM_{2.5}$ in dry season 2011-2017 in Ouagadougou. From this period, it was obviously clear that the industrial areas (Gounghin and Kossodo) are taking the lead in $PM_{2.5}$ concentrations in the city. The minimum estimated $PM_{2.5}$ concentrations in Gounghin and Kossodo were between 65 $\mu g/m^3$ and 70 $\mu g/m^3$, about 16 % increment from their minimum $PM_{2.5}$ concentrations in the dry season of 2006-2011. The mean estimated $PM_{2.5}$ concentrations in these areas were between 85 $\mu g/m^3$ and 95 $\mu g/m^3$ which was about 19 % increment from their $PM_{2.5}$ levels in 2006-2011. Goughin and Kossodo consistently had high $PM_{2.5}$ concentrations in the dry season of 2014, 2015, 2016, and 2017 as shown in **Figure 45, Figure 46, Figure 47, and Figure 48** in appendices. The mean estimated $PM_{2.5}$ concentrations of Tanghin and Ouagadougou International Airport were between 80 $\mu g/m^3$ and 85 $\mu g/m^3$, 6 % increment from their $PM_{2.5}$ levels in the dry season of 2006-2011. However, the mean estimated $PM_{2.5}$ concentrations in Karpala, Dassasgo, Roumtenga, Zagtouli, Koumdanyore, Kamboinsi, and Yagma were not significantly different than their $PM_{2.5}$ levels in the dry season of 2006-2011.

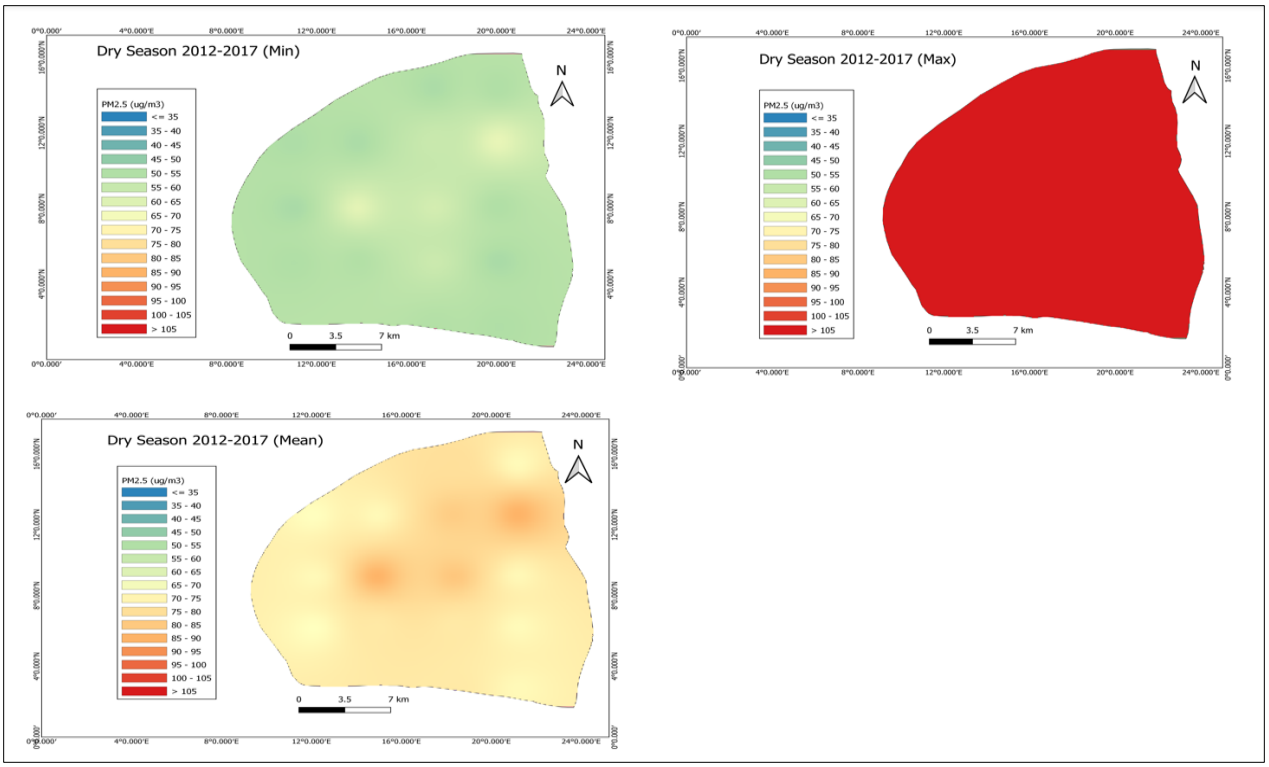

Figure 27: Spatial distribution of estimated $PM_{2.5}$ in dry season 2012-2017

3.8.6 Rainy season 2012-2017

Figure 28 shows the spatial distribution of estimated $PM_{2.5}$ in rainy season 2012-2017. The minimum concentrations of estimated $PM_{2.5}$ in the rainy season were between 40 $\mu g/m^3$ and 45 $\mu g/m^3$ for the industrial areas and areas in the center of Ouagadougou whereas in the remaining areas the minimum estimated $PM_{2.5}$ were below 35 $\mu g/m^3$. The maximum $PM_{2.5}$ concentrations estimated at the industrial areas and the center of Ouagadougou were between 65 $\mu g/m^3$ and 75 $\mu g/m^3$, about 7 % increment from their maximum $PM_{2.5}$ concentrations in the rainy season of 2006-2011. The mean estimated $PM_{2.5}$ concentrations in Gounghin and Kossodo were between 65 $\mu g/m^3$ and 70 $\mu g/m^3$, about 17 % increment from their levels in the rainy season of 2006-2011. These two areas consistently had high $PM_{2.5}$ concentrations in the rainy season of 2015, 2016, and 2017 as shown in **Figure 46, Figure 47, and Figure 48** in appendices. These high concentrations of $PM_{2.5}$ in these areas are due to emissions from industrial activities. Tanghin, Patte d'Oie, and Ouagadougou International Airport had mean estimated $PM_{2.5}$ between 60 $\mu g/m^3$ and 65 $\mu g/m^3$. The high $PM_{2.5}$ concentrations at these areas are due to emissions from heavy traffic since these areas are in the center of the city. However, concentrations of $PM_{2.5}$ in the remaining areas of the city lowered by about 11 % than their levels in the rainy of 2006-2011. This decrease in $PM_{2.5}$

concentrations in the rainy season largely depends on the intensity of rainfall occurring during the period. More rainfall leads to more wet depositions and vegetation growth hence lowering $PM_{2.5}$ concentrations in the atmosphere (Tai et al., 2012).

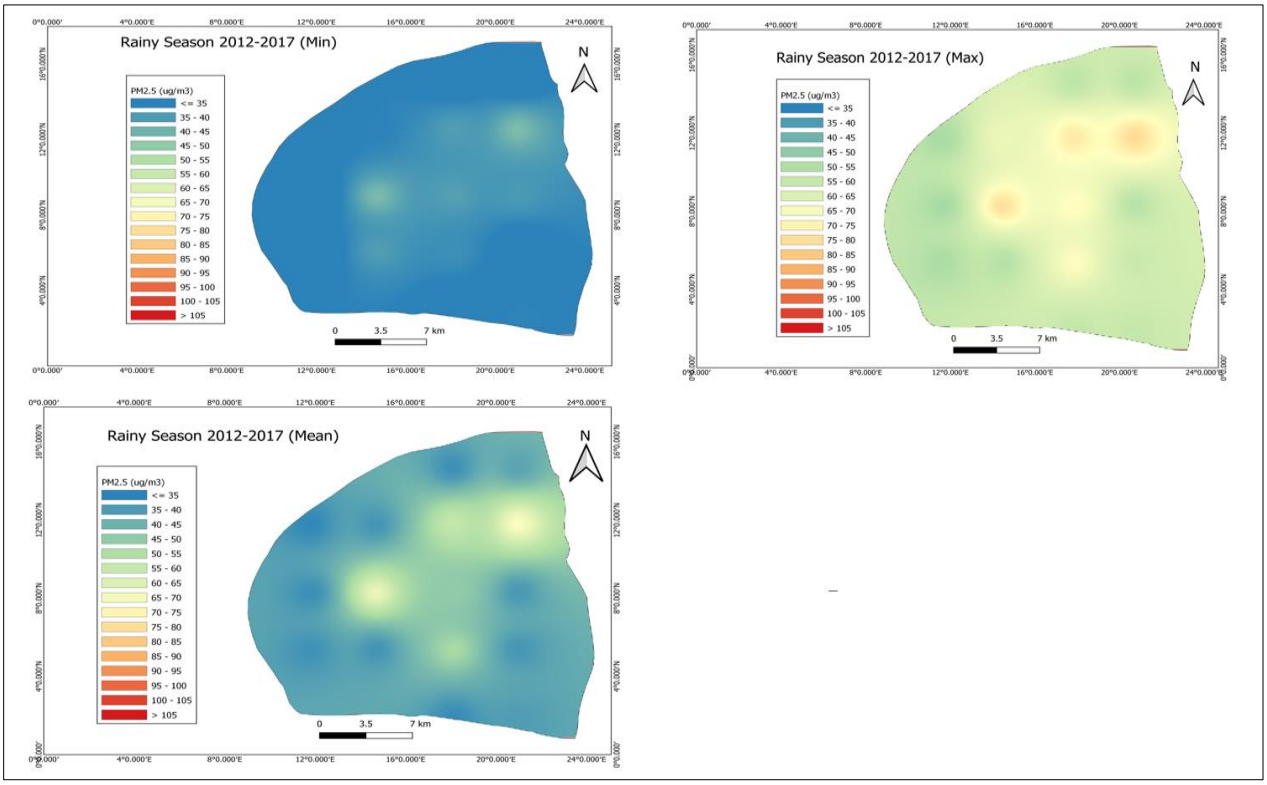

Figure 28: Spatial distribution of estimated $PM_{2.5}$ in dry season 2012-2017

3.8.7 Dry season 2018-2022

Figure 29 shows the spatial distribution of $PM_{2.5}$ in the dry season of 2018-2022. The minimum concentrations of estimated $PM_{2.5}$ concentrations are between 40 $\mu g/m^3$ and 60 $\mu g/m^3$ except Gounghin and Kossodo which are between 65 $\mu g/m^3$ and 70 $\mu g/m^3$. The mean estimated $PM_{2.5}$ concentrations in these areas decreased by 11 % than their levels in the dry season of 2012-2017. This decrease in $PM_{2.5}$ concentrations in the industrial areas are due to the Covid-19 lockdown where industrial activities decreased. Similarly, the estimated $PM_{2.5}$ concentrations in the center of the city decreased by 6 % than their levels in the dry season of 2018-2022. This decrease is also due to the Covid-19 lockdown which constrained the movement of people hence reducing traffic emissions. However, the maximum estimated $PM_{2.5}$ concentrations in the city are still above 105 $\mu g/m^3$ mainly due to the Sahara desert transported dust.

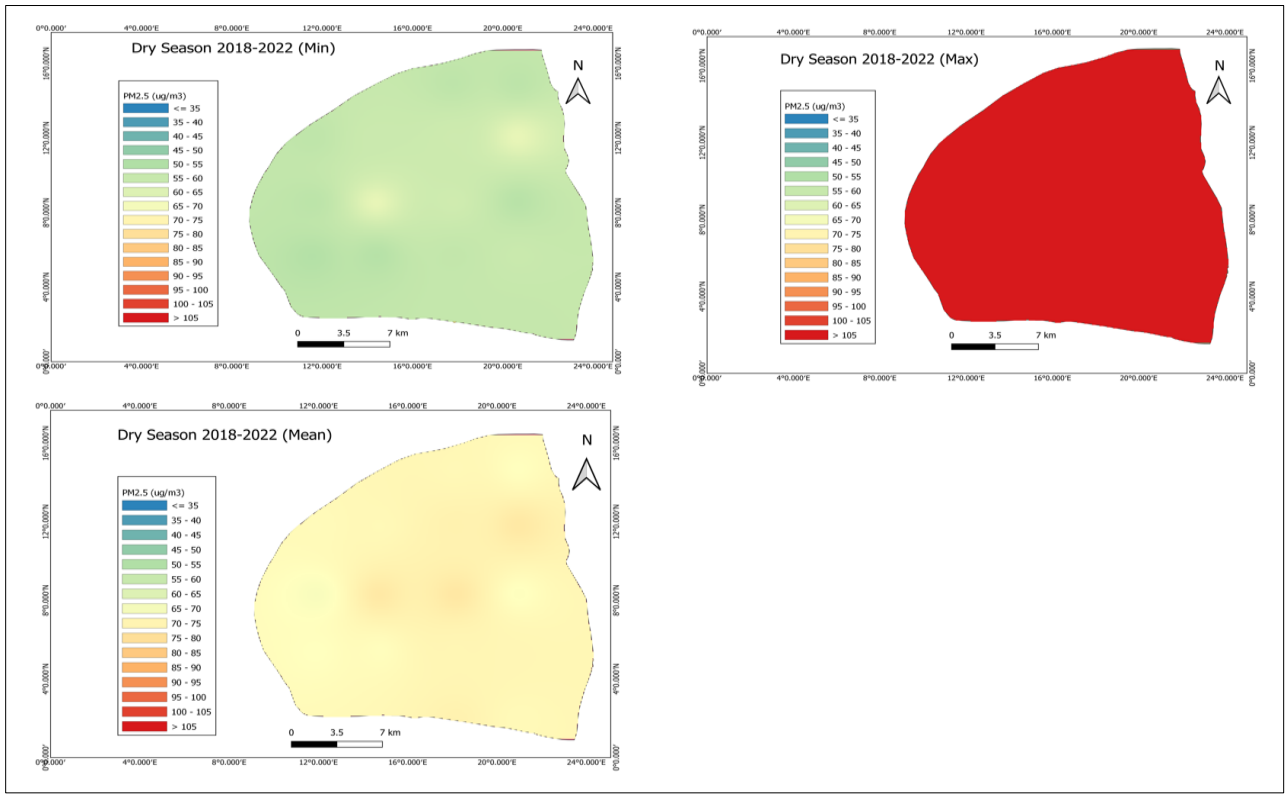

Figure 29: Spatial distribution of estimated $PM_{2.5}$ in dry season 2018-2022

3.8.8 Rainy season 2018-2022

Figure 30 shows the spatial distributions of estimated $PM_{2.5}$ in the rainy season of 2018-2022. The minimum estimated $PM_{2.5}$ concentrations are below 35 $\mu g/m^3$ except Gounghin and Kossodo which has minimum $PM_{2.5}$ concentrations between 35 $\mu g/m^3$ and 40 $\mu g/m^3$. The maximum $PM_{2.5}$ concentrations estimated in all the areas are between 40 $\mu g/m^3$ and 60 $\mu g/m^3$ except Gounghin, Kossodo, Tanghin, Ouagadougou International Airport, and Patte d'Oie which has maximum concentrations between 65 $\mu g/m^3$ and 75 $\mu g/m^3$. The mean estimated $PM_{2.5}$ concentrations in these areas are between 55 $\mu g/m^3$ and 60 $\mu g/m^3$. These higher values of estimated $PM_{2.5}$ in these areas are due to the industrial activities (Gounghin and Kossodo) and the emissions from traffic (Tanghin, Ouagadougou International Airport, and Patte d'Oie). However, the estimated mean $PM_{2.5}$ concentrations in these areas are lower than the estimated mean $PM_{2.5}$ concentrations in these areas in 2012-2017. Gounghin and Kossodo estimated mean $PM_{2.5}$ decreased by 14 % from their levels in the rainy season of 2012-2017. Kossodo, Tanghin,

Ouagadougou International Airport, and Patte d'Oie estimated mean $PM_{2.5}$ concentrations also decreased by 8 % from their levels in the rainy season of 2012-2017. These decreases are due to the Covid-19 lockdown where industrial activities were temporarily closed or reduced and human movement was constrained to comply with social distancing and other safety measures. In 2020, the mean estimated $PM_{2.5}$ in the rainy season were all below 40 $\mu g/m^3$ as shown in **Figure 51** in appendices.

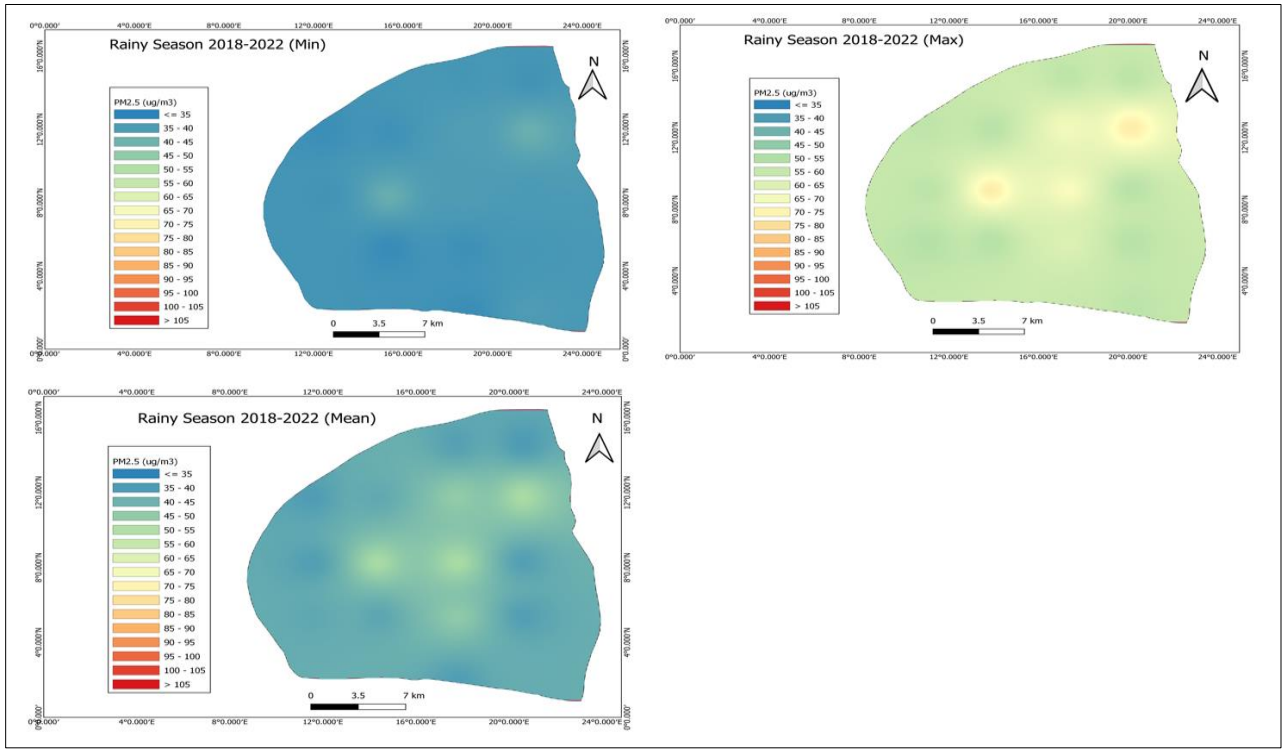

Figure 30: Spatial distribution of estimated $PM_{2.5}$ in rainy season 2018-2022

CONCLUSION

Air monitoring stations in Ouagadougou are sparse despite the fact that the city is considered one of the most polluted cities in Africa. Due to the lack of surface observations of air pollution, few prior studies on fine particulate matter ($PM_{2.5}$) has been done in the city. This work used satellite aerosol optical depth (AOD) and meteorological parameters develop models for estimating daily $PM_{2.5}$ in Ouagadougou. In all the models, AOD is the most important parameter in estimating $PM_{2.5}$ in the city. In the SLR model, AOD explains 0.52 (R^2=0.52) of the variations of surface $PM_{2.5}$ in the city whereas in the DT, RF, and XGBoost model, AOD alone explains 0.56, 0.58, and 0.54 of the variations of surface $PM_{2.5}$ respectively.

Addition of meteorological parameters increase the performance of the models and hence the models' ability to explain the variations of surface $PM_{2.5}$. This indicates that meteorological parameters influence the variability of $PM_{2.5}$ which in a long-run signifies that climate change may have significant impacts on air quality. XGBoost outperforms all the supervised models explaining 0.87 (R^2=0.87) of the variations of $PM_{2.5}$ with an RMSE of 15.8 $\mu g/m^3$. The upgraded XGBoost (semi-supervised XGBoost) model has an R^2 of 0.97 and an RMSE of 8.3 $\mu g/m^3$ indicating that with addition of the lots of unlabeled data, the variability of surface $PM_{2.5}$ in the city can be captured. This confirms the hypothesis that an effective model can be developed for estimating $PM_{2.5}$ from satellite AOD and meteorological parameters using small amount of labeled data and lots of unlabeled data in Ouagadougou by the incorporation of a semi-supervised algorithm. The results from the semi-supervised XGBoost model reveal that estimated $PM_{2.5}$ concentrations in the city are 2 to 4 times higher than the WHO 24-hour limit of 15 $\mu g/m^3$ in the rainy season and 2 to 22 times higher than the WHO 24-hour limit in the dry season. However, in the rainy season, most days have $PM_{2.5}$ concentrations within the US EPA 24-hour standard of 35 $\mu g/m^3$. These variations confirm the hypothesis that $PM_{2.5}$ concentrations vary from season to season. The higher $PM_{2.5}$ concentrations in the dry season are due the dust from the Sahara desert transported by the Harmattan winds whilst the lower $PM_{2.5}$ concentrations in the rainy season are due to wet deposition and gravitational settling of $PM_{2.5}$ particles. The month of March has the highest $PM_{2.5}$ concentrations when the Harmattan reaches its peak whilst the month of August has the lowest $PM_{2.5}$ concentrations when the frequency of rainfall is high. The average annual estimated $PM_{2.5}$

concentrations in the city are 11 to 14 times higher than the WHO average annual standard of 5 $\mu g/m^3$.

However, the yearly $PM_{2.5}$ trend is not direct, it increases and decreases but generally there is a slight increasing trend from 2000-2022. In hourly basis, observed $PM_{2.5}$ concentrations in Ouaga 2000 (U.S embassy) peak between 5:00 am and 9:00 am and 4:00 pm and 11:00 pm. The morning peak is due to emissions from traffic when people are going to work. The evening peak is due to emissions from heavy traffic when people are returning from work. The industrial areas (Gounghin and Kossodo) and areas around the center of the city are the major polluting areas due to combustions from industrial activities and emissions from heavy traffic hypothesis respectively. This verifies the hypothesis that $PM_{2.5}$ concentrations at the industrial areas and the center are higher than the other areas of the city. From these findings, the main hypothesis that $PM_{2.5}$ pollution concentrations can be estimated using AOD and meteorological parameters is confirmed. This research has provided information on $PM_{2.5}$ concentrations in Ouagadougou that would help in epidemiological studies, city planning, and air quality decision-making. The developed semi-supervised XGBoost model can be applied to other areas outside Ouagadougou where $PM_{2.5}$ surface monitoring is limited to estimate daily $PM_{2.5}$.

Based on the findings, the following recommendations are proposed to help address the alarming concentrations of $PM_{2.5}$ in Ouagadougou and enhance air quality management in the city: Firstly, industries should implement cleaner production methods and install emission control systems such as electrostatic precipitators, scrubbers, and bag filters to capture pollutants and minimize their emissions. This would require a concerted effort by industry stakeholders to adopt and enforce stricter emission standards. Secondly, public transportation should be prioritized and the use of electric vehicles should be encouraged. Investing in a well-connected and efficient public transportation system would help reduce the number of private vehicles on the road, leading to lower emissions. Thirdly, there should be an improvement in traffic management and an increase in the number of paved roads within the city. Efficient traffic flow and well-maintained roads can help reduce congestion and the associated emissions from idling vehicles. Lastly, launching public awareness campaigns to educate residents about the health risks associated with high $PM_{2.5}$ concentrations is recommended. Raising awareness about the sources of air pollution

and the importance of individual actions in reducing emissions can lead to behavioral changes and a collective effort to improve air quality.

In the future, inclusion of more $PM_{2.5}$ observed data and other comprehensive data, such as land use and land cover, human flow, Emission inventories and traffic-related factors into the developed model would help to achieve a more rigorous $PM_{2.5}$ estimation model. Additionally, in the future, the semi-supervised XGBoost would be improved to construct a forecasting model for projecting future air quality to help residents in planning outdoor activities reasonably, and to assist policy makers in taking pollution prevention measures in advance.

BIBLIOGRAPHY REFERENCES

Aditya Sai Srinivas, T., Somula, R., Govinda, K., Saxena, A., & Pramod Reddy, A. (2019). Estimating rainfall using machine learning strategies based on weather radar data. *International Journal of Communication Systems*, *33*(13). https://doi.org/10.1002/DAC.3999

Aili, A., & Oanh, N. T. K. (2015). Effects of dust storm on public health in desert fringe area: Case study of northeast edge of Taklimakan Desert, China. *Atmospheric Pollution Research*, *6*(5), 805–814. https://doi.org/10.5094/APR.2015.089

Al-Saadi, J., Szykman, J., Pierce, R. B., Kittaka, C., Neil, D., Chu, D. A., Remer, L., Gumley, L., Prins, E., Weinstock, L., MacDonald, C., Wayland, R., Dimmick, F., & Fishman, J. (2005). Improving National Air Quality Forecasts with Satellite Aerosol Observations. *Bulletin of the American Meteorological Society*, *86*(9), 1249–1262. https://doi.org/10.1175/BAMS-86-9-1249

Altman, N., & Krzywinski, M. (2015). Points of Significance: Simple linear regression. *Nature Methods*, *12*(11), 999–1000. https://doi.org/10.1038/NMETH.3627

Assamnew, A. D., & Mengistu Tsidu, G. (2023). Assessing improvement in the fifth-generation ECMWF atmospheric reanalysis precipitation over East Africa. *International Journal of Climatology*, *43*(1), 17–37. https://doi.org/10.1002/JOC.7697

Babu Saheer, L., Bhasy, A., Maktabdar, M., & Zarrin, J. (2022). Data-Driven Framework for Understanding and Predicting Air Quality in Urban Areas. *Frontiers in Big Data*, *5*, 14. https://doi.org/10.3389/FDATA.2022.822573/BIBTEX

Barnston, & G., A. (1992). Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score. *WtFor*, *7*(4), 699–700. https://doi.org/10.1175/1520-0434(1992)007

Benas, N., Beloconi, A., & Chrysoulakis, N. (2013). Estimation of urban PM10 concentration, based on MODIS and MERIS/AATSR synergistic observations.

Atmospheric Environment, *79*, 448–454. https://doi.org/10.1016/J.ATMOSENV.2013.07.012

Boman, J., Lindén, J., Thorsson, S., Holmer, B., & Eliasson, I. (2009). A tentative study of urban and suburban fine particles (PM2.5) collected in Ouagadougou, Burkina Faso. *X-Ray Spectrometry*, *38*(4), 354–362. https://doi.org/10.1002/XRS.1173

Bougoudis, I., Demertzis, K., Iliadis, L., Anezakis, V. D., & Papaleonidas, A. (2016). Semi-supervised hybrid modeling of atmospheric pollution in urban centers. *Communications in Computer and Information Science*, *629*, 51–63. https://doi.org/10.1007/978-3-319-44188-7_4/COVER

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., & Guo, Y. (2018). A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. *Science of The Total Environment*, *636*, 52–60. https://doi.org/10.1016/J.SCITOTENV.2018.04.251

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Y., Mancini, M., Zhu, X., & Akata, Z. (2022). *Semi-Supervised and Unsupervised Deep Visual Learning: A Survey*. http://arxiv.org/abs/2208.11296

Cohen, A. J., Anderson, H. R., Ostro, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J. M., & Smith, K. (2005). The global burden of disease due to outdoor air pollution. *Journal of Toxicology and Environmental Health. Part A*, *68*(13–14), 1301–1307. https://doi.org/10.1080/15287390590936166

de Sousa, K., Sparks, A., Ashmall, W., van Etten, J., & Solberg, S. (2020). chirps: API Client for the CHIRPS Precipitation Data in R. *Journal of Open Source Software*, *5*(51), 2419. https://doi.org/10.21105/joss.02419

Engel-Cox, J. A., Holloman, C. H., Coutant, B. W., & Hoff, R. M. (2004a). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment*, *38*(16), 2495–2509. https://doi.org/10.1016/J.ATMOSENV.2004.01.039

Engel-Cox, J. A., Holloman, C. H., Coutant, B. W., & Hoff, R. M. (2004b). Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmospheric Environment*, *38*(16), 2495–2509. https://doi.org/10.1016/J.ATMOSENV.2004.01.039

Etyemezian, V., Tesfaye, M., Yimer, A., Chow, J. C., Mesfin, D., Nega, T., Nikolich, G., Watson, J. G., & Wondmagegn, M. (2005). Results from a pilot-scale air quality study in Addis Ababa, Ethiopia. *Atmospheric Environment*, *39*(40), 7849–7860. https://doi.org/10.1016/J.ATMOSENV.2005.08.033

Fisher, S., Bellinger, D. C., Cropper, M. L., Kumar, P., Binagwaho, A., Koudenoukpo, J. B., Park, Y., Taghian, G., & Landrigan, P. J. (2021). Air pollution and development in Africa: impacts on health, the economy, and human capital. *The Lancet. Planetary Health*, *5*(10), e681–e688. https://doi.org/10.1016/S2542-5196(21)00201-1

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, *2*. https://doi.org/10.1038/SDATA.2015.66

Gharibzadeh, M., & Saadat Abadi, A. R. (2022). Estimation of surface particulate matter (PM2.5 and PM10) mass concentration by multivariable linear and nonlinear models using remote sensing data and meteorological variables over Ahvaz, Iran. *Atmospheric Environment: X*, *14*, 100167. https://doi.org/10.1016/J.AEAOA.2022.100167

Gleixner, S., Demissie, T., & Diro, G. T. (2020). Did ERA5 Improve Temperature and Precipitation Reanalysis over East Africa? *Atmosphere 2020, Vol. 11, Page 996*, *11*(9), 996. https://doi.org/10.3390/ATMOS11090996

Gouba, B., Sirima, M. H., & Naon, B. (2021). Measuring the Physico-Chemical Impact of Wastewater from the Open Sewer in an Industrial Area: Case of the Kossodo Industrial Area in the City of Ouagadougou in Burkina Faso. *Environment and Pollution*, *10*(1), 46. https://doi.org/10.5539/EP.V10N1P46

Gupta, P., & Christopher, S. A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research: Atmospheres*, *114*(D14), 14205. https://doi.org/10.1029/2008JD011496

He, Q., Wang, M., Hung, S., Yim, L., & Yim, S. H. L. (n.d.). *The spatiotemporal relationship between PM2.5 and AOD in China: Influencing factors and Implications for satellite PM2.5 estimations by MAIAC AOD*.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., Quattrochi, D. A., Sarnat, J. A., & Liu, Y. (2013). Estimating ground-level PM2.5 concentrations in the southeastern U.S. using geographically weighted regression. *Environmental Research*, *121*, 1–10. https://doi.org/10.1016/J.ENVRES.2012.11.003

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., Estes, M. G., Estes, S. M., Quattrochi, D. A., Puttaswamy, S. J., & Liu, Y. (2014). Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment*, *140*, 220–232. https://doi.org/10.1016/J.RSE.2013.08.032

Hutchison, K. D. (2003). Applications of MODIS satellite data and products for monitoring air quality in the state of Texas. *Atmospheric Environment*, *37*(17), 2403–2412. https://doi.org/10.1016/S1352-2310(03)00128-6

Islam, N., Toha, T. R., Islam, M. M., & Ahmed, T. (2023). Spatio-temporal Variation of Meteorological Influence on PM2.5 and PM10 over Major Urban Cities of

Bangladesh. *Aerosol and Air Quality Research*, *23*(1). https://doi.org/10.4209/AAQR.220082

Jiang, H., Wang, X., & Sun, C. (2022). Predicting PM2.5 in the Northeast China Heavy Industrial Zone: A Semi-Supervised Learning with Spatiotemporal Features. *Atmosphere 2022, Vol. 13, Page 1744*, *13*(11), 1744. https://doi.org/10.3390/ATMOS13111744

Jobson, J. D. (1991). *Multiple Linear Regression*. 219–398. https://doi.org/10.1007/978-1-4612-0955-3_4

Joharestani, M. Z., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere 2019, Vol. 10, Page 373*, *10*(7), 373. https://doi.org/10.3390/ATMOS10070373

Kamarul Zaman, N. A. F., Kanniah, K. D., & Kaskaoutis, D. G. (2017). Estimating Particulate Matter using satellite based aerosol optical depth and meteorological variables in Malaysia. *Atmospheric Research*, *193*, 142–162. https://doi.org/10.1016/J.ATMOSRES.2017.04.019

Kanabkaew, T. (2013). Prediction of Hourly Particulate Matter Concentrations in Chiangmai, Thailand Using MODIS Aerosol Optical Depth and Ground-Based Meteorological Data. *EnvironmentAsia*, *6*(2), 65–70. https://doaj.org/article/2d46c80b90ca4b1da3e63f92435f3b88

Kelishadi, R., & Poursafa, P. (2010). Air pollution and non-respiratory health hazards for children. *Archives of Medical Science : AMS*, *6*(4), 483–495. https://doi.org/10.5114/AOMS.2010.14458

Kumar, K., & Pande, B. P. (2022). Air pollution prediction with machine learning: a case study of Indian cities. *International Journal of Environmental Science and Technology*, 1–16. https://doi.org/10.1007/S13762-022-04241-5/TABLES/7

Kumar, N. (2010). What can affect AOD-PM2.5 association? *Environmental Health Perspectives*, *118*(3). https://doi.org/10.1289/EHP.0901732

Kumar, N., Chu, A., & Foster, A. (2007). An empirical relationship between PM2.5 and aerosol optical depth in Delhi Metropolitan. *Atmospheric Environment*, *41*(21), 4492–4503. https://doi.org/10.1016/J.ATMOSENV.2007.01.046

Lawrence, M. G. (2005). The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications. *Bulletin of the American Meteorological Society*, *86*(2), 225–234. https://doi.org/10.1175/BAMS-86-2-225

Lee, J., Hong, J. W., Lee, K., Hong, J., Velasco, E., Lim, Y. J., Lee, J. B., Nam, K., & Park, J. (2019). Ceilometer Monitoring of Boundary-Layer Height and Its Application in Evaluating the Dilution Effect on Air Pollution. *Boundary-Layer Meteorology*, *172*(3), 435–455. https://doi.org/10.1007/S10546-019-00452-5/FIGURES/6

Léon, J. F., Barthélémy Akpo, A., Bedou, M., Djossou, J., Bodjrenou, M., Yoboué, V., & Liousse, C. (2021a). PM2.5surface concentrations in southern West African urban areas based on sun photometer and satellite observations. *Atmospheric Chemistry and Physics*, *21*(3), 1815–1834. https://doi.org/10.5194/acp-21-1815-2021

Léon, J. F., Barthélémy Akpo, A., Bedou, M., Djossou, J., Bodjrenou, M., Yoboué, V., & Liousse, C. (2021b). PM2.5surface concentrations in southern West African urban areas based on sun photometer and satellite observations. *Atmospheric Chemistry and Physics*, *21*(3), 1815–1834. https://doi.org/10.5194/acp-21-1815-2021

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., & Eck, T. F. (2010). Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *ACP*, *10*(21), 10399–10420. https://doi.org/10.5194/ACP-10-10399-2010

Li, J., Ge, X., He, Q., & Abbas, A. (2021). Aerosol optical depth (AOD): spatial and temporal variations and association with meteorological covariates in Taklimakan desert, China. *PeerJ*, *9*. https://doi.org/10.7717/PEERJ.10542

Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, *6*(3–4), 228–241. https://doi.org/10.1016/J.ECOINF.2010.12.003

Li, L., Zhou, X., Kalo, M., & Piltner, R. (2016). Spatiotemporal Interpolation Methods for the Application of Estimating Population Exposure to Fine Particulate Matter in the Contiguous U.S. and a Real-Time Web Application. *International Journal of Environmental Research and Public Health*, *13*(8), 749. https://doi.org/10.3390/IJERPH13080749

Lin, L., Liang, Y., Liu, L., Zhang, Y., Xie, D., Yin, F., & Ashraf, T. (2022). Estimating PM2.5 Concentrations Using the Machine Learning RF-XGBoost Model in Guanzhong Urban Agglomeration, China. *Remote Sensing 2022, Vol. 14, Page 5239*, *14*(20), 5239. https://doi.org/10.3390/RS14205239

Lindén, J. (2011). Nocturnal Cool Island in the Sahelian city of Ouagadougou, Burkina Faso. *International Journal of Climatology*, *31*(4), 605–620. https://doi.org/10.1002/JOC.2069

Lindén, J., Thorsson, S., Boman, J., & Holmer, B. (2012). *Urban Climate and Air pollution in Ouagadougou, Burkina Faso : An overview of results from five field studies*.

Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A. M., Guo, Y., Tong, S., Coelho, M. S. Z. S., Saldiva, P. H. N., Lavigne, E., Matus, P., Valdes Ortega, N., Osorio Garcia, S., Pascal, M., Stafoggia, M., Scortichini, M., Hashizume, M., Honda, Y., Hurtado-Díaz, M., Cruz, J., … Kan, H. (2019). Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. *The New England Journal of Medicine*, *381*(8), 705–715. https://doi.org/10.1056/NEJMOA1817364

Liu, Y., Paciorek, C. J., & Koutrakis, P. (2009). Estimating regional spatial and temporal variability of PM(2.5) concentrations using satellite data, meteorology, and land use information. *Environmental Health Perspectives*, *117*(6), 886–892. https://doi.org/10.1289/EHP.0800123

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J., & Koutrakis, P. (2005). Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing. *Environmental Science & Technology*, *39*(9), 3269–3278. https://doi.org/10.1021/ES049352M

Liu, Z. N., Yu, X. Y., Jia, L. F., Wang, Y. S., Song, Y. C., & Meng, H. D. (2021). The influence of distance weight on the inverse distance weighted method for ore-grade estimation. *Scientific Reports 2021 11:1*, *11*(1), 1–8. https://doi.org/10.1038/s41598-021-82227-y

Lou, C., Liu, H., Li, Y., Peng, Y., Wang, J., & Dai, L. (2017). Relationships of relative humidity with PM2.5 and PM10 in the Yangtze River Delta, China. *Environmental Monitoring and Assessment*, *189*(11), 1–16. https://doi.org/10.1007/S10661-017-6281-Z/METRICS

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating ground-level PM2.5 in China using satellite remote sensing. *Environmental Science & Technology*, *48*(13), 7436–7444. https://doi.org/10.1021/ES5009399

Malings, C., Westervelt, D. M., Hauryliuk, A., Presto, A. A., Grieshop, A., Bittner, A., Beekmann, M., & Subramanian, R. (2020). Application of low-cost fine particulate mass monitors to convert satellite aerosol optical depth to surface concentrations in North America and Africa. *Atmospheric Measurement Techniques*, *13*(7), 3873–3892. https://doi.org/10.5194/amt-13-3873-2020

McFarlane, C., Isevulambire, P. K., Lumbuenamo, R. S., Ndinga, A. M. E., Dhammapala, R., Jin, X., McNeill, V. F., Malings, C., Subramanian, R., & Westervelt, D. M. (2021). First Measurements of Ambient PM2.5 in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo Using Field-calibrated Low-cost Sensors. *Aerosol and Air Quality Research*, *21*(7), 200619. https://doi.org/10.4209/AAQR.200619

Miles, J. (2014). R Squared, Adjusted R Squared. *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.STAT06627

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G.,

Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/ESSD-13-4349-2021

Nana, B., Sanogo, O., Savadogo, P., Daho, T., Bouda, M., & Koulidiati, J. (2012). Air Quality Study in Urban Centers: Case Study of Ouagadougou, Burkina Faso. *FUTY Journal of the Environment*, *7*(1). https://doi.org/10.4314/FJE.V7I1.1

Ogunjobi, K. O., He, Z., & Simmer, C. (2008). Spectral aerosol optical properties from AERONET Sun-photometric measurements over West Africa. *Atmospheric Research*, *88*(2), 89–107. https://doi.org/10.1016/J.ATMOSRES.2007.10.004

Ouarma, I., Nana, B., Haro, K., Béré, A., & Koulidiati, J. (2020). Assessment of Pollution Levels of Suspended Particulate Matter on an Hourly and a Daily Time Scale in West African Cities: Case Study of Ouagadougou (Burkina Faso). *Journal of Geoscience and Environment Protection*, *08*(11), 119–138. https://doi.org/10.4236/GEP.2020.811007

Paciorek, C. J., & Liu, Y. (2010). AOD-PM2.5 association: Paciorek and Liu respond. *Environmental Health Perspectives*, *118*(3). https://doi.org/10.1289/EHP.0901732R

Petkova, E. P., Jack, D. W., Volavka-Close, N. H., & Kinney, P. L. (2013). Particulate matter pollution in African cities. *Air Quality, Atmosphere and Health*, *6*(3), 603–614. https://doi.org/10.1007/S11869-013-0199-6

Plessis, K. Du, & Kibii, J. (2021). Applicability of CHIRPS-based satellite rainfall estimates for South Africa. *Journal of the South African Institution of Civil Engineering*, *63*(3), 43–54. https://doi.org/10.17159/2309-8775/2021/V63N3A4

Qu, L., Xiao, H., Zheng, N., Zhang, Z., & Xu, Y. (2017). Comparison of four methods for spatial interpolation of estimated atmospheric nitrogen deposition in South China. *Environmental Science and Pollution Research*, *24*(3), 2578–2588. https://doi.org/10.1007/s11356-016-7995-0

Sanjeev, D. (2021). Implementation of Machine Learning Algorithms for Analysis and Prediction of Air Quality. *International Journal of Engineering Research & Technology*, *10*(3). https://doi.org/10.17577/IJERTV10IS030323

Schaap, M., Apituley, A., Timmermans, R. M. A., Koelemeijer, R. B. A., & De Leeuw, G. (2009). Atmospheric Chemistry and Physics Exploring the relation between aerosol optical depth and PM 2.5 at Cabauw, the Netherlands. In *Atmos. Chem. Phys* (Vol. 9). www.atmos-chem-phys.net/9/909/2009/

Schober, P., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Shen, S., & Leptoukh, G. G. (2011). Estimation of surface air temperature over central and eastern Eurasia from MODIS land surface temperature. *Environmental Research Letters*, *6*(4). https://doi.org/10.1088/1748-9326/6/4/045206

Shi, X., & Brasseur, G. P. (2020). The Response in Air Quality to the Reduction of Chinese Economic Activities During the COVID-19 Outbreak. *Geophysical Research Letters*, *47*(11), e2020GL088070. https://doi.org/10.1029/2020GL088070

Sirithian, D., & Thanatrakolsri, P. (2022). Relationships between Meteorological and Particulate Matter Concentrations (PM2.5 and PM10) during the Haze Period in Urban and Rural Areas, Northern Thailand. *Air, Soil and Water Research*, *15*. https://doi.org/10.1177/11786221221117264/ASSET/IMAGES/LARGE/10.1177_11786221221117264-FIG6.JPEG

Somda, D. D. (2018). Inventaire d'émissions de polluants atmosphériques issus du trafic routier à Ouagadougou (Burkina Faso). - References - Scientific Research Publishing. (n.d.). Retrieved April 7, 2023, from https://www.scirp.org/reference/referencespapers.aspx?referenceid=2861701

Song, W., Jia, H., Huang, J., & Zhang, Y. (2014). A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in

China. *Remote Sensing of Environment*, *154*, 1–7. https://doi.org/10.1016/J.RSE.2014.08.008

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130. https://doi.org/10.11919/J.ISSN.1002-0829.215044

Sun, Y. L., Wang, Z. F., Fu, P. Q., Yang, T., Jiang, Q., Dong, H. B., Li, J., & Jia, J. J. (2013). Aerosol composition, sources and processes during wintertime in Beijing, China. *Atmospheric Chemistry and Physics*, *13*(9), 4577–4592. https://doi.org/10.5194/ACP-13-4577-2013

Tai, P. K. A. P. K. (2012). *Impact of Climate Change on Fine Particulate Matter \((PM_{2.5})\) Air Quality*. https://dash.harvard.edu/handle/1/10445627

Tian, J., & Chen, D. (2010). A semi-empirical model for predicting hourly ground-level fine particulate matter (PM2.5) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sensing of Environment*, *114*(2), 221–229. https://doi.org/10.1016/J.RSE.2009.09.011

US EPA, O. (n.d.). *National Ambient Air Quality Standards (NAAQS) for PM*. Retrieved April 7, 2023, from https://19january2021snapshot.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqs-pm

Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

van Donkelaar, A., Martin, R. v., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010a). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environmental Health Perspectives*, *118*(6), 847–855. https://doi.org/10.1289/EHP.0901623

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010b). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environmental Health Perspectives*, *118*(6), 847–855. https://doi.org/10.1289/EHP.0901623

Wan, Z. (2013). *Collection-6 MODIS Land Surface Temperature Products Users' Guide*.

Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21), 2095. https://doi.org/10.1029/2003GL018174

Wang, Q., Zeng, Q., Tao, J., Sun, L., Zhang, L., Gu, T., Wang, Z., & Chen, L. (2019). Estimating PM2.5 Concentrations Based on MODIS AOD and NAQPMS Data over Beijing–Tianjin–Hebei. *Sensors 2019, Vol. 19, Page 1207*, *19*(5), 1207. https://doi.org/10.3390/S19051207

Wang, Z., Chen, L., Tao, J., Zhang, Y., & Su, L. (2010). Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. *Remote Sensing of Environment*, *114*(1), 50–63. https://doi.org/10.1016/J.RSE.2009.08.009

Weber, R. (1991). Estimator for the Standard Deviation of Wind Direction Based on Moments of the Cartesian Components. *Journal of Applied Meteorology and Climatology*, *30*(9), 1341–1353. https://doi.org/10.1175/1520-0450(1991)030

Westervelt, D. M., Horowitz, L. W., Naik, V., Tai, A. P. K., Fiore, A. M., & Mauzerall, D. L. (2016). Quantifying PM2.5-meteorology sensitivities in a global climate model. *Atmospheric Environment*, *142*, 43–56. https://doi.org/10.1016/J.ATMOSENV.2016.07.040

World Population Prospects: The 2010 Revision. (2011).

World's Most Polluted Cities in 2022 - PM2.5 Ranking | IQAir. (n.d.). Retrieved April 30, 2023, from https://www.iqair.com/world-most-polluted-cities

Xi, X., & Sokolik, I. N. (2015). Seasonal dynamics of threshold friction velocity and dust emission in Central Asia. *Journal of Geophysical Research: Atmospheres*, *120*(4), 1536–1564. https://doi.org/10.1002/2014JD022471

Xu, X., & Zhang, C. (2020). Estimation of ground-level PM2.5 concentration using MODIS AOD and corrected regression model over Beijing, China. *PloS One*, *15*(10). https://doi.org/10.1371/JOURNAL.PONE.0240430

Xue, T., Zheng, Y., Geng, G., Zheng, B., Jiang, X., Zhang, Q., & He, K. (2017). Fusing Observational, Satellite Remote Sensing and Air Quality Model Simulated Data to Estimate Spatiotemporal Variations of PM2.5 Exposure in China. *RemS*, *9*(3), 221. https://doi.org/10.3390/RS9030221

Zhai, S., Jacob, D. J., Wang, X., Shen, L., Li, K., Zhang, Y., Gui, K., Zhao, T., & Liao, H. (n.d.). *Fine particulate matter (PM2.5) trends in China, 2013-2018: contributions from meteorology*. http://106.37.208.233:20035/

Zhang, D., Du, L., Wang, W., Zhu, Q., Bi, J., Scovronick, N., Naidoo, M., Garland, R. M., & Liu, Y. (2021). A machine learning model to estimate ambient PM2.5 concentrations in industrialized highveld region of South Africa. *Remote Sensing of Environment*, *266*, 112713. https://doi.org/10.1016/J.RSE.2021.112713

Zhang, L., Gu, Z., Yu, C., Zhang, Y., & Cheng, Y. (2016). Surface charges on aerosol particles - Accelerating particle growth rate and atmospheric pollution. *Indoor and Built Environment*, *25*(3), 437–440. https://doi.org/10.1177/1420326X16643799/ASSET/IMAGES/LARGE/10.1177_1420326X16643799-FIG1.JPEG

Zhang, L., Liu, P., Zhao, L., Wang, G., Zhang, W., & Liu, J. (2021). Air quality predictions with a semi-supervised bidirectional LSTM neural network. *Atmospheric Pollution Research*, *12*(1), 328–339. https://doi.org/10.1016/J.APR.2020.09.003

Zhao, Y., Wang, L., Zhang, N., Huang, X., Yang, L., & Yang, W. (2023). Co-Training Semi-Supervised Learning for Fine-Grained Air Quality Analysis. *Atmosphere 2023, Vol. 14, Page 143*, *14*(1), 143. https://doi.org/10.3390/ATMOS14010143

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., & Qiu, Y. (2017). Analysis of influential factors for the relationship between PM2.5 and AOD in Beijing. *Atmospheric Chemistry and Physics*, *17*(21), 13473–13489. https://doi.org/10.5194/ACP-17-13473-2017

Zhu, X. (Jerry). (2005). *Semi-Supervised Learning Literature Survey*. https://minds.wisconsin.edu/handle/1793/60444

APPENDICES

The spatial distribution of estimated $PM_{2.5}$ in Ouagadougou according to each year is shown;

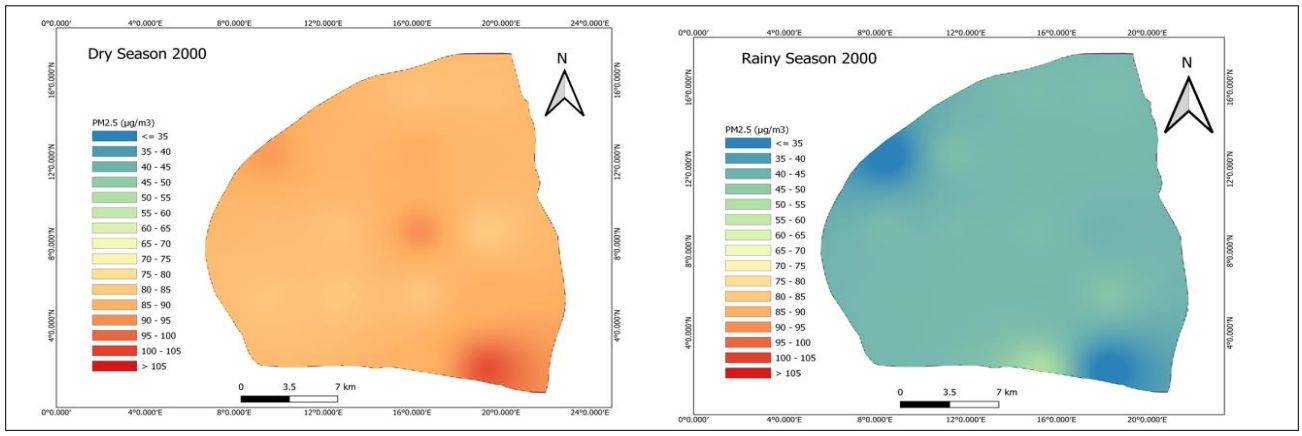

Figure 31: Spatial distribution of estimated $PM_{2.5}$ in 2000

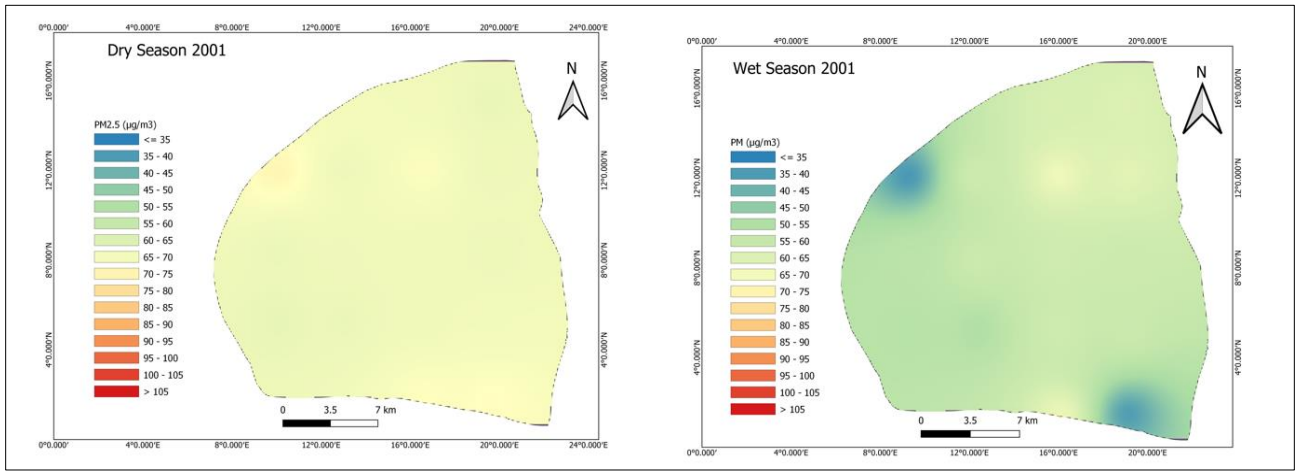

Figure 32: Spatial distribution of estimated $PM_{2.5}$ in 2001

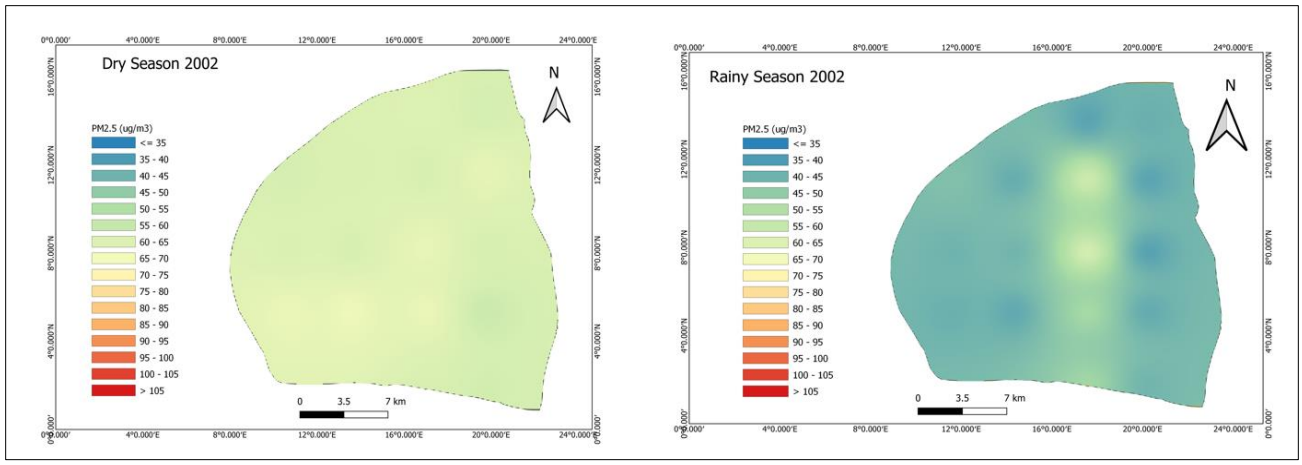

Figure 33: Spatial distribution of estimated $PM_{2.5}$ in 2002

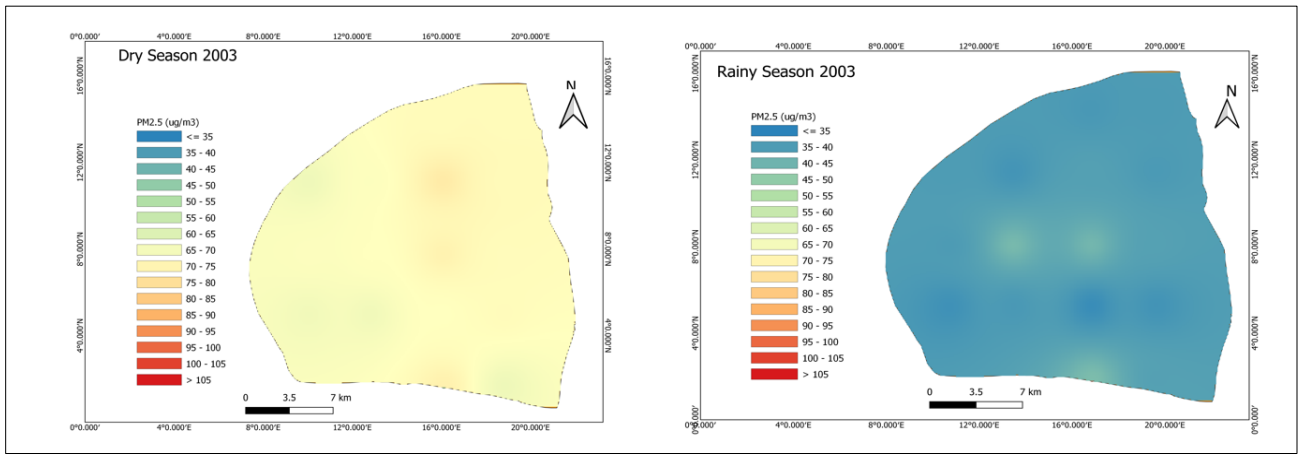

Figure 34: Spatial distribution of estimated $PM_{2.5}$ in 2003

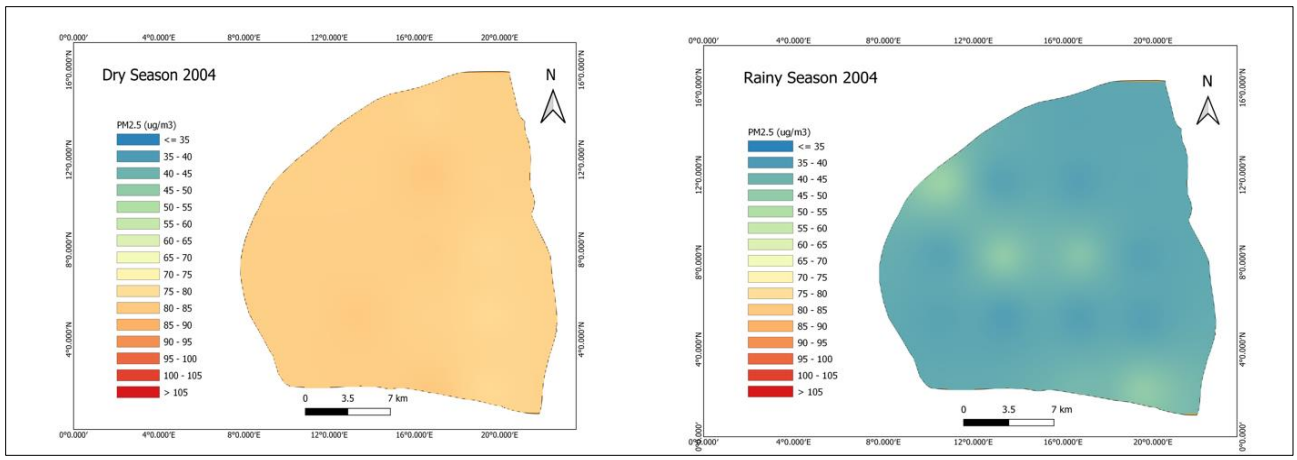

Figure 35: Spatial distribution of estimated $PM_{2.5}$ in 2004

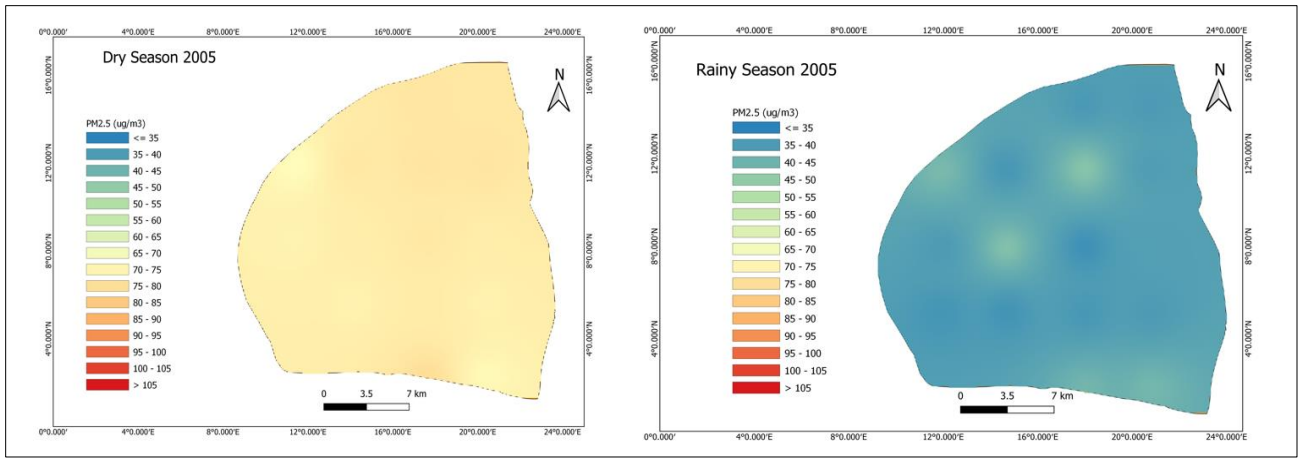

Figure 36: Spatial distribution of estimated $PM_{2.5}$ in 2005

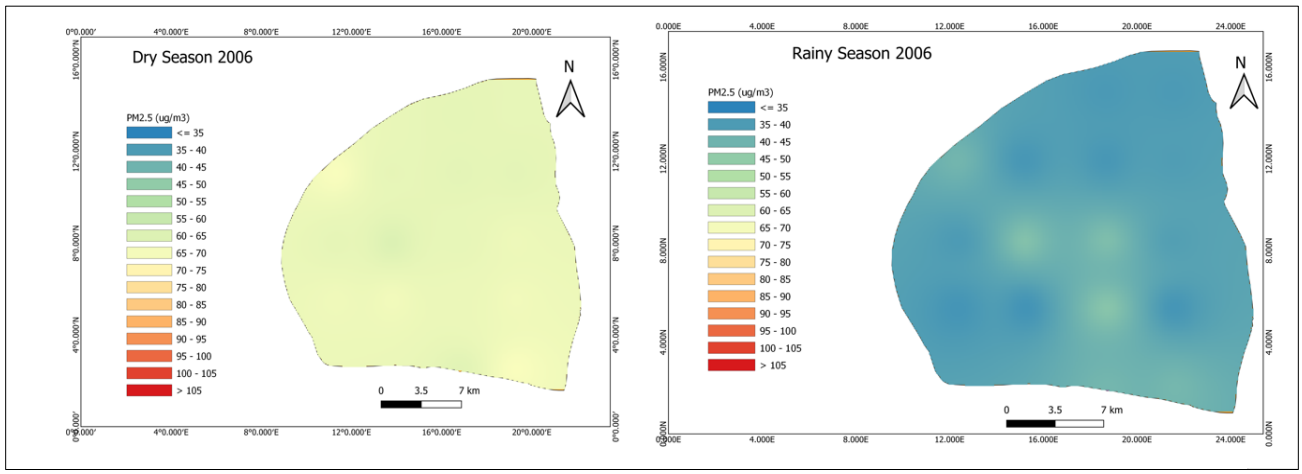

Figure 37: Spatial distribution of estimated $PM_{2.5}$ in 2006

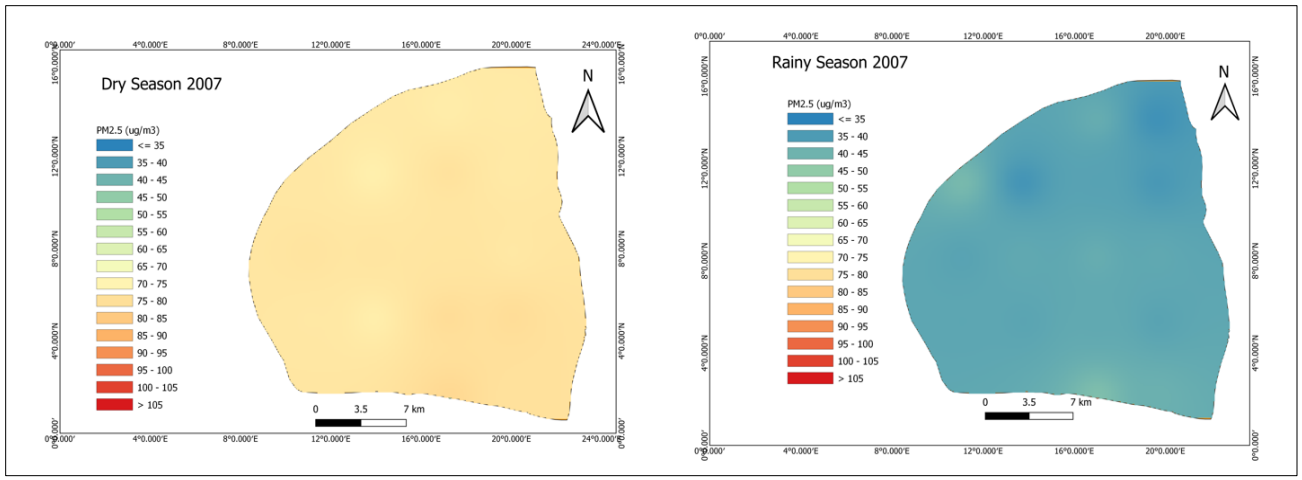

Figure 38: Spatial distribution of estimated $PM_{2.5}$ in 2007

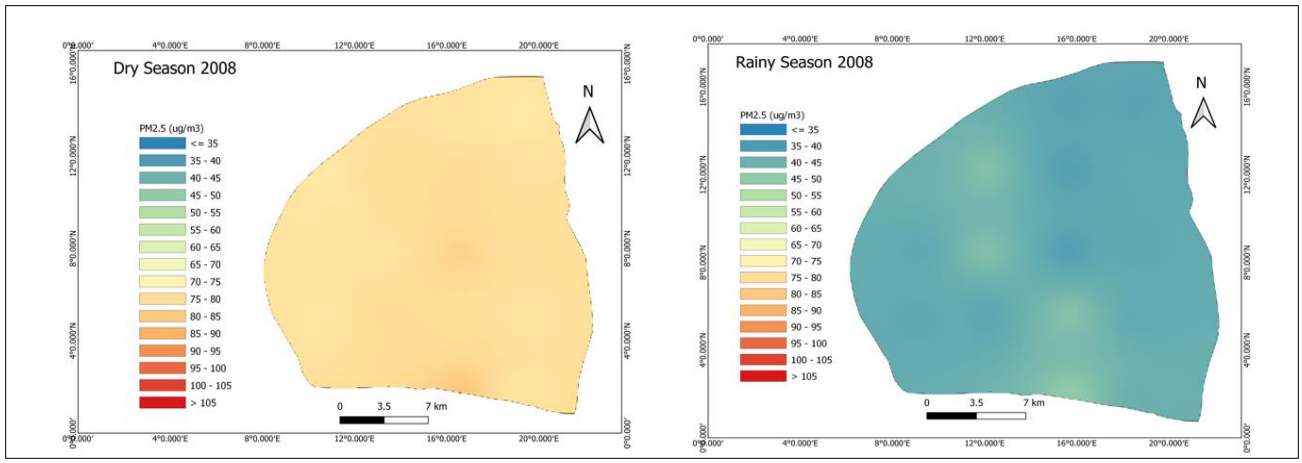

Figure 39: Spatial distribution of estimated $PM_{2.5}$ in 2008

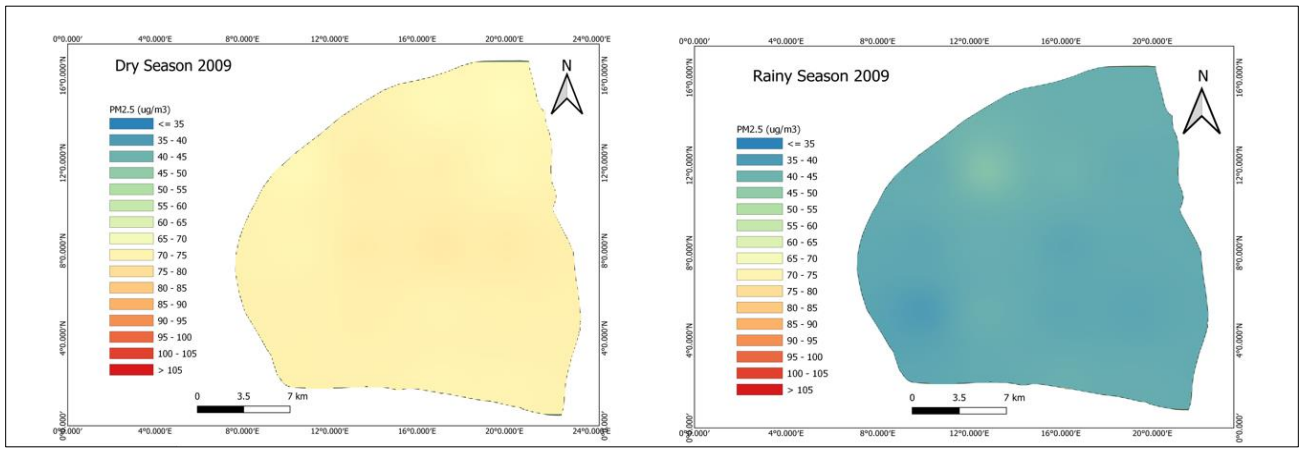

Figure 40: Spatial distribution of estimated $PM_{2.5}$ in 2009

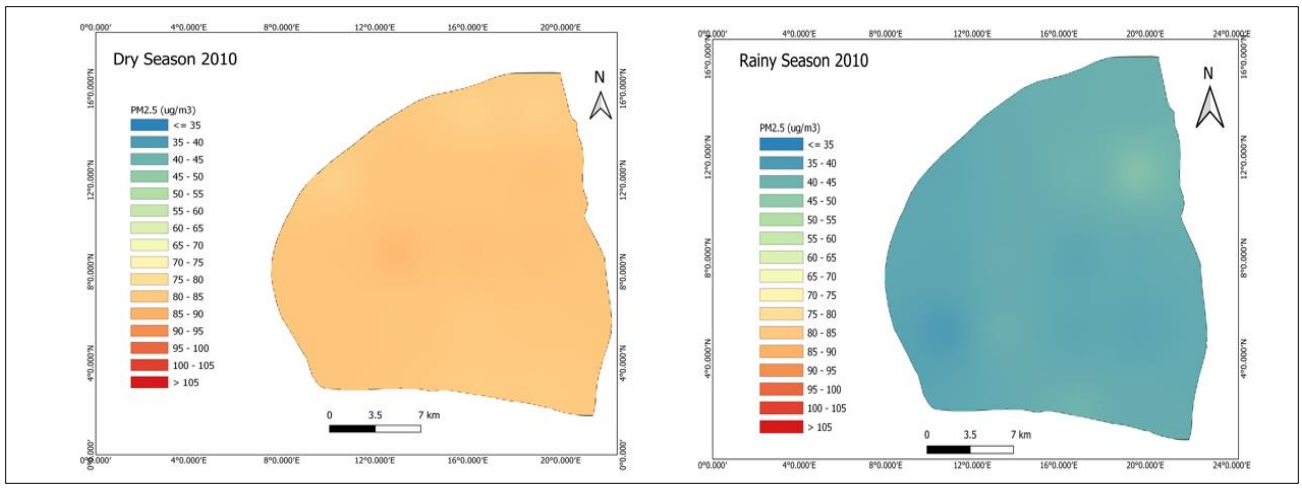

Figure 41: Spatial distribution of estimated $PM_{2.5}$ in 2010

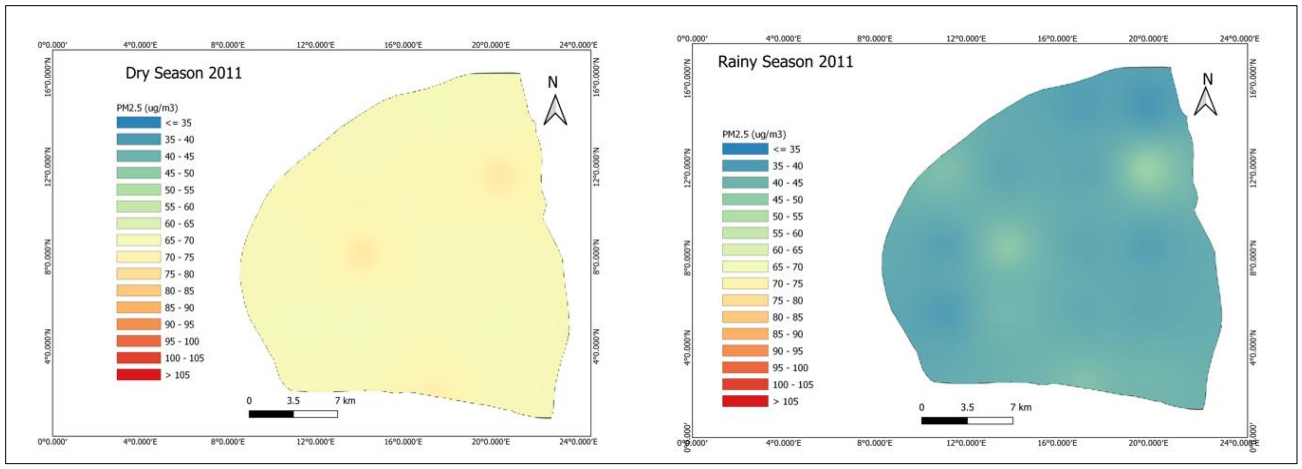

Figure 42: Spatial distribution of estimated $PM_{2.5}$ in 2011

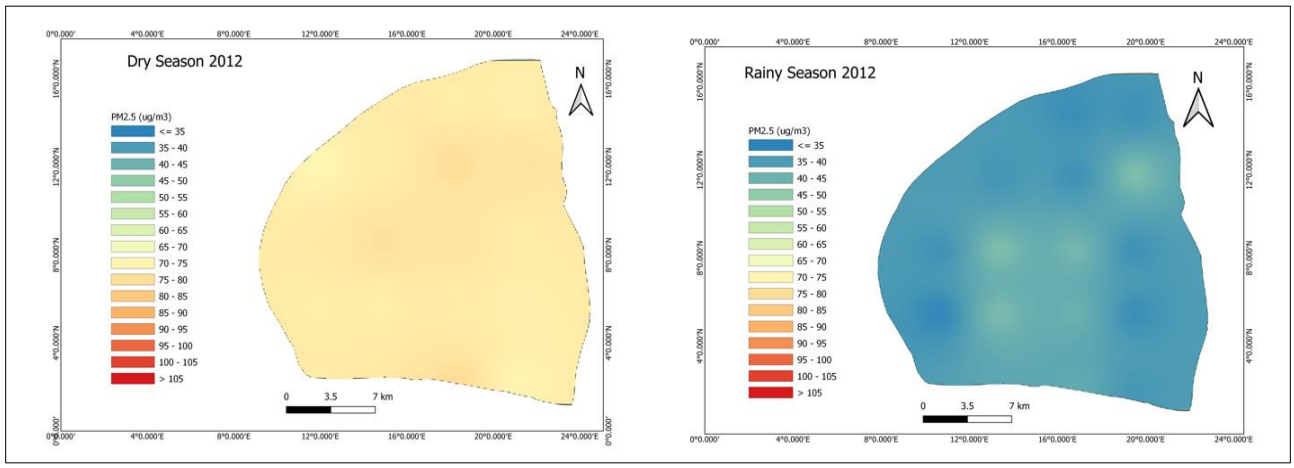

Figure 43: Spatial distribution of estimated $PM_{2.5}$ in 2012

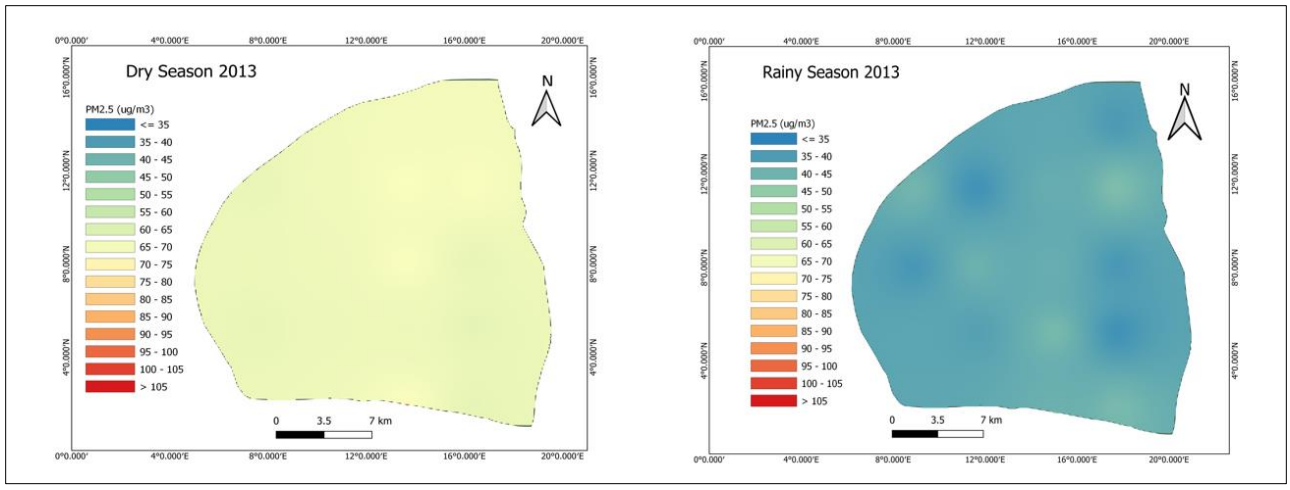

Figure 44: Spatial distribution of estimated $PM_{2.5}$ in 2013

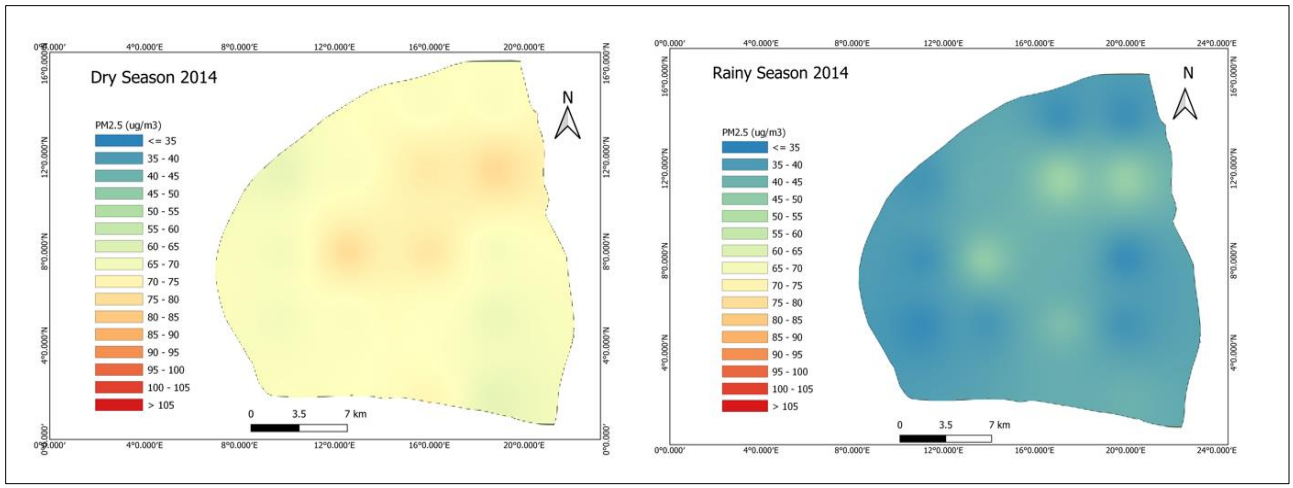

Figure 45: Spatial distribution of estimated $PM_{2.5}$ in 2014

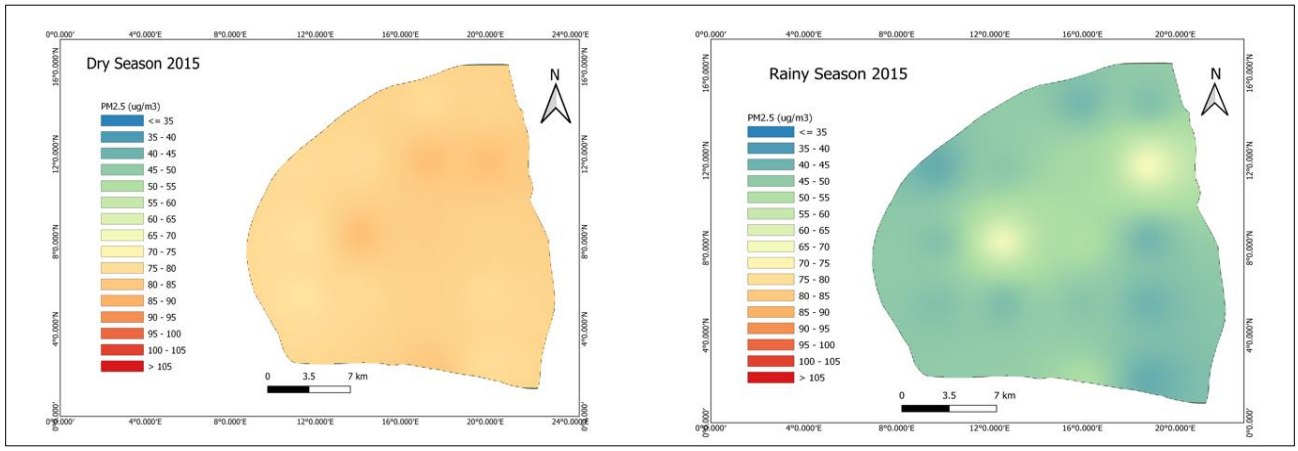

Figure 46: Spatial distribution of estimated $PM_{2.5}$ in 2015

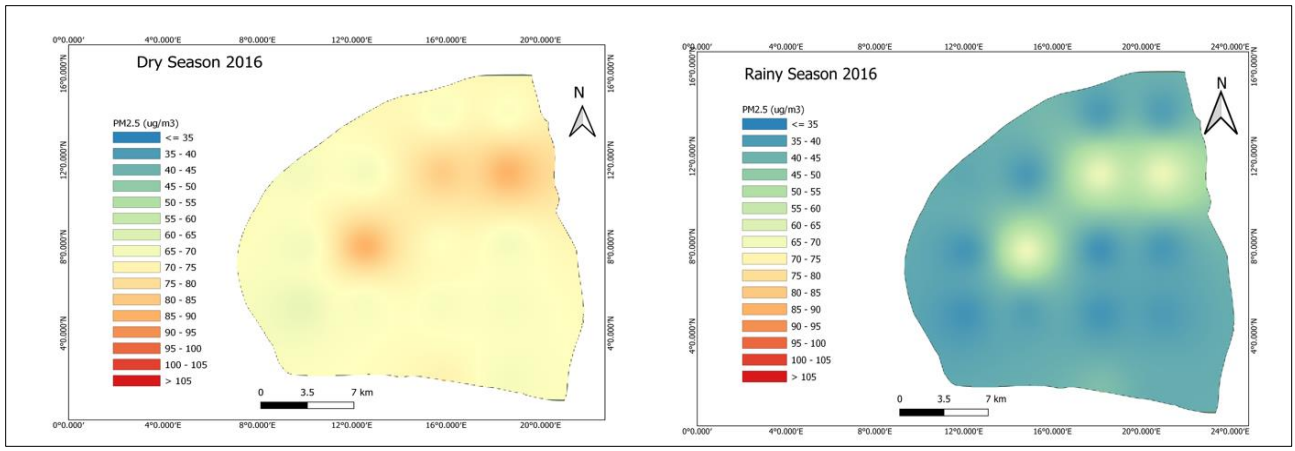

Figure 47: Spatial distribution of estimated $PM_{2.5}$ in 2016

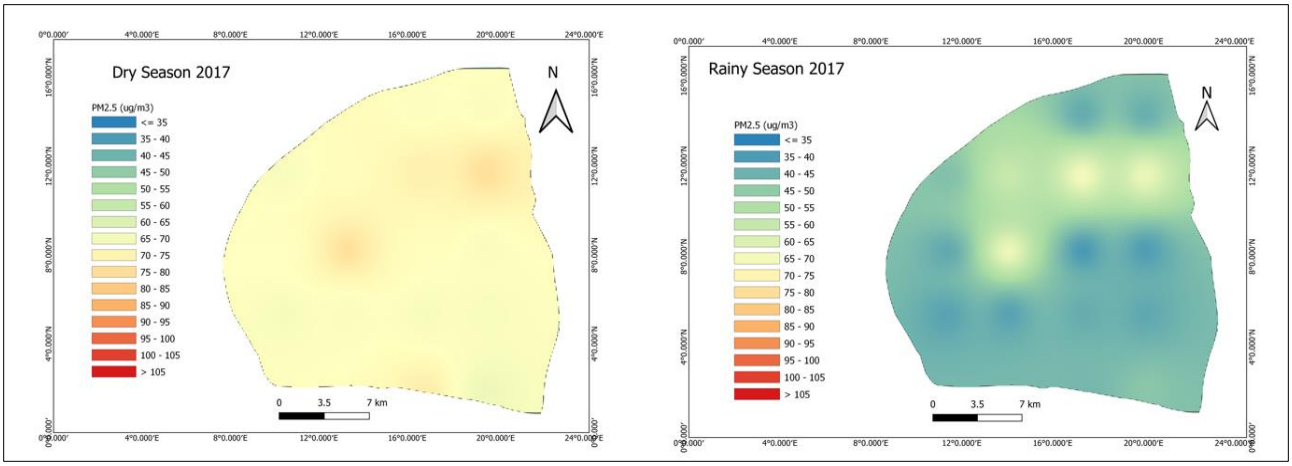

Figure 48: Spatial distribution of estimated $PM_{2.5}$ in 2017

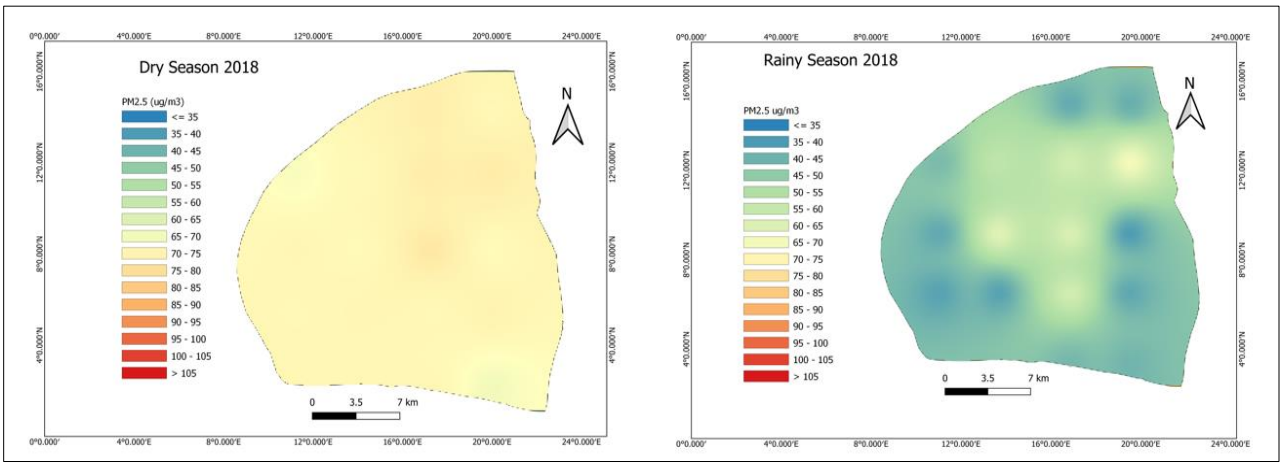

Figure 49: Spatial distribution of estimated $PM_{2.5}$ in 2018

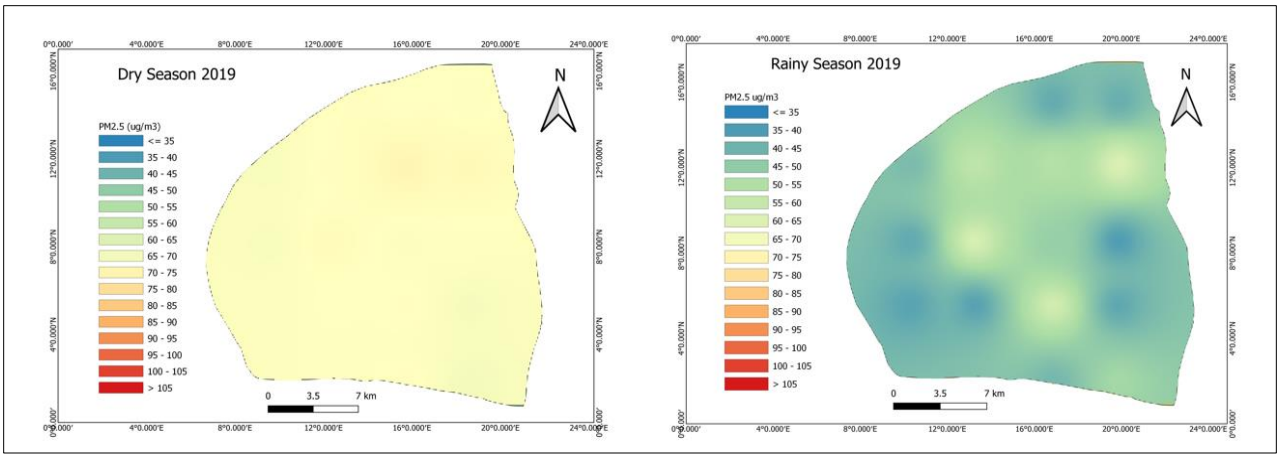

Figure 50: Spatial distribution of estimated $PM_{2.5}$ in 2019

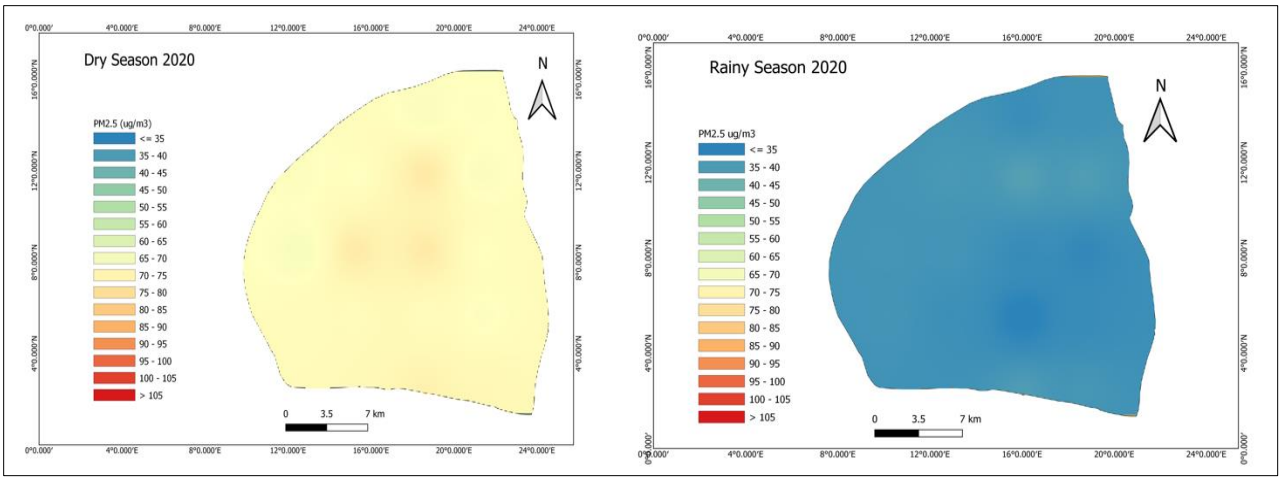

Figure 51: Spatial distribution of estimated $PM_{2.5}$ in 2020

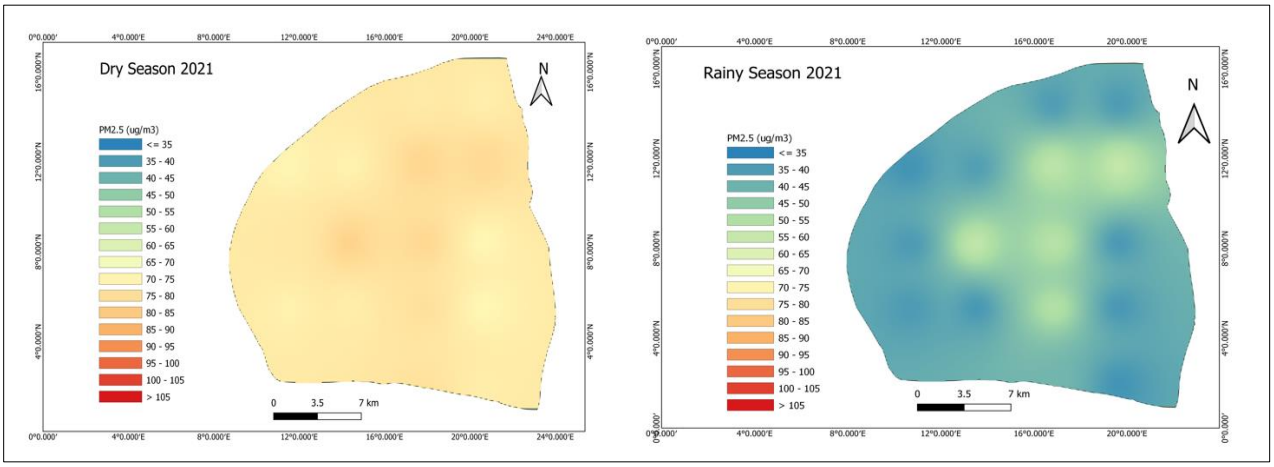

Figure 52: Spatial distribution of estimated $PM_{2.5}$ in 2021

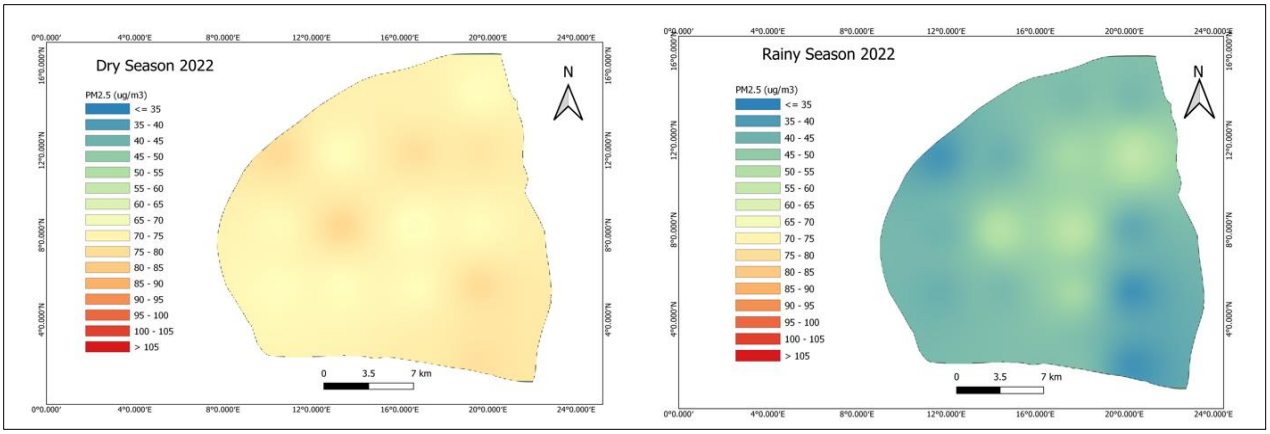

Figure 53: Spatial distribution of estimated $PM_{2.5}$ in 2022

TABLE OF CONTENTS